MULTIVARIATE GENERALIZABILITY OF 2015 NATIONAL EXAMINATIONS COUNCIL SCHOOL CERTIFICATE EXAMINATION OBJECTIVE TEST IN ELECTRICAL INSTALLATIONS AND MAINTENANCE WORKS

By

BAMIDELE, Samuel Tunde
01/680H006

A Ph.D. THESIS SUBMITTED TO THE DEPARTMENT OF SOCIAL SCIENCES EDUCATION, FACULTY OF EDUCATION, UNIVERSITY OF ILORIN, NIGERIA, IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY (Ph D) IN EDUCATIONAL RESEARCH, MEASUREMENT AND EVALUATION

SUPERVISOR: Prof. H.O. OWOLABI

CERTIFICATION

This is to certify this thesis entitled "**Multivaraiate Generalizability of 2015 National Examinations Council School Certificate Examination Objective Test in Electrical Installations and Maintenance Works"** was carried out by **Bamidele, Samuel Tunde (01/680H006)** and had been read and approved as meeting part of the requirements of the Department of Social Sciences Education, Faculty of Education, University of Ilorin, Ilorin Nigeria, for the award of Doctor of Philosophy (ph d) in Educational research, Measurement and Evaluation.

……………………………… ………………………
Prof. H. O, Owolabi Date
(Project Supervisor)

……………………………… ………………………
Prof. Bolanle O. Olawuyi Date
(Head of Department)

……………………………… ………………………
Dr. R. W. Okunloye Date
(Departmental Ph. D PG Coordinator)

……………………………… ………………………
Prof. Sinat M. Salman Date
Dean Faculty of Education

……………………………… ………………………
External Exerminer Date

DEDICATION

This research is dedicated to the Almighty God the Author and the Finisher of my faith who has in His infinite mercy seen me through the programme. May His name be glorified forever.

It is also dedicated to my lovely wife Mrs. Adenike O. Bamidele, and my children Peace, Precious, Praise, Peculiar and Progress.

ACKNOWLEGEMENTS

Firstly, I am grateful to the Almighty God whose guidance, protection, patience and kindness saw me through this programme. I humbly appreciate the tireless efforts of my amiable and indefatigable Supervisor, Professor H.O. Owolabi who encouraged and

motivated me until the successful completion of this work. I pray that God will continue to take him to greater heights.

In the same vein, I specially acknowledge some Lecturers of the Department of Social Sciences Education such as Profs. O.E. Abdullahi, F.O.A. Olasheinde-Williams, A. A. Jekayinfa, B.O. Olawuyi (H.O.D.), Drs. A. A. Ogunlade, R.W. Okunloye (PhD PG Coordinator), O.O.I. Amali, A. Yusuf, M. Bello, D.Daramola, A. Balogun, A. Jimoh and other academic and non-academic staff of the entire Faculty of Education for their support. May God bless them.

I wish to also express my profound gratitude to my parents S. A. Bamidele (late), Felicia Asake (late) and my only precious sister Iyabode Adebayo towards their early caring and education given to me. My appreciations also goes to the following people, Mr. Ojo of NECO office Minna, Mallam Narogo Minna, Elder D.A. Bankole , Dr S.B. Adebara, Mr. M.O. Ayodele, Dr M.D Ibrahim, M.K Raji, Kehinde Ajiboye, Salihu Kolawole, Abraham .Y.G, Mohammed, Ndaba, Moh'd .I . Moh'd (Kalifa), Wuraola J.A. Engr Iyanda Francis and Haruna .I.S.

I will be an ingrate if the efforts of the following colleagues are not appreciated: Mrs. Jumoke Oladele, Winston, Peter Eshun, Cobbina Andrew, Kobbina Paul, Officer Hammad, Dr. Mahmud Jamiu, Osho, Mary, Dr Halimat, Orisamuko Folusho, Dr. Adewuni A.D. ,Dr. Akande Joseph A., Mr. Ojo Rapheal, Ajibade Abraham, Femi Akano, Dr Ogunjimi, Abdulmumuni Yekeen, Rasaki, Banji and Dr Mayowa,I pray that God will continue to be with them all.

My sincere appreciation also goes to my brother Mr. E. A. Orisamuko (late) for his assistance during my undergraduate programme. I appreciate my friends and club members: Salihu Bamidele, Kayode Davis Bamidele, James Kayode Agboluaje, Kolawole Ogunleye, Engr. Adebayo Samuel (Late), Orisamuko J, Sunday, Lanre Owolabi, Kunle Owolabi, Prof. Awodele Oludele, Awodele Oluwole and my able Landlord Mr. Stephen Ogundele, may God be with them for all their support both morally and financially.

Finally I am very grateful to my mother in-law, Mrs. S.M Oluwole for her prayers and words of encouragement as well as Pastor and Mrs Adeoti, Pastor and Mrs. Ojo Abiodun, Pastor and Mrs. Olu Emmanuel, Pastor Augustine, Joy, Adesola, Lola, Segun, Dada and a host of others that I can not mention may God be with them all for their general support.

TABLE OF CONTENTS

Title Page i
Certification Page ii
Dedication
iii

Acknowledgements
iv
Table of Content
vi
List of Tables
viii
List of Figures
ix
Abstract x

CHAPTER ONE: INTRODUCTION

Background to the Study 1
Statement of the Problem 8

Purpose of the Study
10
Research Questions
11
Scope of the Study
12
Operational Definition of Terms
12
Significance of the Study
13

CHAPTER TWO: REVIEW OF RELATED LITERATURE

Achievement Testing

15

Psychometric Properties of Achievement Test

25

Establishment of Public Examination Bodies in Nigeria

33

Psychometric Theory of Reliability

43

Measurement Error and Need for Measurement Errors

55

Random, Systematic Errors

57

Common Sources of Measurement Error

57

Generalizability Theory

63

Designs in Gcneralizability Theory

76

Classical Test Theory and Generalizability Theory

94

Concept of Multivariate Generalizability Theory

99

Empirical Studies on Multivariate Generalizability Theory

105

Appraisal of Related Literature Review

113

CHAPTER THREE: RESEARCH METHODOLOGY

Research Design
115

Population, Sample and Sampling Techniques
116

Instrumentation
117

Procedure for Data Collection
117

Procedure for Data Analysis Techniques
118

CHAPTER FOUR: DATA ANALYSIS AND RESULTS

Results
120

Summary of Findings
125

CHAPTER FIVE: DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS

Discussion of Findings
127

Conclusions
130
Educational Implications
130
Recommendations
131
Suggestion for Further Studies
131
References
132
Appendices
142

LIST OF TABLES

Table
Page

1 Sources of Variability in One-Facet Measurement
6

2 Estimated Variance Components and there corresponding total
variance percentage for 2015 NECO Electrical Installations and
Maintenance Works objective test
120

3 Estimated Variance Component for Persons in 2015 NECO
Electrical Installations and Maintenance Works objective test
121

4 Estimated Variance Components for Items in 2015 NECO Electrical
Installations and Maintenance Works objective
122

5 Estimated Variance Component for persons by items in 2015
NECO Electrical Installations and Maintenance Works objective test
123

6 Relative error variance and Generalizability coefficient for 2015
NECO Electrical Installations and Maintenance Works objective test
124

7 Relative error variance and Dependability coefficient for 2015

NECO Electrical Installations and Maintenance Works objective test 124

LIST OF FIGURES

Figure
Page

1 Comparison of Classical Test Theory and Generalizability Theory 97
2 Venn diagrams for a One-Facet, crossed (p x i) Design
Shavelson &Webb, (1991)
115

ABSTRACT

It is the goal of measurement to minimize errors in test scores. The extent to which errors from likely sources as examiners, items on a test, examiners and test administration occasions affect measures in the Senior School Certificate Examinations conducted by National Examinations Council (NECO) especially in technical subjects is not known. Generalizability coefficient obtained through analysis of variance components helps to address this problem with result of Dependability Index. This study therefore estimated the Generalizability and dependability coefficients of 2015 SSCE objective test in Electrical Installations and Maintenance Works (EI&MW). The objectives of the study were to: (i) estimate the variance in the 2015 SSCE objective test in EI&MW due to persons; (ii) estimate the variance in the 2015 SSCE objective test in EI&MW due to items used in the test; (iii) estimate the variance in 2015 SSCE objective test in EI&MW due to the interaction of persons by items; (iv) obtain the Generalizbility coefficient in the 2015 SSCE objective test in EI&MW; and (v) determine the Dependability coefficient in the 2015 SSCE objective test in EI&MW.

One-Facet Fully Crossed Design was used to carry out the study. The population comprised all the Senior Secondary School Students in Nigeria. Senior Secondary School Three (SS3) students offering Electrical Installations and Maintenance Works participated in the study. Out of the 3,448 students offering Electrical Installations and Maintenance Works in Nigeria, samples of 1,198 were selected. Senior School Certificate Examinations 2015 objective test in EI&MW was adopted as the instrument. The data obtained were analysed using Variance Components (VARCOMP); relative error variance, absolute error variance, Generalizability and Dependabilty coefficients statistics.

The findings of this study were:

i. variance accounted for in 2015 SSCE Electrical Installations and Maintenance Works for persons (σ^2p) = 0.02 equivalent to 8% of the total variance;

ii. variance accounted for in 2015 SSCE Electrical Installations and Maintenance Works items (σ^2i) = 0.03 equivalent to 12% of the total variance;

iii. variance accounted for in 2015 SSCE Electrical Installations and Maintenance Works persons by items (σ^2pi,e) = 0.20 equivalent to 80% of the total variance;

iv. Generalizability coefficient of 2015 SSCE Electrical Installations and Maintenance Works was 0.80 and

v. dependability coefficient of 2015 SSCE Electrical Installations and Maintenance Works was 0.78

The study concluded that Generalizability and dependability coefficients of 2015 objective test in EI&MW were high or acceptable. This revealed that the quality and numbers of items used in EI&MW were of acceptable standard. The study therefore recommended that the quality of the items in Electrical Installations and Maintenance Works objective test should be maintained. Furthermore the Evaluators should endeavour to carry out similar studies in other areas of vocational tests so that inference can also be made.

Word count 455

CHAPTER ONE

INTRODUCTION

Background to the Study

In measurement history, the leading theory for explaining latent trait underlying examinees' test performance is the Classical Test Theory. Classical Test Theory (CCT) describes how error can influence observed scores. It is a simple model based on the true score theory that introduces three concepts-test scores or observed score (X), true score (T) and error score (E). It recognizes that the characteristics of the testing situation can contribute to measurement error and evaluates sources of error separately. This is the case of test-retest, alternate forms and internal consistency option of reliability of measures. The combined effect of error from the various threats to the reliability of an estimated score is rarely considered. Each method of reliability in CCT yields valuable information but provides only a slice of a bigger picture. Each piece of the source of error is estimated in isolation and fulfills a single objective. How all the various sources of error operate at one time and fit together to influence the overall reliability of the instrument cannot be estimated(Shavelson & Webb,1991a).

CCT deals with the reliabilities of relative decisions, where an individual's score is compared with a reference group and used to rank order the individuals as in norm-referenced measurement is not considered. Absolute reliability, where an individual is compared to a well-defined standard and used to provide the absolute value of an attribute as in criterion-referenced measurement is not considered (Shavelson & Webb1991a).

Generalizability Theory (GT) consists of a conceptual framework and a methodology that enables an investigator to disentangle multiple sources of error in a measurement procedure. The roots of generalizability theory can be found in Classical Test Theory and Analysis of Variance (ANOVA). In particular, the conceptual framework in generalizability theory is unique. Historically, in psychology and education, measurement issues have been addressed principally using Classical Test Theory which postulates that an observed score can be decomposed into a "true" score and a single undifferentiated random error term, *E* (Brennan, 2000).

Generalizability Theory liberalizes Classical Test Theory by providing models and methods that allow an investigator to disentangle multiple sources of error that contribute to *E*. In a sense, then, Classical Test Theory and ANOVA can be viewed as the parents of Generalizability theory. More importantly, however, Generalizability Theory has a unique conceptual framework. Among the concepts in this framework are universes of admissible observations and Generalizability studies, as well as universes of generalization and Decision Studies (Brennan, 2000).

Generally, Generalizability Theory (GT) is a statistical theory about the reliability of behavioural measurements. Reliability refers to the accuracy of generalizing from the person's observed score on a test or other measures (e.g. behaviour observation, opinion) to the average score that person would have received under all the possible conditions that the test user would be equally willing to accept. Generalizability Theory provides an all-at-once way of revealing and comparing the sources of error in a common metric. It also provides estimates of the variance contributed by source and in addition presents estimates of the variance associated with interaction between the various sources, for instance if an instrument is administered on two occasions; Generalizability theory provides estimates of the variance contributed by persons, items and occasions, each of the four possible interactions (persons by items, persons by occasions, items by occasions and persons by items by occasions). It also provides helpful forecasts of the improvements in measurement reliability that can be obtained by altering the numbers of persons, items, occasions etc. This theory enables the decision makers to determine how many occasions, test forms and administrators are needed to obtain reliable and valid scores. Therefore Generalizability theory provides a summary coefficient reflecting the level of reliability called Generalizability coefficient that is analogous to CCT reliability coefficient (Shavelson & Webb, 1991b).

GT according to Watkins et al (1980) deal with two main studies, which are the Generalizability (G) study and the decision (D) study. Generalizability (G) study, which is analogue to reliability in Classical Test Theory (CTT), was described by Shavelson and Webb (2005) as the ratio of the universe score variance to the expected observed score variance. It quantifies the amount of variance association with the different facets under the study (Wan, Li, Fan, Yang and Pan, 2014). Generalizability study just like reliability gives it result as Generalizability coefficient. This is to mean that Generalizability coefficient is obtained from G study.

Decision study which is the second study in GT according to Wan et al (2014), gives information about which protocols are optional for particular measurement situation by generating Generalizability coefficient that could be considered as reliability coefficient for various facets of study. D studies make use of information provided by G study in arriving at its own reliability coefficient called index of dependability. It is index of dependability that is used to determine how dependable measurement behaviour is. Dependability is the accuracy of generalizing an individual's average score he would have received under all the possible condition that the examiner would be equally willing to accept (Nie et al, 2007). Therefore, the dependability of a measurement depends on the accuracy of its generalization.

Dependability subsumes all other aspects of Generalizability theory. This is because the result obtains from the Generalizabilty study is used to obtain dependability.

Gerbil (2013), noted that there are four main indices that would be obtained from a decision study which are; relative error variance σ^2 (δ), Generalizability (G) coefficient Ep^2, Absolute error variance (σ^2Abs) and phi coefficient. Relative error variance (σ^2Abs) is used to make relative decision. Relative decision performs the same function as norm reference as it focuses on the rank ordering of examinees (Shavelson and Webb, 2005). For instance, decisions about secondary school admission or employees selections are relative. Absolute error variance on the other hand is used to make absolute decision. Absolute decision which is analogue to criterion reference focuses on the level of an examinees performance independent of others performance. For instance, if the minimum passing score on the drivers examination is 80% correct, regardless of the performance of others, then such decision is regarded absolute as it considered only the performance of the testee without comparing to performance of others. Generalizabilty (G) coefficient which is the third main indices is the result obtains from G study and which is analogue to reliability coefficient in classical test theory (CTT) while the last is Ph coefficient which is also regarded as index of dependability or dependibilty level and this is the result obtained from the D study.

GT as noted by Gebril (2013) is a powerful tool that is used for estimation of consistencies and inconsistencies in test scores, provision of scenario for test developers which will help them make correct decision related to test development and administration, effective investigation of the relative contribution of various facets to test precision, provision of information about the impact of increase or decreasing the number of task, and test validation in order to optimize measurement precision.

Testing has become one of the most important parameters by which the society adjudges the products of her educational system (Emaikwu, 2012). The essence of testing is to obtain the latent ability of the examinees. According to Rivera (2007), a standardized test is an instrument that has been carefully prepared in accordance with scientific techniques to measure intelligence, aptitude or achievement in school subjects. Standardized tests are often considered high stakes because results are used to make important decisions concerning admission, graduation, and certification.

A test consists of a set of questions or tasks to which a student responds independently and the result of which can be treated in such a way as to provide a quantitative comparison of the performance between and among different students (Nworgu, 1992). A test may be defined as a task or series of tasks used to obtain observations presumed representing educational and psychological traits. Test requires examinees to respond to the items from which the examiners refer something about the attribute being measured. It could be said that an educational test is any means of bringing out for observation and assessment specific attributes or characteristics such as abilities, knowledge, skills or feeling of persons individually or in groups (Abiri, 2007).

Validity is a basic fundamental issue in test development and evaluation as well as fairness. Traditionally, validity is defined as the degree to which a test measures what is claims to be measuring. Validity refers to the degree to which evidence supports inferences based on the test scores while fairness means that all examinees are given comparable opportunities to demonstrate their abilities on the construct a test intends to measure (Messick, 1989). Validity is necessary because of the major impact which test results can have on the stakeholders involved. The validity of a test, according to Johann and Fanns (2008) can only be established through a process of validation and this must ideally be done before the results can be used for any particular purpose. The researchers further explained that in order to carry out such validation, a validation study has to be undertaken, on the basis of which one can arrive at a conclusion as to whether the interpretation and use of the test results are valid.

However, evidence that supports a test's validity is gathered from different approaches, the recognized aspects of test validity are content validity, criterion-related validity and construct validity. Content validity includes any validity strategies that focus on the content of the test. To demonstrate content validity, testers or test developers investigate the degree to which a test is a representative sample of the content of whatever objectives or specifications the test was design to measure, while criterion-related validity may be defined as the experimental demonstration that a test is measuring the construct it claims to be measuring such an experiment could take the form of a differential-groups; on that has or possessed the construct and that does not have or possess the construct. Construct validity compose of the analyses of a test's internal constructs in order to confirm that the test indeed functions as it is intended to function. Analyses of construct validity include correlations between items and the test, discrimination between subgroups, factor analysis and multi-trait-multi-method approaches (Crocker & Alginal, 1986). Thus, among other important areas of item analysis is the assessment of Generalizabilty coefficient of test items. Messick (1989) shifted perspective on validity from a property of a test to that test scores interpretation and validity is now closely associated with the interpretation of test scores. He stated that "validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment".

Reliability refers to the consistency of the scores obtained. That is, how consistent the scores are to each individual from one administration of an instrument to another and from one item to another. Reliability is a measure of how stable, dependable, trust worthy and consistent a test is in measuring the same thing each time (Worthen, Borg & White 1993).

This study focuses on Multivariate Generalizability of 2015 National Examinations Council SSCE objective test items in Electrical Installations and Maintenance Works. It also revealed how the estimate variance components and generalizability coefficient were analysed. The universe of admissible observations is defined by completely crossed one facet: item facet (i) by person facet (p). Each combination of conditions of person x item ($p^\bullet$ x i^o) specifies one single observation of the ability of a given person (p) (John & Jeremy 2012).

Multivariate analysis is the statistical process of simultaneously analyzing multiple independent (or predictor) variables with multiple dependent (outcome or criterion) variables using matrix algebra (most multivariate analyses are a correlation). While these analyses have been part of statistics since the early 1900's, the developments of mainframe and microcomputers and related analytical software have made the once tedious calculations fairly simple and very fast (Shavelson & Webb, 1991b). For all these purposes, Multivariate Generalizability Theory decomposes both observed variances and covariance into components. Table 1 shows the sources of variability of the multivariate one-facet crossed persons by items ($p^\bullet$ x i^o) design.

Table 1: Sources of Variability in One-Facet Measurement

Source of variability	**Type of Variability**	**Variance Notation**
Person (p)	Universe score	σ^2p
Items (i)	Conditions	σ^2i
Person by item Interaction	Residual	σ^2pi,e

Source: Shavelson & Webb (1991)

Electrical Installations and Maintenance as one of the Senior Secondary School subjects attracts a very few number of Senior Secondary School Candidates registering Certificate Examinations (SSCE) because the subject is normally offered by the science students. The approved national curriculum for senior school certificate was developed and produced by the National Educational Research Development Council (NPE, 2013). This syllabus is prepared with the aims, objectives and purposes of the Senior Secondary School Certificate Examinations. The multiple choice items in these examinations are often subjected to different processes of validation. Electrical Installations and Maintenance Works objective test has been designed with blue print /table of specifications that indicate six (6) structures of the domains being measured. The aims and objectives of the curriculum are to test candidates' ability to: provide trained manpower in applied science technology and commerce particularly at sub-professional grade; provide the technical knowledge and vocational skills necessary for agricultural, industrial, commercial and economic development; it also aims the ability to provide people that can apply scientific knowledge to the improvement and solution of environment problems for the use and convenience of man, it gives an introduction to professional studies in engineering and other technologies; It also aims the ability to giving training and impart the necessary skills leading to the production of craftsmen, technicians and other skilled personnel who will be enterprising and self-reliant, and it also enable our young men and women to have an intelligent understanding of the increasing complicity of technology (NPE, 2013).

The aims and objectives will be viewed as measures and each of the content areas is a potential measure and in view of the importance of the decisions made on the basis of Senior School Certificate test results, this study assessed the estimates of variance components and Generalizability coefficient of the 2015 SSCE objective test in Electrical Installations and Maintenance Works in Nigeria.

Statement of the Problem

City and Guild of London Institute was working closely with the Board of Education in London, towards the grouped course certificates and later the board endorsed the conduct of internal examining bodies on a regional scale, opened up a new phase of examining in technical education. By and large, City and Guide of London Institute's examinations were transferred to West African Examination Council (WAEC Technical) on its inauguration. But as a result of the leakages of examination papers in 1977 that resulted into the inauguration of Sogbetun Commission of inquiry. Base on the recommendations of this commission, NECO and NABTEB were established to take over City and Guild of London Institute's examinations from WAEC and take over the conduct of technical and business examination from Royal Society of Arts of London. NECO and NABTEB were later established by decree 69 & 70 of 1993 respectively. Though, most of the questions that time were essay and not objective items. However, with the introduction of entrepreneurship subjects into secondary schools curriculum, NECO introduces some trade subjects to be offered by science-based students; Electrical Installations and Maintenance was just introduced in 2014 by NECO as one of the entrepreneurship subject to be offered by science-based students.

Albert, (1984) applied multivariate GT to the assessment of the students' achievement in art education. Twenty-five Art students rated painting of 60 fourth-grade students with regard to three criteria. The results indicated that Generalizability coefficient was low with respect to different raters and moderate with respect to different topics. Noreen and Jonah (1999) investigated the importance of occasion as a hidden source of error variance in the estimation of the Generalizability of science assessment scores and the interchangeability of science test formats. The univariate Generalizability results show that the explicit recognizing occasion as a facet of error variance altered the interpretation about the substantial sources of error in the measurement, and gave lower estimates of the dependability of science scores. Youzhen, (2007) applied the method of multivariate GT to assess the reliability of the student style questionnaire (SSQ). In particular, random effect variance and covariance components were estimated. The results indicated that the G coefficient were acceptable for the total scale and two of the subscales.

Yonyan, Shu and Shum, (2007) investigated the use of GT to evaluate the quality of an alternative assessment (journal writing) in mathematics. Twenty-nine junior college students wrote journal task on the given topics and two raters marked the tasks using a scoring rubric, constituting a two-facet G-study design in which students were crossed with task and raters. The results showed that increasing the number of tasks had a lager effect on the G coefficient and index of dependability, than increasing the number of raters. Tunde, (2015) assessed multivariate Generalizability of NECO’s 2014 SSCE objective test in Electrical Installations and Maintenance Works. The population for the study consisted of all public senior secondary schools in Kwara State. The study further investigated the estimated variance for persons, items, persons by items and G coefficient. The results revealed that all the estimated variance components and G coefficient were low.

In view of diversified findings in empirical studies on Generalizability theory of test items, it is clear that more studies are still required, equally being an indigenous examination body, that, they administered low quality items and from the studies reviewed only Tunde, (2015) has worked on similar study, but the results cannot be generalized due to lower population size and scope used. Thus, this motivated the researcher to conduct a study to investigate Multivariate Generalizability of 2015 NECO SSCE objective test in Electrical Installations and Maintenance Works in Nigeria. This study intends to assessed the estimate of variance components; the Generalizability and dependability coefficients of 2015 NECO SSCE objective test in Electrical Installations and Maintenance in Nigeria.

Purpose of the Study

The main purpose of the study was to investigate the Multivariate Generalizability of Senior School Certificate Examination Objectives Test in Electrical Installation and Maintenance conducted by National Examination Council in 2015 (June/July) using item-information levels. The study was designed to;

a. estimate the variance component due to persons (testees) in the 2015 NECO SSCE objective test in Electrical Installations and Maintenance Works,

b. estimate the variance component due to items used in the test in the 2015 NECO Senior School Certificate examination objective test in Electrical Installations and Maintenance Works,

c. estimate the variance component in the 2015 NECO Senior School Certificate examination objective test in Electrical Installations and Maintenance Works due to person by items interaction,

d. obtain the Generalizability Coefficient of the 2015 NECO Senior School Certificate examination objective test in Electrical Installations and Maintenance Works.

e. determine the dependability Coefficient of the 2015 NECO Senior School Certificate examination objective test in Electrical Installations and Maintenance Works.

Research Questions

This study specifically intends to find answers to the following research questions:

1. What is the estimate of the variance component due to persons (testees) in the 2015 SSCE objective test in Electrical Installations and Maintenance Works?
2. What is the estimate of the variance component due to items used in the test in the 2015 SSCE objective test in Electrical Installations and Maintenance Works?
3. What is the estimate of the variance component in the 2015 SSCE objective test in Electrical Installations and Maintenance Works due to the interactions of persons by items?
4. What is the Generalizability coefficient of the 2015 Senior School Certificate Examination objective test in Electrical Installations and Maintenance Works?
5. What is the Dependability coefficient of the 2015 Senior School Certificate Examination objective test in Electrical Installations and Maintenance Works?

Scope of the Study

This study investigated the estimate of variance components, Generalizability and Dependability coefficients of the Senior School Certificate Objective Test in Electrical Installation and Maintenance Works in Nigeria. The study was carried out among Senior Secondary School Students in Nigeria. The final year students were purposively chosen for the study due to the fact that the examination is majorly meant for the final year students. One thousand, one hundred and ninety-eight (1,198) senior secondary three (SS3) students that are offering Electrical Installation and Maintenance Works participated in the study. The year 2015 NECO SSCE Electrical Installations and Maintenance Works objective test was adopted and used as the instrument for this study since it is related to the work. Thus, NECO SSCE 2015 Electrical Installations and Maintenance Works multiple choice objective tests were used to collect data and it was analysed using Variance Component (VARCOMP) to appraise the estimate of variance components for persons, items, persons by items, Generalizability and dependability coefficients respectively.

Operational Definition of Terms

The following terms are defined in relation to their usage in the study:

Generalizability Study: - This is a measurement that takes care of all sources of errors in 2015 NECO objective test in Electrical Installations and Maintenance Works.

Dependability: it is the level to which the generalization of items in the 2015 NECO SSCE Electrical Installations and Maintenance Works objective test is accurate and consistent. The acceptable dependability level that would be used in this study is 0.7 for both Index of dependability and generalizability coefficient. Because the least acceptable level in the literature is 0.7.

Multivariate Generalizability: - This is the process of estimating variance components and generalizability coefficient.

Persons: - these are 1,198 Senior Secondary three that participated in the study.

Items: - these are 40 items used as the instrument for collecting data for study.

Persons Variance Component: - This is the level in which examinee differs in their responses from one item to another in the 2015 Electrical Installations and Maintenance Works.

Items Variance Component: - This shows how the items differ from one item to another in term of difficulty in the 2015 Electrical Installations and Maintenance Works..

Persons by Items Variance Component: - this is the difference that occurs between the persons scores due to the interaction effect between the persons and items during the administering and other systematic errors not included in the design.

Electrical Installations and Maintenance Works: - As one of the entrepreneurship subject offered by the science-based students at senior secondary schools' level was designed to provide technical training to meet the demands of the electrical industry and the needs of the individuals.

Significance of the Study

The major reason for carrying out educational research is to solve a particular problem(s) in the society. The Findings in this study would be useful to public examination bodies, test item writers, researchers and teachers respectively. Public examination bodies like West African Examination Council, National Examination Council, Joint Admission and Matriculation Board etc. would benefit from the findings of the study, in that, it would provides them with very specific information about measurement errors and how to maintain high quality items.

Item writers and Researchers would equally benefit from the findings of the study as it would enable guide against all possible sources of measurement errors. It would provide them with the knowledge of the relative importance of the various sources of error and procedures for attainment of generalizabilty and dependability coefficients. The study also provided for them avenue for further research on generalizability theory.

To classroom teachers, the findings of the study would be useful to them. Teachers would be acquainted with knowledge about the existence of multiple errors in examinations, hence to maximize reliability and reduce error or eliminate error in examinations, there is need to estimate as many sources of error as is economically viable in order to determine the level of involvement of each source of error in the scores obtained in examinations. It guided the teachers on how to estimate multiple sources of error which are part of the examination process but not related to the construct being measured (students).

CHAPTER TWO

REVIEW OF RELATED LITERATURE

This chapter deals with the related literature to the topic under consideration. It is an essential tool of research; it helps to get well acquitted with other people’s findings and claims in the areas that relate to any research work embarked upon. It equally helps to indicate gaps to be filled as well as the justification for selecting the problem to be studied. The literature was reviewed under the following sub-headings viz:

a. Achievement Testing

b. Establishment of Public Examination Bodies in Nigeria

c. Psychometric Theory of Reliability

d. Measurement Error and Need for Measurement Errors

e. Random, Systematic Errors and Common Sources of Measurement Error

f. Generalizability Theory

g. Designs in Gcneralizability Theory Studies

h. Classical Test Theory and Generalizability Theory

i. Concept of Multivariate Generalizability Theory

j. Empirical Studies on Multivariate Generalizability Theory

k. Appraisal of Related Literature Review

Achievement Testing

Testing has been fully accepted in most modern societies as the most objective method of decision making in schools, industries, and government establishments. It is now used for admission, recruitment, promotion, placement, evaluation, guidance, research and teaching purposes among others (Emaikwu, 2011). Testing has undergone many changes over the years, from oral testing to standardized testing and to authentic assessment; it has continued to change with educational policy and practice, thus testing is a staple of education.

The test is one of the elusive concepts in education that denied a single and universally acceptable definition, thus is not easy to define a test precisely. The meaning of the concept depends on the individual‘s points of view because tests take so many forms. Clark (2008) defined a test as a systematic procedure for comparing the behaviour of two or more persons. Tests are essentially assessment devices which enable the teacher, the curriculum developer, and the educational planner to make a certain decision. Abiri (2007) defined a test as a task, treatment or situation designed to elicit the behaviour or performance of persons or things with the view to determining or drawing inferences about specific abilities or other attributes of these persons or things. He went further to describe the educational test as a means of bringing out for observation and assessment specific attributes or characteristics such as abilities, knowledge skills or feelings of persons individually or in the group. To Airasain (1994) a test is an assessment intended to measure a test taker’s knowledge, skills, aptitude, physical fitness or classification in many topics. While Abodurin (1999) described the test as the formal situation(s) deliberately created by a tester to make the testees respond to stimulus from which desired behaviour could be elicited. Kolo (2001) defined it as a procedure in which a standard series of questions is presented and the subject gives written or oral answer. He equally defined the test in a broader statement “A test is a systematic procedure for observing a person’s and describing it with the aid of a numerical scale or category system.

Wallace (2009) defined a test as a compact task or series of tasks designed to ascertain the merit or quality of something. Educational tests constitute a series of items for which a score is obtained. Depending on how they are constructed, they can serve different purposes while page, Thomas and Marshall (2008) explained test as follows:

(i) Any method with which the presence of quality or geniuses of anything is determined.

(ii) Examination to evaluate the performance and capabilities of a student in a class (e.g knowledge of subject).

(iii) Procedure for eliciting responses upon which appraisal of the individual can be based (e.g intelligence) and

(iv) The process of detecting the presence of an ingredient in a compound or of determining the nature of a substance.

Aggarwal (1997) defined a test as a procedure confronting a subject with a standard set of questions or tasks to which the student is to respond independently and the result of which can be treated in such a way as to provide a quantitative comparison of performance of different students, while Omotoso (1985) saw a test as an instrument which is used to diagnose or analyze a situation within the educational context. To Hassan, (1991), a test is a standard or nonstandard procedure for obtaining a systematic sample of some aspect of behaviour. On a similar note, Gesinde (1999) looked at a test “a set of task or questions/statement designed to elicit particular types of behaviours when presented under standardized conditions". The above explanations give one a notion that a test is a tool, a stimulus presented to the individual in order to elicit a response.

According to Obinne (2011) testing is a fundamental part of the teaching-learning process used not only as a basis for evaluating students at the end of the teaching-learning process but to guide teaching and aid in the development of curriculum as well as in the assessment of needs, learning difficulties, level of mastery and differences among students. Because of the diversity of test, it is been classified in several ways, tests are classified into achievement and psychological tests. Bamidele (2004) noted that an achievement test usually follows around of instruction and is aimed at determining the extent to which learning has taken place. An achievement test has a great value or significance in all types of instructional program. A classroom teacher usually depends on the achievement test for measuring the progress of his students in a subject area. Several educational and vocational decisions about students are taken on their performance in the achievement tests. It is, therefore, necessary that the teachers should be well equipped with the meaning and characteristics and uses of achievement tests (Bamidele 2004).

Gronlund (1993) defined an achievement test as "a systematic procedure for determining the amount a student has learned through instruction" while Gronlund and Linn, (1990) observed, "there typically have been norm-referenced tests that measure the pupil's level of achievement in various content and skill areas by comparing their test performance with the performance of other pupil's in some general reference group". In the word of Wise, Lukin and Roos (1991), achievement test is a measure of knowledge and skills in a content area while Gronlund and Linn (1990) identified the following characteristics of a well standardized achievement test.

(i) A good achievement test is tried out and selected on basis of its difficulty level and discrimination power.

(ii) It should have a description of measured behaviour/

(iii) It should contain a sufficient number of test items for each measured behaviour.

(iv) It should be divided into different knowledge and skills according to behaviour to be measured.

(v) Its instructions in regards to its administering and scoring are so clear that they become standardized

(vi) It is accompanied by norms which are developed at levels and various age groups.

(vii) It provides equivalent and comparable forms of the test.

(viii) It carried with it, a test manual for it administering and scoring.

Psychological tests as the name implies are expected to deal with human behaviour. A psychological test is an instrument designed to measure unobserved constructs or latent variables (Gronlund & Linn, 1990). Kolawole (2001) said that psychological tests are the same as other tests because they are standard instruments for measuring but differ from the test because the focal point of measurement is human behaviour.

In another sense, test is equally classified into essay and objective. An essay test is any written test that requires an examinee to write several paragraphs or passages. Sax,

(1989) stated that an essay test contains “questions requiring the student to respond in writing. Osunde (2009) described essay test as a free response mode of testing an individual’s academic accomplishment, proficiency or learning attainment. It is usually for testing communication skills, ability to recall and express ability to select, organized and integrates data in a general approach to problems. Essay tests emphasis recall rather than recognition of the correct alternative. Essay tests may require relatively brief responses or extended responses.

Ebel and Frisbie (1986) said that “essay test presents one or more questions or other tasks that require extended written responses from the person being testees.” Essay items allow students to select, organize, integrate, synthesize and present the answer in their own way or words, such items allow the student to use information at their disposal in solving problems presented to them in question form or even sometimes present materials in a novel way. In this way, the testee is also disposed to being original or innovative in the approaches adopted in problem situations. It is easy to construct, good for testing comprehension, application, and analysis outcome. It reduces the chance of other examination malpractices.

Essay test can be divided into an extended response and restricted response questions. Items such as short answer or essay typically require a test-taker to write a response to fulfill the requirement of the items. In administration terms, essay items take less time to construct; it can test complex learning objectives as well as processes used to answer the question. The items can also provide a more realistic and generalizable task for the test. Finally, these items make it difficult for test-takers to guess the correct answer and require test takers to demonstrate their writing skills as well as correct spelling and grammar (Ebel & Frisbie 1986).

Objective test has a clear and unambiguous scoring criterion. In a simple language, short-answer items require the examinee to response to the items with a word, short phrase, number or a symbol. The objective test can be well classified as follows: According to Wise, Lukin and Roos (1991) a short-answer item which is sometimes called supply items are considered objective items in that correct response can be secured objectively. Completion tests otherwise called fill-in the gap test provide the testee with a sentence from which a word or phrase is missing and such is to be supplied. Its items differ from the fixed response test items in that the testee does not have any option to choose from. It can be made to cover a large area of content. It is easy to construct and score. It is also applicable to any field in which achievement is measured and encourages the item writer to take sentences verbatim from textbooks. It also had a high discriminating power/value because the guessing factor is minimized in well-prepared items. Completion test items encourage memorizable of facts and it is at times vague especially when the statement is over mutilated and does not contain enough clues for a solution (Abiri 2007).

The true-false consist of a declarative statement on a situation that is either true or false. A true or false item is also known as alternative-Response Test item. Gronlund (1993) said “the alternative response test item consist of a declarative statement that the pupils are asked to mark true or false, right or wrong, correct or incorrect, yes or no, fact or opinion agree or disagree and the likes” the students decide which of the two possible choices is correct and place his answer accordingly. It is probably best known for various types of objective test items it is very popular among the classroom teachers been that it provide a simple and direct means of measuring the essential outcome of formal education. It is very easy to construct and applicable to wide range of subject matter, the objective in scoring and can widely ample knowledge tested per unit of working time.

Interpretive Exercise, according to Abiri (2007) an interpretive exercise is one in which several items are based on a common set of data which may be a reading passage, chart, map, table, picture or graph. The set of data is to be studied carefully and interpreted. A matching item requires the student to match a stem or stimulus the appropriate response or option. It consists of two columns, the premises and the responses which are listed either in cluster or side. In many clusters, the distinction between premises and responses simply given by their names. At times the phrases are the premises while shorter names are responses. Abiri (2007) is of the opinion that “A matching exercise presents the pupil with a list of premises, a list of responses and a set of directions for matching the elements of these two tests.” Arrangement items these are in which certain materials presented are required to be arranged according to a specific criterion. For instance, the material may be events to be arranged in chronological order, objects may be asked to be arranged in order of magnitude, steps in an operation to be arranged in the correct or logical sequence (Abiri, 2007).

Multiple-Choice items are common ways to measure students understanding and recall wisely constructed and utilizes multiple choice questions/test will make stronger and more accurate assessment. The shortcomings of objective tests is that it provides no measure of the student's ability to organized or to arrange his/her ideas and it offers many extraneous clues for the student who is “test wise” but who is ignorant of the subject matter of the test (Abiri 2007)

Historically Multiple-Choice (MC) items were first introduced in the army Alpha test in 1917. In 1926, the first scholastic Aptitudes Test (SAT) was administered consisting of MC Questions. The validity of examination is determined through three main concepts; (1) content sampling (2) higher level thinking and (3) recognition versus production (Messick, 1989). In regard to content sampling, in a short period of time, more test items can be administered, providing a large sample, in regard to higher-order thinking, MC testing was once stereotype as a test format measuring only lower-order thinking such as recall of facts. However MC, tests have been proven to measure higher-order thinking if constructed properly. The Multiple Choice item is generally recognized as the most widely applicable and useful type of objective test. It is a good measuring instrument for measuring complex outcomes in the knowledge, understanding and application areas. A multiple choice item consists of the problem and a list of suggested solutions. The problems are usually required stated as direct question or an incomplete statement which is called “the stem” of the test. In multiple choice items, testees are usually required to tick, shade, or write the correct response. The lists of possible answer are called “options or alternatives” The correct option to an item is regarded as the key to the items and the remaining options are called distracters (Bennett, Rock & Wang, 1990).

Many scholars have suggested three to five options to be written for each item. Olatunji (2007) recommended 3-options multiple choice tests because is as reliable and discriminating as a 4-options test. Haladyn and Dowing (1993) on the other hand supported using four or five options. Multiple-choice tests have been the staple of educational assessment in Nigeria and in administered in a small amount of time, allowing for broad coverage of content domain and higher reliability (Ferrara & Demauro 2006). Many of the strengths of MC items lie in the efficiency of administration and scoring.

Lastly, the issues of recognition versus production are still debated; some test critics believe the process students go through during MC examination is different from constructed response exams. They argued that picking the right answer is a different process than constructing the right answer. They questioned whether MC testing produces invalid results or rather less conclusive results than constructed-response exams (Fiske, 1990).

Testing proponents such as Jones, Jones and Hargrove (2003) find fault with the contention that multiple-choice test does not produce valid results and that more authentic assessment are needed. They are of the opinion that essay test are expensive to score and less reliable due to greater subjectivity in grading procedures. Essay and writing rubrics are typically used to assess writing skills in English classes are a violation of construct validity since they do not measure writing achievement but rather measure compliance to the rubrics themselves.

Iyewarun (1984) emphasized that despite the stated limitations above against objective test it is gaining rapid ground in Nigeria like other countries in the world. It is being used by examining bodies like West Africa Examination Council (WAEC), National Examination Council (NECO), National Business and Technical Examination Board (NABTEB), Joint Admission and Matriculation Board (JAMB) and federal and state government ministries of education for selections into institution of higher learning and certification in various schools or scholarships and employment.

Psychometric Properties of Achievement Test

Attention has focused on developing psychometric or properties that allow for efficient evaluation of MC items (Ferrara & Demauro 2006). These properties include validity, reliability, item difficulty, item discrimination, item distracters and guessing index. Bamidele (2004) defined validity as the extent to which an instrument measures what it purports to measure. Validity is therefore determined in relation to what particular use for which the instrument is being considered. Validity can also be referred to as the appropriateness, meaningfulness and usefulness made of a specific prediction or inference made from test score (Cronbach 1984). Validation involves checking the test score of testee against some scores and other empirical data and logical considerations. It is vital for test to be valid in order for the result to be accurately applied and interpreted. Validation examines the soundness of all the interpretations of descriptive and explanatory interpretations as well as situation bound predictions (Cronbach 1984). Hassan, (1995), Gbaleyi and Akinyemi (1995) as cited in Adewuni (2016), sees validity as the degree to which a test measures what it is supposed to measure. There are four procedures for validation of test and other educational instrument such as questionnaire, rating scale, opinion-naira etc.

Face validity is the extent to which an instrument is judged to be measuring what it purports to measure by mere looking at the test. It is facial appraisal, an instrument to ascertain its claim. Face validity can be ascertained in a way similar to that of content validity. This is done by replying on judgment of experts as to whether the items and the build-up of such instruments have facial relevance and acceptability to what it claims to be measuring (Bamidele 2004).

Content validity refers to the extent which an instrument actually measures or relates to the trait for which it is designed to measure. (Bamidele 2004). When a test has content validity, the items represent the entire range of possible items the test should cover. In the world of Hassan (1991) content validity is the extent to which a test or an instrument adequately covers the domain of behaviour it intends to measure. Airasian, (1994) sees it as the degree to which the sample items tasks or equations on a test are representative of some defined universe or "domain" of content. He stated further that the major question to ask when considering content validity is "Does these cover the entire domain to be measure?

Construct validity can be described as the extent to which the test measure the "right" psychological constructs, it is the degree to which a test measure some hypothetical trait possessed by individuals which is presumed to be reflected in the test performance (Trohim 2002). Messick (1989) described construct validation as an analysis of the meaning of test scores in terms of psychological construct. Some of these constructs which are unobservable are intelligence, anxiety, dominance, achievement, motivation; mechanical validation is its preoccupation with theory, theatrical constructs and scientific or empirical inquiry involving the testing of hypothesized relationships.

Messick (1989) made distinction validity between two types of construct validity as convergent validity and discriminate validity. Convergent validity refers to when a test or other measures of a proposed trait correlates strongly with instruments of the other kinds designed to measure trait, while discriminate validity is shown by the fact that the test correlates little or not at all with measures of other methods. Trochim (2002) is of the opinion that the convergent validity of a test of arithmetic skills can be shown by correlating the score on the test with score on other test that purport to measure basic mathematics ability, where high correlations would be evidence of discriminate validity.

Criterion related is the process of determining the extent to which test performance is related to some other valued measure to performance. It is established by comparing the test score with one or more external variable (criteria) considered to provide a direct measure of the characteristics, behaviour or attribute in question (Hassan 1991). A test is said to have criterion related validity when the test is demonstrated to be effective in predicting criterion or indicators of a construct. There are two different types of criterion related validity; predictive validity and concurrent validity (Lin, Brennan & Hartel 1997).

Prediction validity occurs when the criterion are obtained at a time after the tests. For instance, a test with predictive validity is career or aptitude tests, which are hopeful in determining who is likely to success or fail in certain subject or occupation. Concurrent validity Hassan, (1991) noted that if a new test is constructed for example its concurrent validity may be established by correlation of the examinees scores on the test with the score they recently received in a related subjects rating made by their teachers or score obtained on a similar tests that has been validated rather than waiting several years to ascertain whether a vocational interest test can predict success in a given occupation, an investigator may correlate the vocational interest test and predict successfully in a given occupation or profession. Hence, concurrent validity provides immediate evidence of the usefulness of a test (Edward 2006).

Reliability of an instrument is its measure of consistency, stability, dependability, precision and accuracy (Bamidele 2004). He further elaborated on the above that measurement makes it possible to estimate what proportion of the total variance of test scores is errors variance. Reliability means the extent to which individual differences in test score are attributed to difference in the characteristics under consideration and the extent to which they are attributed to chance errors. Thompson and Vacha-Haase (2002) explained reliability using the analogy of a bathroom scale: “some days when you step on the scale but may not be happy with the resulting score. On some of these occasions, you may decide to step off the scale and immediately step back on to obtain another estimate”. If the second score is half a pound lighter, you may irrationally feel somewhat happier… but if your second score weight measurement yield a score 25 pound lighter than feeling happy, you may instead feel puzzled or perplexed. If you then measure your weight a third time the resulting score is 40 pound heavier, you probably will question the integrity of all the scores produced by your scale. It has begun to appear that your scale is exclusive producing randomly fluctuating scores. In essence your scale measure “noting”

In view of the above, Thompson and Vacha-Haase (2002) also noted “when measurements yield scores measuring “noting” the scores are said to be unreliable”. The reliability coefficient for a set of score from a group of examinees is their co-efficient. It is possible however, for a test to yield highly consistent results from day to day without measuring what is meant to measure. In another words a test may be reliable but may not be valid. Thompson and Levitov (1985) suggested that when computing reliability estimates for a test scores to determine items usefulness to the test as a whole, the total test reliability is reported first and then each items is removed from the test and the reliability for the test less than item is calculated from this test developer deletes the indicated items so that test scores have the greatest possible reliability. Thus co-efficient of reliability are needed in order to apply the correlation for attenuation.

Item difficulty is simply the percentage of students taking the test who answered the item correctly. The larger the percentage of the testees getting an item right the easier the item and lower the percentage of the testees getting an item right the higher the difficulty index, Aggarwal (1997) said that difficulty level of a test is an index of how easy or difficulty the test is from the point of view of the testees. The index is a ratio of the average score of a sample of subjects on the test. It is usually expressed in percentage. Thus the difficulty level of a test is expresses as:

$$P = \frac{\text{Average score on the test}}{\text{Maximum possible score}}$$

The proportion for the item is usually denoted as p and is called item difficulty. An item answer correctly by 85% of the examinees would have answer correctly by 50% of the examinees would have a lower item difficulty, or p-value of .50. a p-value is basically a behavioural measure. Rather than defining difficulty in terms of some intrinsic characteristics of the items difficulty is defined in terms of the relative frequency with which those taking the test choose the correct response (Thorndike, 1951 in Keeves, 1990).

Another implication of a p-value is that difficulty is a characteristic of both the item and the sample taking the test. For example, an English test is very difficult for an elementary student will be very easy for a high school student. A p-value also provides a common measure of the difficulty of items. It is very difficult to determine whether answering a history question involves knowledge that seems more obscure, complex or specialized than that headed to answer a mathematical problem. When p-value are used to define difficulty, it is very simple to determine whether an item on a history test is more difficult than a specific items on a math test taken by the same group of students.

Item discrimination refers to the ability of an item to differentiate among students on the basic of how well they know the materials being tested. If the test and a single item measure the same thing one would expect people who do well on the test to answer that item correctly and those who do poorly to answer the item incorrectly. A good item computerized analysis provide more accurate assessment of the discrimination power of items because they take into account responses of all student rather than just high and low scoring groups (Cronbach 1984).

The method of extreme group can be applied to compute a very simple measure of the discriminating power of a test item. If a test is given to large group of people, the discriminating power of an item can be measured by comparing the number of people with high test scores who answer that item correctly. In computing the discrimination index D, the evaluator needs to first score each student test and rank order the test score, next the 27% of the student at the top and 27% is used because it has shown that value will maximize differences in normal distributions while providing enough cases for analysis. There is need to have as many students as possible in each group to promote stability, at the same time it is desirable to have the two groups as different as possible to make the discrimination clearer.

The discrimination index D is the number of people in the upper group that answered the item correctly minus the number of people in the lower group who answered the item correctly, divided by the number of people in the larger of the two groups. Allen and Yen (2001) stated that “when more in the lower group than upper group select the right answer to an item, this actually has negative validity, the item is not only useless but is actually serving to decrease the validity, of the test”.

The higher is the discriminate in favour of the upper group, which should get more items correct. A negative discrimination index is most likely to occur when an item covers complex material written in such a way that it is possible to select the correct responses without any real understanding of what is being assessed. Poor student may make guess, select those responses and come-up with the correct answer. Good students may be suspicious of a question that looks too easy, may take the harder oath to solving the problem, read too much into the question and may end up being less successful than those who guess. As a rule of thumb it terms of discrimination index 0.39 are reasonable good but possible subject to improvement, 20 to 29 are marginal items and need some revision, below 0.19 are considered poor items and need major revision or should be eliminated (Elbe & Frisbie 1991)

Two indicators of the items discrimination effectiveness are point biserial correlation and biserial correlation coefficient. The choice of correlation depends upon what kind of question we want to answer. The advantage of using discrimination co-efficient over the discrimination index (D) is that every person taking the test is used to compute the discrimination co-efficient and only 54% (27% upper + 27% lower are used to compute the discrimination index D (Matlock & Hetzel, 1997).

The point Biserial (R_{pbis}) correlation use to find out the right people are getting the items right and how much predictive power the item has and how it would to predictions. While besirial correlation (r_{bis}) is computed to determine whether the attribute or attributes measure by the criterion are also measure by the items and the extent to which the item measure them. Elbe and Frisbie, (1986) stated that rbis simple described the relationship between scores on a test e.g. 0.1 and scores 0 or 1 and score 0 or 1 on the total test for all examinees.

Analyzing the distracters is useful in determining the relative usefulness of the decoys in each item. Item should be modified if student consistently fail to select certain multiple choice alternatives. The alternatives are probably totally implausible and therefore of little use as decoys in multiple-choice item. Newogel (1992) stated that a discrimination index and discrimination co-efficient should be obtained for each option in order to determine each distracter's usefulness. The discrimination value of the correct answer should be positive; the discrimination values for the distracters should be lower and preferable negative distracters looks extremely plausible to the information reader and when recognition of the correct response depends on some extremely subtle point, it is possible that examinees will be penalized for partial knowledge. At the same time Aggarwal (1997) noted that a detail item analysis will reveal the facility value of each of the items and the discrimination index of those questions/items. It is through items analysis that a teacher or a paper setter comes to know whether the question/items has the right level of difficulty and whether there was discrimination between more able and less able students. Even though there is nothing like level of difficulty. It is usually accepted values in the range of 20% to 80% are in order for an achievement test or examination. However, the overall facility of test for the entire population may have to be around 50%. A high facility value indicates that the item is very easy for the group while a low facility value indicated a very difficulty item.

Guessing could be a strategy employed by examinees to earn marks. Guessing means giving an answer or making a judgment about something without being sure of all the facts. Guessing is a standard test-making strategy presented to examinees taking multiple choice assessments (Obinne, 2011).

Establishment of West Africa Examination Council (WAEC): The Jefferey report which was formally submitted in 1950 was occasioned by joint discussions started in 1948 among the university school examinations matriculations council, that is, the University of Cambridge Local Examination Matriculation Council, with the West Africa Department of Education. This discussion was centered on the future of school examinations in West Africa, Jeffery, who was then. Director of University of London Institute of Education was invited in 1949 by the British secretary of state to the colonies to visit West Africa, to study and advice on a proposal to establish a West Africa school Examination and Council. Between December 1949 and March 1950, Jeffery toured the Anglophone, countries of West Africa – Ghana, Nigeria Sierra Leone and Gambia. He thereafter submitted his report supporting the establishment of West Africa examination council. The report was adopted and the council was established with Ghana. Nigeria, Sierra-Leone and Gambia. He thereafter submitted his report supporting the establishment of West African Examination Council. The report was adopted and the council was established with Ghana. Nigeria, Sierra-Leone and Gambia as member countries and the headquarters in Accra Ghana. The body had its first meeting at Accra between 24th- 27th March 1953 and since then been growing in strength every year. The council was charged with the responsibility of conducting examinations and awarding certificate authorized in the United Kingdom. In 1966, the Test Development and Research Office (TEDRO) of the council were established and that marked the beginning of the use of multiple choice objectives test in public examinations in West Africa. A first school learning certificate examination was conducted for 73,340 pupils in four of the twelve states of Nigeria in December 1970, while 66,794 pupils took the Ghana middle school leaving certificate examination on August 28. 1970. Liberia becomes a full member of the organization (WAEC) in March 1974. Bathes-Weilson (1974) suggested that the council should consider providing a completely new examination for secondary school leavers who are seeking immediate employment instead of higher education. And that this examination should include aptitude test and should be an international one" this probably led in the taking over of the conduct of City and Guild Certificate Examination from the City and Guild Institute, London by WAEC.

Another development was the establishment of International Centre for Educational Evaluation (ICEE), Institute of Education, and University of Ibadan in 1973. ICEE functional as experiment from September 1972 to June 1973 and now an acute organized research unit of the Institute of Education. University of Ibadan. In 1976, the Joint Admission and Matriculation Board (JAMB) was established with the responsibilities of conducting examinations for selecting candidates for admission into undergraduate courses in all federal and state Universities in Nigeria.

Establishment of National Examination Council (NECO): Kolawale (2001) regarded the birth of the National Examination council (NECO) as the climax of an evolution process. Between 1997 and 1980 there are cases of leakages of West African School Certificate Examination papers. This has never happened before. The leakages attracted public outcry and criticisms. This result into setting up Sogbetun Commission of Inquiry to investigate the matter and give appropriate recommendations to the Government. The commission recommended for the establishment of a new parallel examination body, on the ground that West African Examination Council was too overloaded. Meanwhile NECO was established on the recommendation of Sogbetun commission of inquiry (1977), which recommend that the workload of WAEC be drastically shelved to other examination bodies to be set up. Angulu Panel of 1982 set up by the Federal Government supported Sogbetun's recommendation and recommended the establishment of three regional examination bodies for the conduct of senior school certificate examination and one Board to conduct the G.C.E type for private candidates.

Fafunwa (1991) also stated that for how long this country will continue to lament the fact of examination leakages, a situation which may be due to the handing over of the conduct of almost all the school exams to, for example West African Examination Council. In that same review of the reports of the two former commissions. Hence the report of Osiyale committee led to the Establishment of two examination bodies.

(i) National Board for Education Measurement under Decree 69 of August 1993

(ii) The National Business and Technical Examinations board (NABTEB) under Decree 70 of August 1993. The National Board for Education Council (NECO) 1999 with the responsibility of conducting the senior school certificate examination. Their first SSCE Examination was conducted in May/June 2000.

Establishment of National Business and Technical Examination Board (NABTEB): The scope of business Education encompasses Technical Education, commercial Education and Vocational Education which make the individual competent and be in a position to contribute to national development through the acquisition of business skills.

Osuala (1985) described Business Education as a programme of instruction which consists of office education for careers in office through initial, refresher and upgrading of education and the general Business Education to provide students with information and competencies needed for managing personal business and the services of the business world. According to Osuala, Business Education is that aspect of the total education programme that provides the knowledge, skills understanding and attitude needed to perform in the business world as a producer and consumer of goods and services of that business. Isu (2000) then observed that the function of education for which business education is an integral part, has been conceived to the adjustment of man to his environment to the end that enduring satisfaction may accrue to the individual and the society. The Federal Board for Vocational Education in American in 1917 saw Vocational education as an act of training that must be for the common wage earning employment. While Isu (2000) defined vocational education as a phase of education where in emphasis is laid on preparation in occupations of social value. He viewed vocational education as a very inclusive term and broadly to cover all those experiences where by an individual learned to carry on successfully any useful operation.

The international Labour Organization defined vocational education as activities which essential aim at providing the skills; knowledge and aptitude required for employment in a Particular occupation group of related occupation or a function in any field of economic activity including agricultural, industry, commerce, hotel, catering and tourist, industries, public and private services, from the definition above, it is noted that vocational education is concerned with the whole hierarchy of occupation from those requiring relatively short periods of training to that of long period of three or four years of training.

Fafunwa (1967) in his own contribution declared that if education at the elementary level is to take a new and dynamic significance, it must be aimed at training the child for some skill. Technical education and the two terms are often used interchangeable to refer to the same type of education. UNESCO, (1978) was just trying to distinguish the two by stating that vocational education require the study of technology or related sciences while technical education require the study of science subjects that qualifies someone as technologist or to enter into his chosen occupation or career. Still on scholars ideas of vocational and technical education and technical as a form of education whose primary purpose is to prepare persons for employment in recognized occupation. This definition has to do with the recognition of apprenticeship certificate, whether it is government owned or private owned. Meanwhile Ayodele (1999) described vocational education and organized education which is directly related to the preparation of individual for paid or unpaid employment or for additional preparation for a career. This definition regards vocational and technical as what is needed by all, for preparation for sale –able skills or for self sufficiency and for societal development. Kolawole (2001) also believed that certificate obtained from NABTEB should be self-employment. National Business and Technical Examination Board is the examination body in charge of assessment of technical and vocational education studies. Vocational and technical education have been in keeping with the changing employment needs of different nations as they pass through stages in their social and political development.

The origin of technical and vocational education can be traced down to Britain with the establishment of the Royal Society of Arts (RSA) in 1754 in London. The society was founded by Sir Harry Chester with a number of gentlemen interested in promoting art, industry, commerce and invention by granting rewards and premiums (Victor, 2005). Alexander, (1966) stated that faith rested on the knowledge that Britain has led the world in industrialization for a consecutive hundred years. In 1848, the Royal Society of Arts resolved that any mechanics should be entitled to join the society for the same subscription as an individual so that its members might enjoy certain of the advantages of membership. Since then technical-vocational education has started growing tremendously.

Gill and Dar (2000) stated that technical and vocational education is distinguished from general by its higher cost of delivery, especially at the secondary levels. Fluitman, (2000) stated that the effectiveness of school-based vocational education programme appear to depend on their objectives. Hence, the most common objective of vocational education is;

(a) To keep the less gifted student out of higher education and off the streets.

(b) To keep people temporarily out of labour market

(c) To provide employers with skilled workers and technicians.

(d) To provide students with general vocational skills to prepare them for lifelong learning or for post secondary specialized training.

According to Thakur and Ezenne (1980) people in Nigeria had been engaged in vocational education of some sorts from time immemorial. Towards the end of 19th century, Hope Waddell Training Institute of Church of Scotland Mission was established. In 1908, the Survey School at Lagos was established to train surveyors and the school was moved to Ibadan and later to Oyo. Higher College Yaba started as a vocational institution in 1932 and was officially opened in 1934 comprising medical school, schools of agriculture, school of forestry, Vetinary School; the survey school was bound to have demoralizing effects on the Yaba graduates. Meanwhile, University College, Ibadan came into existence in 1948, on the recommendation of the Elliot Commission, the higher college. Yaba was transferred to University College Ibadan and its students became the foundation students of the university. Since then, a variety of technical and vocational institutions at different level have come into being (Adewale 2005). The reasons behind such development have been the need to diversify education and to meet the demand from industries and government. In the western part of the country, there are secondary modern schools offering a three-year course for non-academic type of children and in addition to usual subjects, needle work, domestic science, hand crafts, rural science, music and art are also included in their curriculum. While in the Northern parts of the country, there are craft schools offering three years of courses in woodwork, and metal work technical and vocational courses are offered in institutions all over the country. At the post secondary level, technical colleges operate and offer a great variety of course in Accountancy, Commerce, and Civic, Mechanical and Electrical studies. The course led to the qualifications of professional bodies like the City and Guild of London. However, the culmination of higher stages of technical and degree courses in civic, electrical and mechanical engineering. The growing government interest in technical and vocational education is evident from the financial outlay of the third National Development plant. The plan allocated N277.326 millions for, technical education during the plan period ending in 1980.

The National Policy on Education (2013) has been candid in that it has deplored the general public attitude which regards technical education as somewhat inferior to other types of education. The policy has identified the aims of technical education as:

(i) Providing trained manpower in applied science, technology and commerce.

(ii) Providing technical knowledge and vocational skills necessary for agriculture, industrial, commerce and economic development.

(iii) Producing people who can apply scientific knowledge to the improvement and solution of environmental problems for the use and convenience of man.

(iv) Giving an introduction to professional studies in engineering and other technologies and giving training and imparting the necessary skills leading to the production of craftsmen, technicians and other skilled personnel who will be enterprising and self-reliance.

The curriculum of technical college prepares candidates for the award of the National Technical Certificate (NTC), National Business Certificate (NBC) and Advanced National/Business certificate (ANTC/ANBC). Presently there are one hundred and fifteen 115 government (both Federal and state) Technical colleges in Nigeria.

The Genesis of evaluation in the Technical and Vocational Sector revealed that in December 1853. Harry Chester, the founder of Royal Society of Arts (R.S.A), suggested the establishment of a system of examination for the benefit of members of the affiliated institutions with Sir Henry Truman Wood (Adewale, 2005). The historian of the society opined that the union, with the society’s examination paved way for the latter state organization of technical education by encouraging the establishment and growth of technical institutions. The examination conducted by the then society started badly in the spring of 1854. The syllabus was so ambitious, elaborate and comprehensive that only one candidate could be induced to enter. The council of the society (RSA) later defined the nature of its examinations in admitting public from commercial and trade schools and was intended that the tests should be confined to a certain class of persons. Shorthand was introduced as a subject in 1876 and typewriting in 1891 as a demand for the skills dictated. Thereafter, commercial certificates were being awarded to students passing three or more subjects. In 1882 the examination were thrown open to everybody and the union of institution disappeared (Adewale 2005).

In 1878, City and Guilds of London Institute removed the burden of the examination from the R.S.A and was incorporated in 1880 for the advancement of technical education under the directorship of Philip Magnus. This body took over the technological examinations of the society of Arts and their first examinations were held in 1879 with 151 successful candidates out of 184. In 1911, the City and Guide of London Institute was working closely with the Board of education, London towards the grouped course certificates and the board later endorsed the conduct of internal examining bodies on a regional scale, opened up a new phase of examining in technical education (Adewale, 2005). By and large, City and Guild examinations were transferred to WAEC on its inauguration. But as a result of the leakages of examination papers of 1977 that resulted into the inauguration of Sogbetun Commission of Inquiry. Based on the recommendation of this commission, The National Business and Technical Examination Board (NABTEB) was established to take over City and Guide examination from WAEC and take over the conduct of technical and business examination from Royal Society of Arts of London (RSA) other responsibilities include the conduct of examinations leading to the award of National Business Certificate (NBC), Advanced National Technical Certificate (ANTC) and Advanced National Business Certificate (ANBC). NABTEB was established by degree 70 of 1993 which was promulgated and signed into law on 23rd August 1993, with its headquarters in Benin City, Edo State (Adewale 2005).

Psychometric Theory of Reliability

The classical method of determining the reliability of a test is for the researcher to obtain two scores for a group of subjects on a test. These two scores may come from two separate scorings of the instrument from administration of two parts or forms of the instruments to the subjects or from two administrations of the same instrument to the subjects. The central theoretical concept that underlies the psychometric view of reliability is that every test score is composed of two parts; a true sore, which reflects the presence or extent to some trait, characteristic or behaviour plus an error score, which is random and independent of the true score. The proportion of variance accounted for by each of these parts is estimated from the correlation between the two scores obtained on the instrument.

The variance attributed to individual differences is usually given the same interpretation, regardless of how the two scores used to compute it were obtained. It reflects stable differences among individuals – the true score part of the data. The variance that is attributable to measurement error, however, is subject to varying interpretations, depending on how the two scores were obtained.

However, no measurement is perfect. For some measurement, a source of imperfection is obvious (Traub & Rowley 1991). Any particular observation has some unknown amount of error associated with that measurement for “all measurement is befuddled by error (McNemar, 1946). Test reliability is about the relative consistency of test scores and other educational and psychological measurements. One of the most important requirements for educational and psychological measurements is reliability (Chen & Fan, 1998, 1999). Reliability is an indicator of consistency, i.e. an indicator of how stable a test score or data is across applications or time. A measure should produce similar or the same results consistently if it measures the same “thing”. A measure can be reliable without being valid. A measure cannot be valid without being reliable. Reliability is the property of a set of test scores that indicates the amount of measurement error associated with the scores.

According to Ebel and Frisbie (1986), reliability is the name given to one of the properties of a set of test score – the property that describes how consistent or error-free the measurement is. We know that some tests can be fairly precise measuring tools, but we also realize that sometimes, the scores they yield are not so dependable; students can obtain scores that are either higher or lower than they really ought to be. Consequently, it is important for teachers to determine how consistent the scores from their tests are so that those scores can be used wisely to make instructional decisions about students.

Historically, the study of reliability has been linked to the study of individual differences and has been largely restricted to standardized tests of intelligence, achievement, and personality (Mitchell, 1979). A reliable instrument is one with small errors of measurement, one that shows stability, consistency and dependability of scores for individuals on the trait, characteristics or behavior being measured.

The importance of reliability pivots around the need for assurances that measurements are designed and used in ways that minimize unstable response patterns and corresponding individual and collective examinee scores. Reliable measurement is also a necessary condition for measurement of validity – although it is not the only condition. Without reliability, it is impossible to determine whether a test accurately measures student achievement. Reliability is generally described in terms of score consistency. AERA, APA & NCME (1999) define reliability as “the consistency of measurements when the testing procedure is repeated on a population of individuals or groups.” Reliability typically refers to the measurement error that is introduced into the entire measurement process, limits the degree to which generalizations can be made beyond the specific testing event, and quantifies the confidence that can be held in the value assigned to any performance. Reliability ultimately bears on the repeatability of the behaviour elicited by the test and the consistency of the resultant scores.

Reliability is related to measurement error, which almost always refers to the random component of error (Feldt & Brennan 1989). It is difficult to understand or appraise the concept of reliability without discussing the concepts of error scores, paralleled forms, reliability coefficients, and standard error of measurement.

Error Score

One of the most traditional conceptualizations is in terms of the true score, “a personal parameter that remains constant over time required to take at least several measurements” and “the limit approached by the average of observed scores as the number of these observed scores increases” (Feldt & Brennan, 1989). Unfortunately, it is impossible to know a person’s true score; it must be estimated from the observed score, which provides imperfect information. Therefore, in addition to the observed score, an error score must be theorized. A very simple concept of observed score, true score, and error score is captured in the equation;

Observed score = true score + error score - - (1)

$$X = T + E$$

Where X is the observed score, T is the true score and E is the error score. The central theoretical concept that underlies this psychometric view of reliability is that every test score is composed of two parts; a true score, which reflects the presence or extent of some trait, characteristics, or behaviour, plus an error score, which is random and independent of the true score (Nunnally 1967). Both the true score and error score are unobserved and must be estimated. The proportion of variance accounted for by each of these parts is estimated from the correlations between the two scores obtained on the instrument. The variance attributable to individual differences is usually given the same interpretation, regardless of how the two scores used to compute it were obtained. It reflects stable differences among individuals – the true score part of the data. The variance that is attributable to measurement error, however, is subject to varying interpretations, depending on how the scores were obtained.

The concept of error score is at the heart of reliability. The goal of a good measurement design is to minimize the error component. In the simple model equation, error is thought to occur randomly. The importance of random error may be recognized if an assessment is used repeatedly to measure the same individual. The observed score would not be the same on each repeated assessment. In fact, scores are more or less variable, depending on the reliability of the assessment instrument. The best estimate of an examinee's true score is the average of observed scores obtained from repeated measures. The variability around the mean is the theoretical concept of error, also called error variance. As noted earlier, measurement error can occur in the form of either systematic bias, which deals with construct validity, or random error, which deals with reliability. Random error can never be eliminated completely (Brennan 2000).

Parallel Forms

A formal concept of error is developed largely around assumptions pertaining to parallel forms. It is more effective to use parallel forms of the assessment. Parallel forms of a test are test comprising different items, but the items are designed so that they can be assumed to be randomly sampled from the same domain of comparable difficulty. The correlation $r_{X1}X_2$ of scores from any two parallel forms x_1 and x_2, are highly correlated only if the assessment is highly reliable. The concept of parallel forms lets us continue the definitions of psychometric reliability. The equation below describes $rx_{1\,,}x_2$ in terms of observed score variances V_{x1} and V_{x2} and their covariance Vx_1x2.

$$r_{x1}x_2 \quad = \quad {Vx_2 X_2}/{sd_{x1} sd_{\times 2}} \quad - \quad - \quad - \quad (2)$$

According to Feldt & Brennan (1989); Chatterji (2003), the equation above can be written in terms of true score and observed score variance

$$r_{x1}x_2 \quad = \quad {Vtrue}/{Vobserved} \quad - \quad - \quad - \quad - \quad (3)$$

Equation (3) shows that the observed correlation of two parallel forms provides information for estimating test reliability. Substituting Equation 1 in equation 3, we have

$$r_{X_1X_2} \quad = \quad {}^{V\,true}\!\Big/_{V\,true + V\,error} \qquad - \qquad - \qquad - \qquad (4)$$

This equation shows that observed score variance is composed of true score and error score variance. As error score diminishes, the ratio of true score and observed score variance approaches a value of 1. So, if the correlation of parallel forms $r_{X_1}x_2$, approaches one, then the error variance must be small. Conversely, if $r_{X_1}x_2$ is small, the error variance must be large (Brennan 2000).

Standard Error of Measurement

The reliability coefficient can be used to produce the standard error of measurement (SE_M), which sets the band of error tolerance that should be allowed in interpreting individual scores. Where reliability is high, the Se_M or band of error will be small. However, for an unreliable scale, the band of error is correspondingly wider. The formula for the standard error of measurement (SE_M) is;

$$SE_M \quad = \quad SD\sqrt{1-r}$$

Where;

SD = Standard deviation of the scale.

r = reliability coefficient of the scale

Subtracting the SEm from the score provides a lower band for the true score, adding the SE_M to the score provides an upper band. Thus a band of error is created. Under standard assumptions, this band provides a statistical confidence interval, or tolerance figure, for the score. There is a 68% chance that, if the measurement was repeated, the new score would lie within one SE_M of the original score. For greater confidence, a wider band may be used. There is a 95% probability that a second score will lie within plus or minus two SE_Ms of the original score.

Reliability coefficients are sampled dependent. They are affected by the homogeneity of the scores of the group from which they are estimated. The raw score standard error of measurement is a measure for an individual score and is therefore, consistent over different samples. When comparing the reliability of a test for different samples, SE_M is a better comparator of the reliability coefficient.

It should be of note that these measures are estimations. Theoretically, each time a test is administered, a different measure is likely to be obtained. The degree of difference depends on the reliability or error in measurement. Furthermore, reliability as a concept has several useful properties namely:

- It is a dimensionless number (i.e. it has no units)
- The maximum value of the coefficient of reliability is one, when all the variance of observed scores is attributable to true scores
- The minimum value of the coefficient is zero, when there is no true score variance and all the variance of observed scores is attributable to errors of measurement.
- In practice, any test that we may use will yield scores for which the reliability coefficient is between zero and one; the greater the reliability of the score, the closer to one the associated the reliability coefficient will be.

It is common for test users and developers to see reliability as an important property of the scores examinees attain on a test, and to see the reliability coefficient as a vital indicator of test score quality (Allen & Yen, 1979). Similarly, it would be rare for publishers of tests not to provide data on reliability in their test manuals, especially if they aspire to any degree of respectability for their tests. We should know that while accepting that reliability is an important property, we should not imagine that a high reliability coefficient alone is sufficient to demonstrate the high quality of a set of test scores. A test that yields highly reliable scores may measure abilities that are not considered important, and the test scores may be interpreted incorrectly or used for inappropriate purposes (Allan & Yan, 1979; Crocker & Algina 1986). Moreover, we should be conscious of the fact that reliability is not simply a function of the test. It is an indicator of the quality of a set of test scores; hence reliability is dependent on characteristics of the group of examinees who take the test, in addition to being dependent on characteristics of the test and the test administration.

Reliability study designs and corresponding reliability coefficients have shown that some primary sources of random error that can jeopardize the quality of tests. These sources need to be documented and monitored to provide procedural evidence that any measures of behaviour are replicable. Corresponding to many of these error sources is the type of reliability design. Conventional reliability indices such as Cronbach's alpha and Kuder-Richardson formulas, KR20 and KR21, are based generally on the concepts of observed score variance, true score and error score variance (Feldt & Brennan, 1989). In a typical test situation, four types of reliability coefficients are considered, each associated with different sources of error; - test – retest, - parallel form, - split half and – inter-rater agreement. If error is believed to be due to occasion or time, we use test-retest, if error is believed to be due to form used, parallel form is used, if error is believed to have been introduced by the specific sample of items, tasks, or behaviours, split half is more appropriate while inter-rater agreement is used if there is any reason to question the judgment or rating of performance. To estimate test-score reliability, at a minimum, one needs at least two scores on the same set of persons. The correlation between one set of scores with the other then provides a reliability coefficient.

The type of reliability estimation procedure employed is driven by the intended use of the test score. Reliability indices indicate a test's degree of consistency in assessing examinee performance. It is also an indication of the amount of measurement error present in scores generated by a test. In fact, the reliability coefficient is used to quantify a measure's reliability. It can range from 1.00, indicating perfect reliability or no measurement error, down to 0.00, indicating that the presence of random error is the only reason why students obtained scores that differed from one another. It has the same interpretative properties as the Pearson's r. For example, if a test publisher computes r and reports a reliability coefficient of 0.81, this tells us that 81% of the observed score variance is attributable to true score variance for the examinee group. However, for reliability standards; instruments where groups are concerned, 0.80 or higher is adequate – for decisions about individuals; 0.90 is the bare minimum, .95 is the desired standard (Feldt & Brennan 1989).

Factors Affecting Reliability of a Test

If we understand the factors that contribute to test score inconsistencies, and if we compute reliability estimates for the scores from our tests, we should be able to use and interpret the test scores prudently. But that is not enough; we must be able to build tests that will help us achieve score reliability estimates that are at least minimally acceptable, and we must be able to revise our tests so that “improved” versions will yield more reliable scores in the future. To this end, Frisbie (1988) identified, the test itself, the testing conditions and the group of examinees being tested as the possible factors that can affect reliability of a test.

Test and Testing Condition

Test length: Score from a longer test are suitable to be more reliable than the scores from a shorter one. This is true because the longer test is likely to yield a greater spread of scores. It is argued that a more dependable, more reproducible rank ordering of students can be achieved with a 10-item, for example than a 5-item test.

Test content: Test that measures the achievement of a somewhat homogeneous set of topics is likely to yield more reliable scores than tests that measure somewhat unrelated ideas. For example, a test that has items that measure reading comprehension, computational skills and knowledge of the principles of test construction probably will yield less reliable scores than a test of comparable length that measures only one of these traits.

Item difficulty: All the items in a test need to be in the moderate range of difficulty, neither too hard nor too easy for the group, to help identify differences in achievement among students. An item that everyone in a class answers correctly does not help to show who has achieved more or less; neither does an item that everyone misses. Consequently, in the small amount of time available for testing, the very easy or very difficult test items do little to further our purpose for testing. In fact, they take up valuable testing time and return very little information that helps us rank order individuals precisely. Item discrimination; items that discriminate properly are answered correctly by most of the students who earn high scores on the test and are missed by most of those who earn low test scores. Items that discriminate properly help to accumulate high scores for those who have learned and keep low achievers from obtaining high scores on the test (Bamidele, 2004).

Highly discriminating items help to distinguish between examinees of different achievement levels and consequently, they contribute substantially to test score reliability. In fact, the single most useful action to take in an attempt to improve the reliability of scores from a certain test is to improve each item's ability to discriminate. The test with the highest average item discrimination index is likely to yield scores of highest reliability.

This is the physical conditions under which the test is administered e.g. the time limit, security precautions among others.

Time Limits: It is normal for achievement tests to be administered with generous time limits so that nearly all, if not all, students can finish. However, when time becomes a factor, when the test can be regarded as speeded, the result is a reliability coefficient that somewhat misrepresents score accuracy.

Security Precautions*:* Occurrences of cheating by students during a test contribute random errors to the test scores. Some students are able to provide correct answers for questions to which they actually do not know the answers. Copying of answers, use of cribs or cheat sheets, and the passing of information give unfair advantage to some and cause their scores to be higher than they would be on retesting. The passing of information from class to class when the same test will be given to different classes at different times also reduces overall score reliability (Bamidele, 2004).

The Group of Examinees

Group heterogeneity*:* The reliability estimate will be higher for a group that is heterogeneous with respect to achievement of the test content than it will be if the group is homogeneous. When a group is very homogenous, it is more difficult to achieve a spread of scores and to detect the small differences that actually exist. The scores we obtain in such situations usually are so similar to one another that we are not sure if the differences are real or due strictly to random error. When inter-individual differences are greater, as a more homogeneous group, the rank ordering of individuals is likely to be replicated more easily on a retest (Bamidele 2004)

Student Motivation*:* If students are not motivated to do their best on a test, their scores are not suitable to represent their actual achievement levels very well. But when the consequences of scoring high or low are important to examinees, the scores are likely to be more accurate. Indifference, lack of motivation, or under enthusiasm, for whatever reasons, can depress test scores just as much as anxiety or overethusiasm may.

Students' Testwiseness: When the amount of test taking experience and levels of testwiseness vary considerably within a group, such backgrounds and skills may cause scores to be less reliable than they otherwise would be when all examinees in the group are experienced and sophisticated test takers or when all are relatively naïve about test taking, such homogeneity probably will not lead to much random measurement error. The rank order of score is likely to be influenced only when there is obvious variability in testwisenes within the group. Students, who answer an item correctly because of their testwiseness rather than their achievement of content, cause the item to discriminate improperly. As we have said earlier, poor item discrimination contributes to lower reliability estimates (Bamidele 2004).

Measurement Error and Need for Measurement Errors

Do measurement errors actually exist? In test, which we administer to students, there is nothing about a single test score or a pair of scores that implies the presence of measurement errors. However, we observe that if two scores are taken to be measures of the same variable for the same person, we expect them to be equal, and if they are not equal, our data are inconsistent with our conceptual framework. This dilemma can be resolved if we assume that one or both of the measurements contain error. Measurement errors play a vital role in quantitative analyses, by making it possible to model data without immediately running into inconsistencies (Kane 2008).

Kane (2008) illustrated the need for measurement error using a simple example. According to him, suppose that we have made observations of the performance of four students on some task (e.g. on a multiple-choice or performance test) and found that the four students got scores of 65, 77, 79 and 49 respectively. At this point, there is no reason to assume that these scores contain any measurement error. The scores are accepted bearing in mind that there were no mistakes made both in observing the performances and in reporting the scores. Suppose that on another day, we obtain a new observation for the four students using the same procedures, and we find that the scores are now 69, 80, 75 and 46.

Looking at the two scores, we might expect a student's level of skill to improve overtime as a function of instruction and practice.

In another context, we might expect performance to vary as a function of time, but not necessarily to follow a particular trend. If we assume that the attribute is likely to change from one observation to the next, because of fluctuations in the attitude (e.g. attitudes, moods) changes in the person's scores from day to day are likely to be interpreted as changes in the attribute of interest. However, in some cases, it may be reasonable and desirable to assume that the scores for each person should be the same on the two days – which the attribute of interest is stable across days; it means any changes in observed scores for a person from one day to the next day do pose a problem. In this case, the variability in the observed scores for a person is inconsistent with our expectations about the attribute of interest. Measurement errors therefore are introduced to eliminate this inconsistency (Kane 2008).

Based on the explanation above, two options abound. First we can simply accept the fact that each person's performance may vary across conditions of observation (occasions, tasks, context, etc) and perhaps, study how score vary as a function of different kinds of conditions of observation (e.g. how the scores change overtime). Second, we can assume that the attribute has a definite value for each person and treat the variation over conditions of observation as due to random errors of measurement. The need for measurement errors arises from the inconsistency between our assumptions that the construct of interest is invariant over conditions of observation (e.g. test forms, occasions) and observed scores, which do vary over the conditions of observation. To paraphrase Hamlet, there is more variability in our observation than is dreamt of in our theories (Kane 2008).

Measurement errors arise when we adopt a conceptual framework that presumes that the construct being measured is invariant over some conditions of observation. If we interpret our observations in terms of general attributes or constructs of persons that should not vary over certain conditions of observation, and the scores to vary over these conditions of observation, we need errors of measurement to resolve the discrepancies.

Random and Systematic Errors

Measurement errors fall into two categories namely; random and systematic errors. Random errors are errors that are easier to deal with because they cause the measurements to fluctuate around the true value. If we are trying to measure some parameter X, greater random errors cause a greater dispersion of values, but the mean of X still represents the true value for that instrument while systematic errors can be trickier to track down and is often unknown. This error is often called a bias in measurement. For example, in a chemistry class, if a teacher tells a student to read the volume of liquid in a graduated cylinder by looking at the meniscus, a student may make an error reading the volume by looking at the liquid level near the edge of the glass. Thus this student will always be off by a certain amount for every reading he makes. This is a systematic error. Instruments have both systematic and random error. Systematic errors refer to issues related to content validity while random error refers to reliability issues (John &Jeremy, 2012).

Common Sources of Measurement Error

What sorts of things create measurement error? Error can result from the way the test is designed, factors related to the individual students, the testing situation, and many other sources (Johnson, Dulany & Banks 2000). Some students may know the answers to questions posed to them, but fatigue, distractions, and nervousness affect their ability to concentrate. Students may know correct answers but accidentally mark wrong answers on an answer sheet. Students may misunderstand the instructions on a test or misinterpret a single question .Scores can also be an overestimate of true achievement. Students may make random guesses and get some questions right.

There are also test specific sources of error. For example, if the test uses reading selection as the basis for some questions. If a class happened to have previously studied the text passage being used, that class will probably do better than a class of students who have never seen the text before. For some tests, we know that changing the order of the items on the test leads to higher or lower scores. This implies that the order of the items is causing measurement error. Some test items may be biased in favour of or against particular groups of students. For example, if the reading passage contains a story that took place on a farm, students from the township like Port Harcourt may be at disadvantage in making inferences based on the story.

Thorndike (1951) pointed out that measurement error varies across the score scale despite the existence of various reliability measures. Feldt and Brennan (1989), Bachman (2004), and Webb, Shavelson & Hartel (2007) discovered an important limitation in the use of classical test theory in estimating reliability.

Specifically measurement error arises from student variables, task sampling, scoring process, test administration. These different sources affect the process at different times in the development and implementation of test scores; therefore, the need arises for the scores to be documented and monitored throughout the entire measurement process. Hence, the effect can be minimized to provide more stable and dependable estimates of students' performance.

The opportunities for measurement error are likely to expand with increased flexibility. As a consequence, assessment design and reliability estimates need to take into account the multiple factors that can bring measurement accuracy. The challenges with isolating and controlling sources of measurement error are complicated by the relationships among error sources as will be seen below:

Student Variables, Task Sampling and Scoring Process

Students come to school situations from a variety of home environments, all of which can affect their performance in school. For example, students come to school hungry, tired, or fatigued, and so forth. As they interact with classroom tasks and received feedback, students come to have expectations of success or failure, reflecting motivation, self efficacy that may interact differentially with the kinds of tasks they are given. All these co-native factors may influence the results of tests in unsystematic (i.e. random) ways (McGrew, Johnson, Cosio & Evans 2003).

Samples of performance tasks must be prepared so that they are parallel in format and difficult. That is the tasks are ideally comparable to the extent that a student would not perform differently with one another because they are both of equal difficulty. The sample of tasks is apt to be more or less variable with respect to difficulty and representation of the performance domain. Using multiple forms, individuals can be assessed overtime or compared to another. The extent of which tasks differ is obvious consequences because, with more variation, the change overtime or comparisons over multiple individuals is less trustworthy score variability that is attributable to task differences needs to be identified with carefully controlled studies in which parallel tasks and forms are used.

Irrespective of any errors made in collecting assessment data or as estimated with reliability coefficients, different or unique errors can also be made when making judgments. This type of random error refers to ratings and classifications made for students, such as pass/fail or below basic, proficient, and advanced. In this instance, the focus is less on the actual score consistency than on the consistency of judgments about states of mastery. Two types of judgments can contain error; at the score level, the focus is on partial correct responses; at the classification level, the focus is not only on the final decision to classify a student's performance but also on the standard setting process itself. The analysis therefore, needs to consider both the individual judgments made for a student as well as the overall process for making classification decisions. Although score errors need to be addressed, classification errors are far too serious, are more difficult to defect, and require more resources to resolve. Furthermore, whereas score error is usually minimized at the cut score, judgment error is most problematic at the cut score.

Test Administration and Reliability of Composite Scores

One reason for using standardized procedures in multiple systems is to minimize measurement error from external sources. Testing personnel (most often teachers), however, can introduce error (random error variance) through the way that they administer or score the test. Ironically, few states in the country have training systems for test administration. Educators assume that the conditions as noted in the test booklets are the same as those enacted in the classroom. Significant deficits are evidence in teacher knowledge concerning high-stakes testing. Most teachers' knowledge about testing and measurement comes from "trial-and-error learning in the classroom" (Wise, Lukin, & Roos 1991).

Feldt and Brennan (1989) provided the basic statistical theories about composites that are composed of linear combinations of weighted components, which can be used to study reliability of composite scores within the CTT framework. For a composite L composed of n weighted components, $\left(L = \sum_{i=1}^{n} WiXi, \right)$ where X_i is the score on component i and W_i is the assigned weight), assuming that the errors between the components are linearly independent, the composite reliability r can be assessed as (Feldt and Brennan, 1989; Thissen and Wainer, 2001; Webb, Shavelson & Hartel, 2007).

$$r = 1 \frac{\sigma^2_{C,e}}{\sigma^2_{Ce}} = 1 - \frac{\sum_{i=1}^{n} W_i^{\,2} \sigma^2_{e,Xi}}{\sum_{i=1}^{n} W_i^{\,2} \sigma^2_{Xi} + \sum_{i=1}^{n} \sum_{j(=i)=1}^{n}}$$

$$= 1 - \frac{\sum_{i=1}^{n} W_i^{\,2} (1 - r_1) \sigma^2_{Xi}}{\sum_{i=1}^{n} W_i^{\,2} \sigma^2_{Xi} + \sum_{i=1}^{n} \sum_{j(\mp i)=1}^{n} WiWjr_i, j\sigma_{x1}\sigma_{xj}}$$

$$= 1 - \frac{\sum_{i=1}^{n} W_i^{\,2} r_i \sigma_{Xi}^2 + \sum_{i=1}^{n} \sum_{j(\mp i)=1}^{n} WiWjr_i, j\sigma_{x1}\sigma_{xj}}{\sum_{i=1}^{n} W_i^{\,2} \sigma_{Xi}^2 + \sum_{i=1}^{n} \sum_{j(\mp i)=1}^{n} WiWjr_i, j\sigma_{x1}\sigma_{xj}}$$

r_i = the reliability of component i;

σ_{Xi}^1 = the variance of component i,

$\sigma_{e,Xi}^2$ = the error variance of component i;

$\sigma_{Xi,Xj}$ = the covariance between component i and component j.

r_{ij}= the correlation between component and component;

σ_c^2 = the variance of the composite scores;

$\sigma_{c,e}^2$ = the error variance of the composite score

The equation above shows that the reliability of the composite score is a function of the weights assigned to the individual components, the reliability measures and variances of the component scores and the correlations between the component scores.

Procedures involved in calculating the reliability of composite scores using the equation above include;

- Estimating the reliability of individual components;
- Calculating the variance of individual components;
- Calculating the correlation coefficient between components;
- Assigning weights to individual components to form the composites;

Using the equation above to calculate the reliability of the composite score, and, if required, manipulating the equation to optimize the reliability of the composite score to determine the weights for individuals’ components; calculating the standard error of measurement of the composite score. In conclusion, many aspects of classical test theory will continue to pertain to the former type of measurement concepts of reliability and validity, true and error scores, and parallelism of equivalent measure will continue to be applicable to these types of variables. It would be a mistake to abandon the fundamental principles and techniques of classical test theory, because they are the only principles and techniques that are applicable to variables that arise from single observations or judgments. However, even though the measurement of height comes from a single observation, one could be concerned with the reliability of such measurements by obtaining similar measures on multiple occasions and performing appropriate statistical procedures on them (Carroll 1990).

In this realm, one can foresee the future development and application of generalizability theory, concerned with the degree of which measurement errors can be identified as attributed to different sources (Brennan, 1983). For example, the reliability of teachers rating of essays can be studied as a function of teacher, students, essay, subject matter etc (Carroll, 1990).

Generalizability Theory

The classical theory of psychological tests was developed at the beginning of the 20th century, before statistical inference itself was conceived. In the second half of the century Generalizability Theory reformulated classical theory, by distinguishing between observed sample and parent population. Therefore the G theory definition of an observed test score is the sum of an unobservable true scores T and multiple error components each denoted Ei:

$X=T + E1 +E2+\ldots\ldots. +Ek.$

The observed score is thus considered as the mean score achieved by a subject (typically a student) on a random sample of test questions presented under some particular conditions of observation. The true score is defined as the mean score the subject would achieve if given the opportunity to attempt all possible questions in the population concerned. More precisely, it is the subject's expected mean score for the whole set (universe, domain) of permissible questions and conditions of observations. Measurement error, in this new theory is the result of random fluctuations due to the choice of a particular sample of questions and conditions of observation. Optimizing the sampling strategy will improve measurement precision.

Thus, G theory shares the same theoretical basis as the theory of experimental designs, i.e, and statistical inference. Yet it differs from experimental design theory in several respects. Firstly it focuses on the quantification of sources of variance, estimation of confidence intervals for means, etc, rather than on traditional significance testing. Secondly the designs that G theory is concerned with are such that each cell contains only one observation (designs without replication), because repeated measures would create a new facet (Brennan 2000).

Generalizability theory provides a framework for examining the dependability of behavioral measurements (Cronbach, Gleser, Nanda & Rajaratnam 1972). It is a statistical theory for evaluating the dependability "(reliability)" of behavioural measurements (Brennan, 2001; Shevelson & Webb, 1991). Generalizability theory consists of a conceptual and statistical framework and a methodology that enable an investigator to disentangle multiple sources of error in a measurement procedure (Gao & Brennan, 2001). Classical test theory and ANOVA can be regarded as the parents of Generalizability theory (Brennan, 2001). This is based on the fact that G-theory employs ANOVA procedures with models that are extensions of the models used in classical test theory. G theory is not a replacement for classical test theory, although it does liberalize the theory. Also not all of ANOVA is relevant to G theory; indeed some perspectives on ANOVA are inconsistent with G-theory (Brennan, 1984).

However, Generalizability theory has a conceptual framework that is not part of either classical test theory or ANOVA. A Generalizability theory analysis begins with the specification of a universe of admissible observations. A G-study is employed to estimate variance components for this universe and a relevant population. These G-study estimated variance components are used to estimate results (error variances, Generalizability coefficients/indices, etc) for one or more decision (D) studies associated with a pre-specified universe of generalization. D-study may differ in terms of sample sizes and or design structure. Specifying a universe of generalization requires identifying which facets are random and which are fixed. The most important and unique feature of Generalizability theory is its conceptual framework which focuses on certain types of studies and universes.

To estimate different sources of measurement error, Generalizability (G) theory extends earlier analysis of variance approaches to reliability and focuses heavily on variance component estimation and interpretation to isolate different sources of variance in measurement and to describe the accuracy of generalization made from observed to universe scores of individuals. G theory provides estimates of the variances contributed by each source and also provides estimates of the variance associated with the interaction between the various sources. It should be noted that the variance components contributing to measurement error are somewhat different for relative and absolute decisions; for relative decisions, variance components that influence the relative standing of individuals contribute to error while in absolute decisions, all variance components except the object of measurement contribute to measurement error (Shavelson and Webb 1991a).

Consequently, relative decisions are highly relevant since we are interested in how students perceive their scores in mathematics relative to other students' mathematics scores in public examinations. Generalizability theory provides a unified approach to understanding the dependability of measures (Brennan, 1993; Shavelson & Webb, 1991; Vanleeuwen, Barnes, & Pase, 1998; Vanleeuwen, 1997) and allows accurate assessment of the reliability of complex measures as well as measures used for either relative decisions or criterion – referenced decisions. Generalizability theory is a multifaceted extension of the classical test theory or a test theory that provides a framework for thinking about the dependability of measurements in a much broader sense through the application of certain analysis of variance (ANOVA) procedures (Kane, 1993; Vanleeuwen, Barnes & Pase 1998; Feldt & Brennan 1989; Lord & Novick 1969). Generalizability theory provides a flexible alternative to classical test theory that allows multiple sources of error to be estimated separately (Shavelson, Webb & Rowley, 1989).

Generalizability theory allows the impact of a variety of different types of sources of error, such as items, occasions, forms, or raters, on the reliability of measurements to be examined within a unified framework. The direction of Generalizability (G) theory methodology is variance components estimation; while Generalizability theory provides coefficients that are analogous to the classical test reliability coefficient, much more emphasis is placed on examining the magnitudes of the error from the different sources (Brennan, 2003). The Generalizability theory literature emphasize that reliability is a characteristics of the data, not a given test or instrument (Eason 1989; Thompson 1991, 1992). In order to evaluate the dependability of behavioural measurements, a Generalizability (G) study is designed to isolate and estimate as many facets of measurement error as is reasonably and economical feasible. A G-study makes an explicit separation of empirical information into facets of observation and objects or targets of measurement respectively. Classical test theory postulates that an observed score can be decomposed into a true score and a single undifferentiated random error term. By contrast, Generalizability theory liberalizes classical theory by employing ANOVA methods that allow an investigator to disentangle the multiple sources of error that contribute to the undifferentiated error in classical test theory (Brennan, 2003).

According to Brennan (2003), there is some truth to the assertion that Generalizability theory is the application of ANOVA to measurement problems, but this assertion is perhaps more misleading than informative. To him, it is misleading in that on a superficial level, Generalizability theory pays no attention to hypothesis testing; rather, Generalizability theory focuses on the use and estimation of variance components. An individual observation or measurement is merely a sample estimate of an individual's true score and are part of a universe of admissible observations. Forms, items, occasions and raters called facets which can include any characteristics of a measured procedure that is a potential source of error. The levels within each facet (e.g. the different items or different occasions of measurement are conditions that can be infinitely large (Webb, Rowley & Shavelson, 1988). The object of measurement is usually persons which is typically not considered a source of error, because people vary and their true score differences are real, systematic and of great interest to investigators (Eason 1991; Kieffer 1999). The objects of measurement do not create error variance and therefore, not considered a facet. Anything that generates systematic variance can be the object of measurement. Other possibilities include schools, businesses, work groups or occasions. For each person (which, for simplicity, will be the objects of measurement in this study), the mean of the score from the various conditions and facets provide the best estimates of the person's true, or universe score. This grand mean is always a flawed estimate of the universe score, because of the errors contributed by the measurement facet. G-theory, decomposes, estimates, and reveals these measurement errors which are termed variance components. A G-coefficient is produced for each data set, representing the universe score variance divided by the observed score variance. G coefficient range between zero and one. When the G-coefficient for a data set is high, investigator can generalize the obtained scores across the study facets-hence the term Generalizability theory.

G-theory as a test theory has been used in many studies in different areas of academic endeavour. Many areas in research have explored the use of Generalizability (G) theory to better explain sources of error that may occur during assessments, such areas as marketing research (Huges & Garret 1990), clinical assessment in athletic therapy (Ragan & Kang, 2005), and performance assessment in foreign language (Kozaki 2004). However, a study carried out by Akeju (1972) on the reliability of General Certificate of Education Examination, English composition papers in West Africa by W A E C showed that such factors as poor matching of questions, poor reader agreement and inadequate examiners were the major factors that affected the reliability of the examination. Despite the fact that this study dealt with multiple factors, Generalizability theory analysis was not employed to estimate the multiple sources of error in the study. The Generalizability theory analysis would have been employed to estimate the contributions of each sources of error identified to measurement error. Therefore, the need arises for the use of G-theory in studies especially those involving multiple factors so as to make full explanation of errors that occur in assessments.

The Universe of Generalization: This is the conditions of a facet to which a decision maker wants to generalize. In short, the universe of generalization is defined as the set of facets and their levels (e.g., items and occasions) to which a decision maker wants to generalize. A person's universe score (denoted as μp) is defined as the long-run average or more technically, expected value of his or her observed scores over all observations in universe of generalization

Decision (D) Study: The decision (D) study deals with the practical application of measurement procedure. A decision (D) study uses variance components information from a Generalizability Study to design a measurement procedure that minimizes error for a particular purpose. A decision (D) study estimates examine universe scores (true scores) along with various reliability and dependability indices. “Generalizability analyses are useful not only for understanding the relative importance of various sources of error but also for designing efficient procedures” (Brennan, 2001). In the decision (D) study, decisions are based on the mean over multiple observations (e.g. test items) rather than on a single observation (a single test item).

The most important D study consideration is the specification of a universe of generalization. Together, a G-study and a D-study aid measurement developers and users with very specific information about measurement error and optimal measurement design (Shavelson & Webb 1991a). A relative decision concerns the relative ordering of individuals (e.g. norm-referenced interpretation of test scores). The variance of errors for relative decision is:

$$\sigma_{\delta}^{2} \quad = EpE_{1} \;\; \delta_{p1}^{2} = \sigma^{2}{}_{p1,\;e} = \frac{\sigma^{2} pie}{n^{i}{}_{i}}$$

Similarly, an absolute decision focuses on the absolute level of an individual’s performance independent of others’ performance (c.f. criterion-or domain-referenced). The variance of errors for absolute decisions is;

$$\sigma_{\Delta}^{2} \ = \ EpE_{1}\,\Delta_{p1}^{2} = \sigma_{1}^{2} + \sigma^{2}{}_{p1,\ e} = \frac{\sigma_{i}^{2}}{n^{i}{}_{i}} + \frac{\sigma^{2}_{pie}}{n^{i}{}_{i}}$$

Note that with absolute decisions, the main effect of items – how difficult an item is does influence absolute performance and so is included in the definition of measurement error (Webb, Shavelson & Hartel, 2007).

Facets: Facets are those variables that potentially influence our observed measurements. To get the best estimates of true score variance and error variance,we need to identify as many of the facets that are at play in our measurement application as we can, and to classify these as contributors to one or other type of variance. The universe of admissible observations typically is discussed in terms of measurement facets. “A facet is simply a set of similar conditions of measurement” (Brennan, 2001), e.g., items, occasions, raters. It should be noted that the term is not applied to population (persons or students), who serve as the primary objects of measurement. Basically, various facets are identified for data collection using a carefully designed study. Then, ANOVA procedure is used to estimate how much variance from separate facets as well as variance from the interaction of facets. In generalizability (G) studies, the focus is on understanding (estimating) the amount of variance associated with universe of admissible observations (operationalized through facet). After the Generalizability (G) study is completed, the variance components have been estimated, and then a decision study is conducted.

Decision studies “emphasizes the estimation, use and interpretation of variance components for decision-making with well-specified measurement procedures” (Brennan, 2001). Basically, the focus is on making generalization (over replication) based on the result of the G-study. Facets can be fixed or random, crossed or nested.

Random and Fixed Facets: A facet is random if its conditions can be exchanged with another of the condition from the same facet (Kieffer, 1999). Similarly, a facet can have an infinite number of conditions in the universe or a finite number of conditions in the universe but not all conditions are included in a measurement design. In this case, the facet is random, one the conditions in a particular measurement design are a sample of all possible conditions. On the other hand, a measurement can exhaust all possible conditions of a facet in a G study and therefore the facet is a fixed one. There is no variance component for a fixed facet in a G study statistically; G-study treats a fixed facet by averaging over the conditions of a facet. Shavelson and Webb (1991) points out that 'if it does not make conceptual sense to average over the condition of a fixed facet, or if conclusions about such average are of little interest, separate G-studies should be conducted within each condition of a fixed facet".

Crossed and Nested Facet: When all conditions of one facet are observed with all conditions of other source of variation, the design is a crossed design. In our simple scenario, where each individual responds to all the items in the achievement test, the design is a crossed one and can be denoted by P x i. In a generalizability (G) study design, it is also possible that one facet is nested within another. Nesting happens when two appear with one and the same condition of another facet (Shavelson & Webb, 1991). For example, items in a test may be nested within the subtests facet when each subtest has two or more distinct items. The notational form of this design is P x (i: t), where t represents the facet subtests.

Variance Component: If we consider a two-facet crossed person x item x occasion, G study design where items and occasions have been randomly selected, an observed score for a particular person on a particular item and occasion is decomposed into an effect for the grand mean, plus effects for the person, the item, the occasion, each two way interaction and a residual (three way interaction plus unsystematic error). The distribution of each component or 'effect', except for the grand mean, has a mean of zero and a variance σ^2 called the Variance Component.

In practice, the parameter values and expected values are unknown, therefore, G theory typically uses analysis of variance (ANOVA) procedures to compute terms that can be used to obtain estimates of variance components. Unlike ANOVA, G theory is typically not concerned with tests of statistical significance but employs ANOVA sums of squares and mean squares to obtain estimates of variance component.

The variance component for the person effect is called the universe-score variance. The variance components for the other effects are considered error variation. Each variance component can be estimated from a traditional analysis of variance (ANOVA) (Shavelson, Webb & Rowley 1989). The magnitude of the variance components tells us how much each facet contributes to measurement error. To estimate variance components, a G study has to be conducted in which data are collected on a representative sample of persons or items as the case may be.

Error Variances

Absolute Error Variance: Absolute error variance, $\sigma^2(\Delta)$, is the error involved in using an examinee's observed mean score as an estimate of his or her universe score. It is simply the difference between a person's observed score and universe scores. For person P, absolute error variance is defined as:

$\Delta_P = X_{PTR} - \mu_P$ where

Δ_P = absolute

X_{PTR} = a person's observed score over tasks and raters.

μ_P = universe score

Absolute error variance is often associated with domain – referenced (or criterion-referenced) interpretations of scores. From the formula above, the variance of absolute errors, $\sigma^2(\Delta)$, is the sum of all the variance components except the variance of P.

Thus:

$$\sigma^2(\Delta) = \sigma^2(1) + \sigma^2(P1)$$

$$= \sigma^2(i)/n^1_i + \sigma^2(pi)/n^1_{i.}$$

$\sigma^2(\Delta)$ = absolute error variance

$\sigma^2(i)$ = variance component for item

$\sigma^2(Pi)$ = variance component for persons and item

Relative Error Variance: Relative error variance $\sigma^2(\delta)$, is defined as the difference between a person's observed deviation score and his or her universe deviation score, defined as; $\delta_P = (Xpi - \mu_i) - (\mu_P - \mu)$ where

$\delta_P =$ relative error

Xpi =observed score

μ_i = expected score

μ = universe score

Relative error variance is therefore;

$\sigma^2(\delta) = \sigma^2(Pi) + \sigma^2(i)$

Relative error variance is similar to error variance in CTT. In general, relative error variance is less than absolute error variance because it includes fewer variance components. This suggests that relative interpretations about person's score are less prone to error than absolute interpretations (Brennan, 2001a).

Generalizability Coefficient / Indices

Generalizability Coefficient: Cronbach, Gleser, Nanda and Rajaratnam (1972) defined reliability – like coefficient called a generalizability coefficient, which is denoted E_P^2. A generalizability coefficient can be viewed as the ratio of universe score to expected observed score variance. The expected score variance includes both the universe score variance and the relative error variance. The formula for Generalizability Coefficient is appropriate when making a relative decision.

In an achievement test with items as the only facets, the relative error is;

$$\sigma^2_{rel} = \frac{\sigma^2_{pie}}{ni}$$

where n_i is the number of items in the measurement. This is true because $\sigma^2_{pi,e}$ is the error variance for a single item. The amount of error for an instrument is inversely proportional to its number of items. The formula for calculating Generalizability Coefficient is;

$$\Sigma P^2 \quad = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_{rel}}$$

The G-coefficient shows how accurate the generalization is from a person's observed score, based on a sample of a person's behaviour, to his or her universe score (Shavelson & Webb, 1991a). It reflects the proportion of variability in individuals' scores that is systematic and attributable to universe-score.

Index of Dependability: If an absolute decision is to be made, index of dependability is the proper coefficient to be used. Index of dependability is the ratio of universe score variance to the sum of universe score variance and absolute error variance.

$$\phi = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_{abs}}$$

In calculatingϕ, not only the residual variance (interaction and unidentified error) but also the items variance contributes to the absolute error

In the former scenario of achievement test,

$$\sigma^2_{abs} \qquad \frac{\sigma^2_i + \sigma^2_{pi,e}}{ni}$$

The difference between relative and absolute decisions is reflected in how the relative and absolute decisions are determined. Relative error only involves interactions that include variances except for the universe score variance. Both Generalizability coefficient and index of dependability involves number of items n, therefore, we can determine how many items are needed in a measurement in order to reach a particular Generalizability (ΣP^2) coefficient and index (ϕ) of dependability.

Designs in Gcneralizability Theory Studies

Generalizability study (G study) is designed specifically to isolate and estimate as many facets of measurement error as reasonably and economically possible (Shavelson & Webb, 2005). Designs in GT could be classified as: Univariate and Multivariate; Balanced and Unbalanced; Crossed and Fixed; and Random and Nested Designs. Univariate and Multivariatc Gcneralizability Designs: Brennan (2003), stated that univariate GT has only one universe score for the object of measurement while multiple universe scores could be used for the object of measurement in a multivariate GT. Univariate GT uses particular set of scores to describe individual's personality, skills, aptitudes or performance but multivariate GT uses multiple scores to describe an individual's personality, skills, aptitudes, or performance. Also, multivariate GT decomposes observed variances and covariances into components (Cuiyun, Yuanhang. Qiang & Jianyi, (2010). The use of multiple scores in multivariate GT makes it appropriate for use when a test has a number of sections or subsets such that there would be need to investigate composite score reliability (Gebril, 2013). Brennan (2000b) indicated that in a multivariate design, an examinee would have multiple universe scores such that each has one condition of a fixed facet.

According to Webb et al (2007), there are three purposes for which multivariate GT could be conducted.

a. estimation of the reliability of different scores, observable correlations, universe scores and error correlations for different decision studies with different sample sizes;

b. estimation of the reliability of a profile of scores using multiple regression of universe scores on the observed scores in the profile; and

c. Production of a composite of scores with maximum GT.

They used a one facet crossed multivariate design involving teacher behaviour which was divided into behaviour in Reading and behaviour in Mathematics while the same raters rated the teachers in Reading and Mathematics. GT was developed as a random effect theory and that pan of its limitation is the need for fixed effect. Keller, Clauser and Swanson (2010), maintained that multivariate GT helped to overcome this limitation by accurately modeling the particular levels of fixed facet as separate dependent variables in the design. For instance, a proper modeling of a test constructed from a table of specifications would need a multivariate design such that different levels included in the table of specifications are the separate dependent variables. They explained that multivariate design is theoretically more appropriate for use when assessing reliability of a test that is constructed based on table of specifications.

Furthermore, Brennan (20I0c), stated that multivariate GT has multiple universes of generalization such that each examinee has multiple universe scores. He explained that statistically, multivariate GT involves both variance components and covariance components. If a particular test say essay has narrative and descriptive types of essays, then each essay type would be designated as YI and V2 respectively. If the design is a fully crossed as p x t x r (persons crossed with times and crossed with raters) then there would be seven variance components for each of the essay types. Conclusively, the basic difference between univariate GT and multivariate GT is the presence of covariance and use of multiple scores in multivariate generalizability design.

Balanced and Unbalanced Designs: A balanced GT design requires that each facet involved in the design should be equal such that the same number of observations would be employed. Also, there would be no missing value in a complete balanced design. On the other hand, unbalanced design occurs when the facets in the design have different levels and there are missing values in the raw scores (Briesch et al, 2014).

Crossed and Nested Designs: Shavelson and Webb (1991 a), stated that a design is crossed if all raters rate all the students at all occasions (px r x o) or all raters rate all the students using all the forms of lest and at all the occasions (p x r x f x o). Bolus et al (2006), stated that when all the subjects (persons) were evaluated by each rater the same number of times then raters are crossed with subjects. It is a type of design where all conditions of one facet (items) are observed with all condition of another source of variation (persons), such that all persons responding to all forms of items, at all occasions and are rated by all raters (Brennan, 2010c). The crossed design is symbolised by **x" such that p x r (persons crossed with raters), p x r x o (persons crossed with raters and crossed with occasions), or p x r x f x o (persons crossed with raters crossed with forms and crossed with occasions) represent different crossed designs (Brennan, 2010c).

Nested design: is another division of GT design whereby some conditions of one

facet (e.g. items) are observed with some conditions of another source of variation (persons). One facet is said to be nested in another facet when two or more conditions of the nested facet appear within one and only one condition of another facet (Shavelson & Webb, I991b). It is symbolized by ":" For instance, i;p (items nested in persons), i:p:o (items nested in persons and nested in occasions), or p:r:i :o (persons nested within the raters, nested within the items and nested in occasions), represent different nested design. Shavelson and Webb (199la), stated that nested design is a GT design whereby different raters rate different students, or one group of persons take Paper A and another group of persons take Paper B or one group of markers mark Questions 1 and 2 and another group of markers mark Questions 3 and 4.

Brennan (2010a), observed that a design is said to be nested if each person (examinee) is administered a different sample of the same number of items (or is rated by different raters) with all items sampled from the same population (universe). There are two conditions for nested designs as noted by Shavelson and Webb (1991a), and they are: (a) multiple levels of A are associated with each level of B; and (b) different levels of A are associated with each level of B. These two conditions would mean that, if raters are nested within the students for instance, there would be different raters of different levels to rate each student at each occasion. They explained further that there are two reasons for having nested design in Generalizability Theory.

Random and Fixed Designs: Shavelson and Webb (199la), stressed that a design is regarded as random if the sample is less than the universe (population) and that the researcher is willing to exchange the sample with another sample from the universe. This means that the sample is not special to the user as it does not possess any characteristics that other samples from the universe may not have. A random design, according to Shavelson and Webb (2005), is created through random sampling levels of a facet. They observed that a design could still be called random even if the levels of a facet have not been selected randomly from the universe of admissible observations but the intended universe of generalization is infinitely large, then the concept of exchangeability or interchangeability may be employed to consider the facet as random. Webb ct al (2007), noted that a design could be regarded as random even if conditions of the facet was not randomly sampled but there is exchangeability of unobserved conditions with observed conditions.

A facet is random if its conditions can be exchanged with any other conditions from the same universe. So, it is a design that allows conditions in the sample to be exchanged with another set of same-size conditions from the universe. For instance, if 20 items in an Economics test that has 60 items can be exchanged with another set of 20 items then the items are random. Random design is necessary and appropriate when the researcher intends to generalize the outcome of the research to the universe and beyond. GT is essentially a random measurement theory therefore, it is good to have only designs in which at least one facet of error is random (Brennan, 20I0a).

A fixed design on the other hand, is a design in which the sample is equal to the universe (population) and that the issue of exchange or interchange of sample does not arise (Shavelson & Webb, I991a). The major shortcoming of fixed design is that it cannot be generalized. This is because the sample is equal to the universe to which generalization would be made. To Shavelson and Webb (2005), a fixed facet in GT is analogue to a fixed factor in ANOVA. This is true because a fixed factor in ANOVA exhausts all levels in the universe to which a decision is made just as the conditions of a fixed facet in a GT exhausts all possible conditions of interest in the universe. Facets are considered fixed when their levels are not exchangeable and are of specific interest to the decision makers and that generalization is restricted to the conditions of the research. Fixed design occurs when conditions of the facet exhausts the conditions in the universe to which researchers want to generalize. For example, an Economics test with subsections covering all aspects of Economics does not give room for the issue of exchange of sample as the sample in itself is same as the population. Fixed design is used when the researcher is not interested in the' generalization of the outcome of the research beyond the sample.

Shavelson and Webb (1991 a), identified two reasons for having a fixed design in GT which include: intentionally selection of certain conditions from the universe for the study by the decision maker and that he is not interested in generalizing beyond them or is not reasonable to do so, and secondly, if the entire universe of conditions is small such that all conditions are included in the measurement design. In their own explanation of fixed facet, Webb, Shavelson and Haertel (2007), corroborated Shavelson and Webb (1991 a), by stating that a design is regarded as fixed if the decision maker: intentionally selects certain conditions and he is not interested in generalizing beyond the conditions selected; sees it is as undesirable to generalize beyond the conditions that have been observed; or noted that the entire universe of conditions is small and the design include all the conditions.

Classification of Generalizability Theory based on Number of Facets in the Design: GT could also be classified based on number of facets involved in the design. In this categorization, GT is divided into three namely: One-Facet; Two-Facet; and Multiple-Facet Designs.

One-Facet Design: A one-facet design is described by one source of measurement error, that is, by a single facet (Shavelson & Webb. 1991 a). This is to say that one facet design occurs when the decision maker is dealing with only one condition of measurement or a universe defined by a single facet. Brennan (2010a), stated that a single-facet design occurs when the universe of admissible observation and the universe of generalization involve conditions from only one facet.

One facet design has four types of designs and is grouped into crossed and nested designs. Each of the group has two designs! Crossed design is divided into one-facet crossed random and one-facet crossed fixed while nested design has one-facet nested random and one-facet nested fixed design.

One-Facet Crossed Random Design: Shavelson and Webb (1991a) explained that a design would be regarded as one facet, if it has one source of measurement error and it is crossed if all conditions of one facet (item) are observed with all condition of another source of variation (person). It is random if decision maker intends to generalize from one set of test items to a much larger set of test items. The item universe is defined by all admissible items or if Shavelson and Webb intends to generalize from one test type to a much larger set of test types, the universe will be defined by all admissible test types (Wan et al, 2014; Webb et al 2007). So, the presence of one facet of measurement (one facet) coupled with its observation by all the conditions of the facet, with all conditions of variation (crossed) and its generalization to a larger population (random) make a design to be regarded a one-facet crossed random.

Lei, Smith and Suen (2007), used this design as occasions crossed with observers (o x r). In their study, occasions as object of measurement were ten while the facet of measurement were two observers. The observers observed all the occasions to make it a crossed design. In another work, Cuiyunetal (2010), used persons crossed with raters (p x r) as one facet crossed random design where sixteen entrepreneurs served as persons while six entrepreneurial evaluation experts served as raters. Each expert rated all the sixteen entrepreneurs using a like type (I-IO points) evaluation method. Other designs are denoted by p x i (persons crossed with items), p x f (persons crossed with forms), or i x r (items crossed with raters).

One-Facet Crossed Fixed Design: A one-facet crossed fixed design is differentiated from a one-facet crossed random by the absence of generalization (Shavclson & Webb, 199Ib). The design is fixed if the conditions of one facet exhaust the conditions in the universe to which researchers want to generalize. That is, the sample is equivalent to the population and that there is absence of generalization. For instance, a one-facet design that has an English Language test with subsections covering all aspects of English Language, which is crossed with examinees (persons) cannot be generalized beyond the sample as it has covered all aspect of the item (English Language). It is denoted by p x i (persons crossed with items), p x r (persons crossed with raters), p x f (persons crossed with forms), i x r (items crossed with raters), i x o (items crossed with occasions), etc. There is no difference in the denotation representation of one-facet crossed random and one-facet random fixed but the difference is the presence of generalization in random and absence of generalization in fixed design. Shavelson and Webb (1991 a) argued that GT is not meant for fixed design due to the absence of generalization except if the design is a combination of crossed and fixed.

As regard the sources of variability of a one-facet crossed design, Shavelson and Webb (I991b), and Webb, Shavelson and Haertel (2007), stated that there are four sources of variability (either random or fixed). These sources of variability are:

a. The differences among students' achievement in a test. (The difference may be as a result of their knowledge, skills, etc. This is usually called object of measurement. This is the person's ability denoted by $o^2{}_p$);

b. The differences in the difficulty of test items (some items would be easy while some may be difficult. This is denoted by $o^2{}_i$);

c. The differences in the educational and experimental histories that students bring to the test (this is the interaction between person and item (p x i), which is denoted by $o^2{}_{pi}$);and

d. The random error or unidentified effects which may affect the observation (it is denoted by $o^2{}_e$);

Shavelson and Webb (199l a), stated that number c and d are combined to form the residual as they are not easily accounted for separately and is denoted by $o^2{}_{pi,e}$ The sources of error (variability) in one facet crossed (random or fixed) design according to Li and Lautenschlager (1997) and Lin and Zhang (2014) are three because the variance for residual represents both interaction between the person and item and random error. This also supports the view of Brennan (2010a) that the interaction between persons, items and error be combined.

One-Facet Nested Random Design: One-facet design is said to be nested if some conditions of one facet appears within one and only condition of another facet (Shavelson & Webb, 199la). For instance, one group of persons taking Test Form 1 and another group of persons taking Test Form 2 or one group of markers mark questions 1 and 2 and another group of markers mark questions 3 and 4. They maintained that a one-facet design would be regarded as nested only if multiple levels of A are associated with each level of B and different levels of A are associated with each level of B. If an achievement test with several subsets has different sets of items (i) is associated with each subset (s) then items are said to be nested within subset and this is represented by i:s or i(s) which means facet ^{4}i" is nested within facet^{k}s' (Brennan, 20IOa), The randomization in the design is the willingness of the decision maker to generalize the result and that any sample chosen is not unique which could be exchanged with any other sample within the universe (population).

One-Facet Nested Fixed Design: Absence of willingness of the decision maker to generalize the result obtained beyond the sample makes the design to be fixed. Therefore, a one-facet nested fixed design is a type of design that involves one object of measurement (person) and one facet of observation such that some conditions of the facet of observation are observed with some conditions of the object of measurement and that the decision maker is not interested in generalizing the result beyond the sample (Brennan, 2010a). The presence of one facet of observation, different levels of conditions of measurement with object of measurement and absence of generalization makes a design to be regarded as a one-facet fixed nested design. It is symbolised by p;i (persons nested within items), i:r (items nested within raters), i:o (items nested within occasions), etc. The symbols for nested designs are the same whether random or fixed. The main difference between one-facet nested random and one-facet nested fixed is the presence of willingness to generalize in the former and absence of it in the latter.

There are two sources of variability in a one-facet nested (random or fixed) design according Shavelson and Webb (199la), and which are:

a. The differences among students' achievement in a test (the difference could be as a result of their knowledge, skills and behaviour. It is usually regarded as object of measurement. This is the person's ability denoted by o^{2}_{i}); and

b. The differences in the difficulty of items coupled with the nesting interaction between items and persons, and error (this is represented by $o^{2}_{i,p,e}$).

Shavelson and Webb (199lb), and Brennan (2010a), corroborated Shavelson and Webb (1991a) that one facet nested (random or fixed) design has two sources of variability the person effect $\sigma^2_p$and residual $\sigma^2_{i,p,e}$.

Two Facets Designs: Due to the complexity of the measurement in social sciences, it is usually contain more than one facet (Shavelson & Webb, 1991a), A GT design is said to be two-facet, if it contains one object of measurement (person) and two facets of observations (raters and occasions). Its universe of admissible observations is defined by two facets (items and occasions) (Shavelson & Webb, I991b). There are two groups under this design. These are crossed and nested designs. Crossed design is divided into two-facet crossed random and two-facet crossed fixed while nested design is sub-divided into two-facet nested random (fully), two-facet nested random (partial) and two-facet nested fixed design.

Two-Facet Crossed Random Design: Shavelson and Webb (1991a), described a two-facet crossed random design as a type of GT design whereby the object of measurement (person) is observed by all conditions of two facet of observations (raters and occasions, items and forms or raters and items). For instance, if the decision maker intends to generalize from set of the items and test types to a much larger set of test items and test types such that the universe of admissible observations would be defined by the two facets (items and test types) taken together then it is two-facet crossed random. In this design, it could be that all the raters rate all the students at all occasions, all raters rate all the students on all the types of test or all examinees partake in all the test types at all the occasions, It is regarded as crossed since the conditions of the two-facet of observations are observed by the object of measurement while its randomization is due to the willingness of the decision maker to generalize beyond the sample to the universe (population) and that he is ready to exchange the sample with any of the same sample size from the universe because the chosen sample is not important to him (Shavelson & Webb, 1991 b). It is symbolized by p x r x o (persons crossed with raters and crossed with occasions), p x r x f (persons crossed with raters and crossed with forms), p x I x r (persons crossed with items and crossed with raters), p x i x f (persons crossed with items and crossed with forms), p x i x o (persons crossed with items and crossed with occasions), etc.

Watkins, Lee and Erich (1980) stated that a completely crossed, two-facet random model design was used as they apply GT to the Matching Familiar Figure Test (MFF) to analyse the dependability of the MFF as a measure of reflection-impulsivity in which four grade levels: second, third, fourth and fifth were used. Burns (1998), while identifying sources of variation in the dependability of a modified form of the Habitual Physical Activity Index (HPAI), which is a commonly used self-report physical activity questionnaire, made use of items and occasions as the major potential sources of error and this he claimed is a two-facet crossed design because the HPAI consists of 8 item indices (one for work and one for leisure) which was administered to 45 persons on two occasions, 2 weeks apart. He also established that since he accepted the principle of exchangeability the design is random. So, this is also a two-facet crossed random design. Also, Shavelson and Webb (2005), used two-facet crossed of person x item *x* occasion design where items and occasions have been randomly selected in their study.

Furthermore, a two-facet crossed random design was employed by Bolus, Bridgeman and Bailey (2006), in their study: introduction of generalizability theory in second language it-search, as they used person x rater x occasion design. In evaluating quality of journal writing in Mathematics in Singapore, twenty-nine junior college students wrote journal insks on the given topics and two raters marked the task using a scoring aibric and this according Nie et al (2007), is a two-facet crossed random design where students were crossed with task and raters (s x t x r). Schunemann et al (2007), made use of two-facet crossed random design where 91 patients with chronic obstructive pulmonary diseases (COPD) were rated by three different clinical markers states (CMS - mild, moderate, find severe diseases) twice several weeks apart. Here, the design is persons x items x occasions. This design (a two facet crossed random) is the most popular design of all the GT designs has it is mostly used by researchers.

Two-Facet Crossed Fixed Design: When the design of a Generalizability theory has one object of measurement (e.g. persons), which is observed by two-facet of observations (e.g. raters and occasions) such that the decision maker is not willing to generalize beyond the sample (since the sample is equal to the universe) then it is said to be two-facet crossed fixed (Shavelson & Webb, 1991a). The major difference between a two-facet crossed random design and a two-facet crossed fixed is the presence of randomization in the former and its absence in the latter, which could be as a result of the small size of the population or that it is not reasonable to generalize beyond the sample (Shavelson & Webb, 1991 b). The symbols used for two-facet crossed random are the same for a two-facet crossed fixed.

There are seven sources of variation in a two-facet crossed (random or fixed) design as observed by Nie et al (2007): and Iramaneerat et al (2007). This is also corroborated by Keller et al (2010), as they observed seven sources of variability in a two-facet crossed design of p x c x r where "p" is person, "c" is case and "r" rater. It should be noted here that different researchers used this design (two-facet crossed) by combining different facets such as p x i x o (persons crossed items and crossed with occasions) (Shavelson & Webb, 2005); p x t x r (persons crossed with tasks and crossed with raters) (Brennan, 201 Ob); p x r x i (persons crossed with raters and crossed with items) (LoPilato, Carter & Wang, 2014); p x r x o (persons crossed with raters and crossed with occasions) (Briesch et al, 2014); m x r x g (models crossed with raters and crossed with grade-levels) (Lin & Zhang, 2014).

Two-Facet (Fully) Nested Random p;r:o Design: A design in which different raters (r) rate different students (p) at different occasions (o) is regarded as a two-facet nested design (Shavelson *&* Webb, 199la). It is a fully nested design because the two facets of observations (raters and occasions) are nested within one and another and also with the object of measurement (persons). This design is regarded as nested because each person is rated by different raters and different raters rate different students at different occasions as opposed to crossed design which requires the same set of students to be rated by the same raters at each occasion. So, occasions are nested within the raters as raters are nested within the persons and occasions are also nested within persons. The design is also said to be random in as much the decision maker is willing to generalize beyond the sample and can exchange the sample with any other same sample size from the universe. It is symbolised by p:i;o (persons nested within items and nested within occasions), p:i:r (persons nested within items and nested within raters), p:r:o (persons nested within raters and nested within occasions), p;i:f (persons nested within items and nested within forms), etc.There are three sources of variability in a two-facet (fully) nested random design as observed by Brennan (2010a); and Shavelson and Webb (I991a).

Two-Facet (Partial) Nested Random Design: A nested design is said to be partial if one of the facets is crossed and the other is nested (Shavelson & Webb, 199 la). It could be raters nested within occasions but crossed with persons p *x* (r:o), occasions nested with persons but crossed with raters (o:p) x r, occasions nested within persons and raters but persons are crossed with raters o:(p x r), or occasions crossed with raters but nested within persons (o x r):p (Brennan, 2010a). These show that there are four different designs under partial nested designs. The sources of variability of each of the four designs differ.

a. Raters Nested within Occasions but Crossed with Persons p *x* (r:o): In this design, different raters are present at different occasions to examine the same set of students at each occasion. So the students are the same while the raters differ as well as occasions. There are five sources of variations in this design according to Shavelson and Webb (199 la). Clauseretal (2014), used items crossed with judge and panel while judge is nested within panel. In this design (i x (J:P))> there is no object of measurement as it is applicable in the educational measurement application. This according to them is because their research work was on the estimation of the replicability of the standard setting results in the medical field. However, Hagtvet and Hanin (2007), used this design p x (ire) for best and worst performance of athletes where item indicators (i) are nested (:) within global emotions (e), but crossed with athletes (p). Athletes on the other hand are crossed with global emotions. They identified five sources of variability.

Shavelson and Webb (1991 a), used persons (p) as object of measurement while Hagtvet and Hanin (2007), used emotion (e) as object of measurement. So, there is no permanent source of variability that should be object of measurement. It is the design that determines which of the sources of variability would be object of measurement. In other word, Follesdal and Hagtvet (2009), used this design as p(i:s) where "p" represents person, "i" for item and "s" for stimuli. In the design (p(i:s)), persons are crossed with items and stimuli while items are nested within stimuli. They corroborated Shavelson and Webb (199la), and Hagtvet and Hanin (2007) that there are five sources of variability in a two-facet partial nested design of p x (r:o), px (i:e)and p(i:s).

b. Occasions Nested within Persons but Crossed with Raters (o:p) x r: Shavelson and Webb (1991a), explained that the design involves different persons at different occasions who are examined by the same set of raters at each occasion. This means that all the raters are present at every occasion and rate all the students. Brennan (20I0a), observed that there are five sources of variability in this design, which are the universe score due to students' different ability, difference in the rating of the raters, interaction between the raters and persons, the nesting of occasions within the persons; and residual which combines the nesting of persons, raters and occasions.

Furthermore, Coates and Thoresen (1978), in their study titled "Using generalizability theory in behavioral observation" used subjects, times and observers for this same design. In the study observers were crossed with times and subjects while times were nested within subjects. That is (t:s) x o design. Observers are same with raters, as times equal occasions while subjects are same with persons. They maintained that five sources of variance could be estimated because the subjects and times interaction has been confounded as a result of nesting of time within subject. Occasions Nested within Persons and Raters but Persons are Crossed with Raters o:(p x r); In this design, each person is rated by all the raters but at different occasions. Persons are crossed with raters but occasions are nested within the combination of persons and raters. So, the students are rated by all the raters but not all the students are present at all the occasions. Coatcs and Thoresen (1978), used two designs in the same study. In the second design, (the first design was (o:p) x r) observers were crossed with subject but nested within the times (o x s):t.

There are four sources of variations in this design as observed by Shavelson and Webb (1991a). Occasions Crossed with Raters but Nested within Persons (o x r):p: Shavelson and Webb (199la), noted that in this design all raters are present at all occasions but each person is rated by different raters at different occasions. This means that all the raters are present at every occasion but rate different students. According to them, there are four sources of variability in the design. These are persons effect nesting of raters within persons and rater effect, nesting of occasions within persons and occasions effect, and residual which combines interaction effect between occasions and raters, interaction effect among persons, occasions and raters and error.

Two-Facet Nested Fixed Designs: All the two-facet nested random designs would be regarded as fixed whenever the decision maker intends not to generalize beyond the sample which may be as a result of the small sample size or that it is not reasonable to generalize beyond the sample (Shavelson & Webb, 199la). The design and sources of variations are the same for all the five types of two-facet random designs that have been discussed. So, the fixed designs would also have two-facet (fully) nested fixed (p:r:o); raters nested within occasions but crossed with persons p x (r:o); occasions nested within persons but crossed with raters (o:p) x r; occasions crossed with raters but nested within persons o:(p x r); and occasions crossed with raters but nested within persons (o x r):p.

Multi-Facet Designs: GT designs can go beyond one-facet and two-facet designs. There could be three-facet design, four-facet design, and so on, depending on the decision maker's design. Also, the design could be crossed, nested, random or fixed depending on the presence or absence of generalization (random and fixed) and cost or logistic considerations (for crossed or nested) (Shavelson & Webb, 199la). A three-facet crossed random design was employed by Sudweeks et al (2005), in their study: A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing using nine raters to rate each of the 48 essays twice, written by 24 students at each occasion. The design was students by tasks by raters by occasions (p x t x r x o). In this design, there are fifteen sources of variability. Also, students by raters by items by code (language or dialect) (s x r x i x c) random design, which is a three-facet crossed random design was employed by Solano-Florcs and Li (2006) in their study: The use of generalizability theory in the testing of linguistic minorities. They also corroborated Sudweeks et a) (2005), that there are fifteen sources of variability in the design which ranges from variance for person to variance for residual.

However, Follesda) and Hagtvet (2009). used item nested within stimuli and task but crossed with person i.e. p x (i;s;t) design which *is a* three-facet (partial) nested random design. Also, Taylor and Pastor (2013), used tasks nested within students, while student is crossed with rater and attempts in a three-facet partial nested design (t:s) x r x a. In their work titled "Reliability of observers' subjective impressions of families: A generalizability theory approach", Stora, Hatgvct and E4eyerdahl (2013), employed a three-facet partial nested design of mother nested in father and crossed with raters and items i.e. (m:f) x r x i. There are eleven sources of variability in this design which ranges from variance for persons to variance for residual.

A Four-Facet Crossed Random Design was used by Hagtvet and Hanin (2007), as athletes were crossed with items, games and occasions (e x a x i x g x o). In this study, emotions were the objects of measurement while athletes, items indicators, games and occasions were facets of observations.

Classical Test Theory (CTT) and Generalizability (G) Theory

Statistical theories of measurement scores have evolved over the last 100 years. Spearman (1904) provided the basic theoretical foundation for the true score model. Many developments occurred in the next several decades. Gulliksen (1950) summarized all developments into a single coherent system. This system of statistical theorems and equations has become known as the classical test theory of measurement. Cronbach, Rajaratnam & Gleser (1963) and Cronbach, Gleser, Nanda, & Rajaratnam (1972) established the Generalizability (G) theory. Classical test theory and Generalizability theory are more robust against the violation of their model assumptions. Technically, classical test theory and Generalizability theory are somehow similar to each other except for little dissimilarity which will be discussed later. The score of interest in both classical test and Generalizability theory is the observed score from the test. The primary goal of both classical test theory and Generalizability theory is to evaluate the quality of the observed test score by estimating reliability coefficients and standard errors. Again the unit of analysis for both classical test theory and Generalizability theory is the overall test (Brennan, 2001).

Vanleeuwen (1997) identified advantages of Generalizability theory over classical test theory. Generalizability theory considers multiple sources of error simultaneously and allows more accurate modeling of the measurement situation than methods modeling only a single course of error. Classical test theory considers only single sources of measurement error for relative decisions.

- Generalizability theory provides a unified approach to viewing various types of error. The same methodology can be applied, whether the source of error is items, occasions, forms or raters. Thus Generalizability theory can consider any of the number of different sources of error either by combination with one another or by themselves.

- Generalizability theory provides a unified approach for assessing the reliability of measurements taken for either relative decisions (norm-referenced measures) or absolute (criterion referenced measures). Relative decisions are based on an individual's ranking within a group rather than on an absolute score. Absolute decisions, on the other hand are based on an absolute score with no comparative reference to the scores of others (Ary, Jacobs & Razavich, 1996).

- Generalizability theory simultaneously estimates various sources of error including interclass (Thompson, 1992, 1991). Classical test theory assumes that sources of error do not consider the possibility of interaction as to create additional measurement.

- Classical test theory assumes facet effects are zero. For example, if items are the source of error, classical test theory assumes that all items are equally difficult. These assumptions are relaxed under Generalizability theory. The removal of these assumptions allows Generalizability theory to consider both relative and absolute decisions.

Those who think that classical test theory and Generalizability theory are put under one umbrella of "true score model" are partly correct in that classical reliability can be considered a special case of Generalizability theory reliability in which most facets are fixed. However, fixed facets may not be realistic for many measurement applications. Hence, classical reliability estimates tend to overestimate true reliability. Thus, only the

Generalizability theory is sufficiently comprehensive to accommodate hypothetical, as well as practical measurement situations. Those who consider Generalizability theory as a clarification of the classical test theory with an additional component of formative evaluation of measurement errors are also partially correct. The Generalizability theory conceptualization of reliability indeed forces one to consider different sources of error simultaneously and estimate the effect of each source. In addition, one can plan the measurement procedure more effectively according to the relative magnitude of variance due to different sources. As an example, if item variance constitutes the largest source of error whereas rater variance constitutes the least, one can allocate the resource more efficiently by increasing the number of items without increasing the number of raters to achieve a desired level of reliability (Brennan, 2001a).

Classical test theory (CTT) is most commonly used as a framework for the detection of rater variation and estimating reliability in performance assessment situations. However, Generalizability (G) theory is a more powerful approach than CTT for the detection of rater variation and estimating reliability (Shavelson, Baxter &Gao, 1993). G. theory extends the framework of CTT in order to take into account the multiple sources of variability that can have an effect on test scores. While CTT provides a single estimate of error, G-theory can be used to identify not only multiple sources error but also the impact of these sources of error on the overall accuracy (Shavelson & Webb, 1991b).

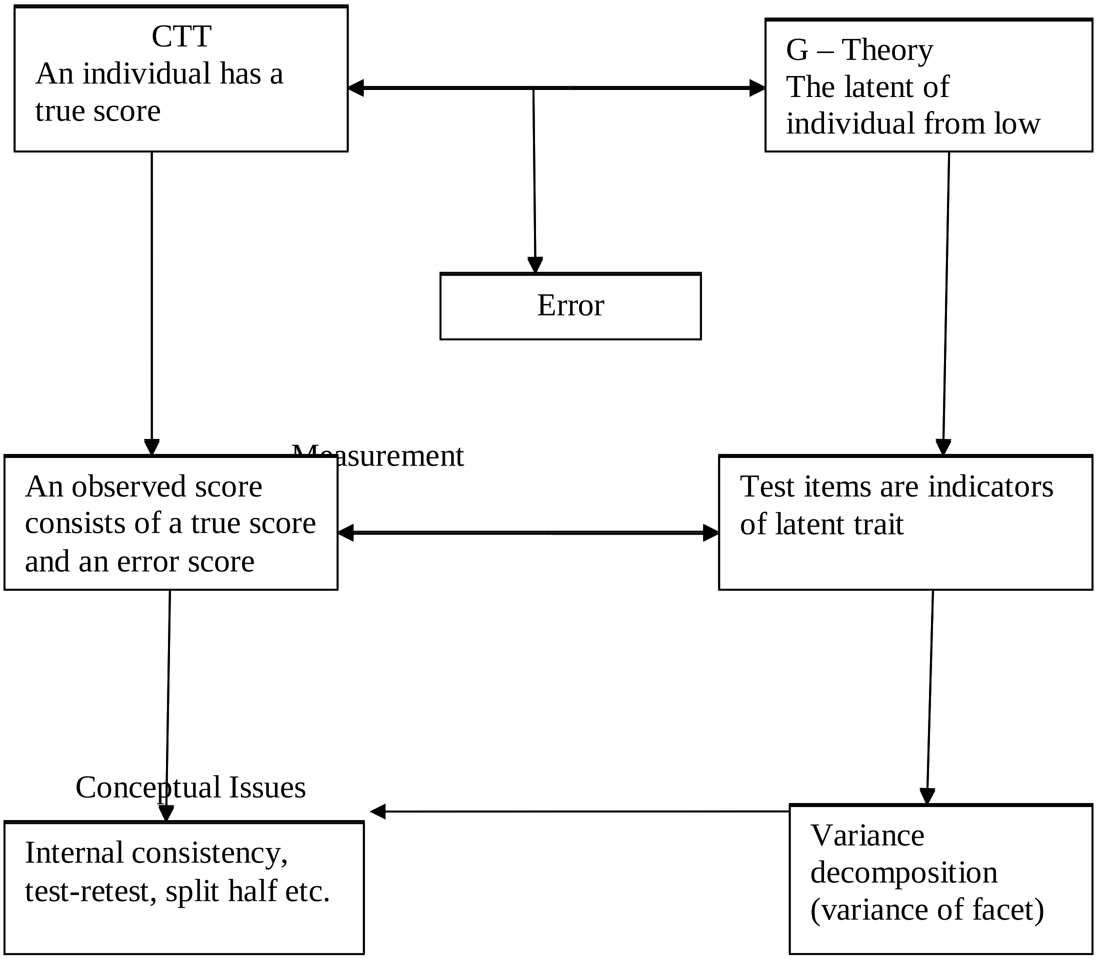

CTT
An individual has a true score
G – Theory
The latent of individual from low
Error
Measurement
An observed score consists of a true score and an error score
Test items are indicators of latent trait
Conceptual Issues
Internal consistency, test-retest, split half etc.
Variance decomposition (variance of facet)

Figure 2: Comparison of Classical Test Theory and Generalizability Theory.

Source: Wang (2005, 50); Estimating reliability under a Generalizability theory model for writing scores in C –Base.

However, beyond the above points, classical test theory and Generalizability theory differ both theoretically and practically in some very important ways. First, the underlying statistical assumptions are different. Classical parallel test assumes that true score, variance, error variance and correlation with an external criterion are the same across test forms. These assumptions are often untenable. Generalizability theory assumes that tests are randomly parallel. That is, test content is assumed to be random sample from a defined domain or universe. This random sampling assumption is more realistic and is commonly made in statistical analyses. Another theoretical difference is that error in classical test theory is a random term of undefined source. As such, classical test theory cannot accommodate the idea of reliability due to specific sources such as stability, internal consistency and inter-rater consistency. These concepts, however, are part of the conceptualization of measurement within Generalizability theory (Brennan, 2000).

In practical terms, an implicit condition for classical test theory is that all potential sources of measurement error are fixed except the one for which a sample is drawn. For example, when a number of items are used, raters and occasions as well as any other potential sources of error are considered fixed with no generalization beyond the exact rater, exact occasion, and so on, actually used in the data collection procedure. This is unrealistic and has serious limitation in terms of meaningfulness. In that respect, classical test theory would not be appropriate for authentic or performance assessment. In contrast, the Generalizability theory can accommodate any assessment situations, restricted only by the practical limitations of data collection and software. Finally, in criterion – referenced testing, both systematic measurement errors and random measurement errors needs to be considered. Classical test theory cannot accommodate systematic measurement error and is, thus, appropriate only for norm-referenced testing. The Generalizability theory, however, is appropriate for either norm-referenced or criterion-referenced testing because, it has the flexibility to accommodate both relative and absolute measurement errors (Brennan, 2000).

Concept of Multivariate Generalizability Theory

To address the challenge coming from CTT, Cronbach, Gleser, Nanda and Rajaratnam, (1972) first introduced Generalizbility Theory (GT) as a statistical theory for evaluating the reliability of measurements. Later, Shavelson and Webb, (1991) further developed GT and made the theory more understandable with their published book—Generalizability Theory: A Primer (1991), the theory reached its climax and was accepted by most researchers nowadays after Brennan produced his book—Generalizability Theory (Brennan 2001a).

Comparing with CTT, GT is more flexible and powerful. In particular, instead of decomposing an observed score as a true score and an error score, GT considers both systematic and unsystematic sources of error variations and disentangles them simultaneously, so the observed score can be decomposed into as many possible effects as specified by the measurement design. For example, in a writing test where raters and prompts must be considered, an examinee's score can be decomposed into a grand mean in the population and universe, and seven other effects, due to person, rater, prompt, person-rater interaction, person-prompt interaction, rater-prompt interaction, and person-rater prompt interaction. By examining all possible sources of error, a researcher can easily identify where large error sources come from and make appropriate decisions to decrease the error variance. As an extension of classical test theory, GT shares some concepts and assumptions with classical test theory. For instance, the *universe score* in GT has the same implication as the *true score* in classical test theory, errors are assumed to be uncorrelated and independent of true scores, samples selected and used to estimate the error variances are randomly selected from the population (Brennan 2001a).

Many reliability studies utilizing Generalizability Theory (GT) is called Univariate Generalizability Theory (UGT), because only one *universe score* is associated with the object of measurement. For example, in math achievement test, each examinee only has one math score, that is, only one universe math score is associated with each person. Increasingly, however, datasets in the form of multiple subtests are more common to test developers and users. For example, in SAT test, each examinee has two scores representing verbal and math abilities, and a total score for the whole test. Of course, we could analyze the SAT data using UGT where the universe score of each examinee is regressed on his or her total score. However, we lose information about two specific subtests. In addition, if different set of items is used to measure the ability and number of items is not equal, then, we have unbalance data problem, which leads to the difficulty of variance components estimation. Even worse, some measurement may not have a composite score, that is, they are only profiles for sub-scores.

All these facilitated the development of Multivariate Generalizability Theory (MGT), where two or more universe scores are associated with the object of measurement and covariance components in addition to variance components are taken into account. In sum, under certain circumstance, both UGT and MGT can analyze the same given data and the results from univariate and multivariate analysis can be similar, however, multivariate analysis provides more information that can be used by the test developers and users (Shavelson & Webb, 1991a; Brennan 2001a).

Educational and psychological measurements often involve multiple scores describing individuals' aptitudes or skills. To assess the reliability of such measures, the vast majority of studies take a univariate approach. The most common procedure is to determine the reliability of each score separately. Another method, sometimes used in Generalizability studies which take into account multiple sources of error variation, is to determine the Generalizability of a particular composite of the scores. These univariate methods do not, however, allow the investigator to assess sources of error covariation among the multiple scores. Such information is important for designing an optimal D study, and permitting the decision maker to determine the composite with maximum Generalizability. For these purposes, a multivariate analysis would be more appropriate. The Generalizability procedures outlined by Cronbach, Glesser, Nanda, and Rajaratnam (1972) and Joe and Woodward (1976) are applied to General Educational Development (GED) ratings of jobs in the U.S. The results of univariate and multivariate Generalizability analyses of the GED ratings are presented (Shavelson & Webb, 1991; Brennan, 2001a).

A major contribution of Generalizability theory, then, is that it allows the researcher to pinpoint the sources of measurement error (i.e., rater, occasion, or both) and increase the appropriate number of observations accordingly. In extending the notion of multifaceted error variance to multivariate designs, Cronbach (1972) focus on methods of obtaining and interpreting variance components. Analogous to univariate G theory, multivariate G theory decomposes the observed score variance-covariance matrix into matrices of components of variance and covariance for universe scores and sources of error variance. To obtain the matrices of components of variance and covariance, the expected mean-square and crossproduct equations are solved in a fashion analogous to their univariate counterparts. Although Cronbach et al. discuss components of variance and covariance at great length in their illustrations and interpretations of multivariate Generalizability analysis; they do not develop a multivariate Generalizability coefficient. Multivariate analogues of reliability are developed by Bock (1966, 1963; see also Haggard, 1958) and Conger and Lipchitz (1973; Conger, 1974) for designs that do not differentiate sources of error variation, but the only multivariate reliability coefficient anchored in Generalizability theory is developed by Joe and Woodward, (1976). Their approach distinguishes between G and D studies and can be generalized to a variety of multifaceted designs with crossed and nested facets. The value of their approach is that it allows the investigator to maximize a Generalizability coefficient by assessing the magnitude of different sources of error and so design D studies that reduce the large sources of error variation and covariation. Joe and Woodward, (1976) multivariate coefficient is a direct extension of Cronbach, (1972) univariate coefficient. From a random effects multivariate analysis of variance, the canonical variates are determined to maximize the ratio of universe-score variation to universe-score plus error variation.

For the last few decades, test developers and users have attempted to investigate the reliability of a measurement where responses of multiple subtests (or profiles) are obtained for each object of measurement. Such data have the following characteristics: (1) each examinee (object of measurement) has two or more universe scores representing subtests or profiles; (2).The conditions of subtests (or profiles) are fixed, that is, the selected conditions are our interest and will not be generalized to other conditions. (3) The number of items in each condition of the subtests (or profiles) is not the same, which means the data are unbalanced. Furthermore, the researchers concern about not only each universe score but also the composite (or profiles) of universe scores for the whole test. Multivariate Generalizability theory (MGT), in contrast with univariate Generalizability theory (UGT), was developed to meet the challenge (Rajaratnam 1965; Shavelson & Webb, 1991a; Brennan 2001a).

MGT is not complete without comparing with UGT. The difference between MGT and UGT can be described like this: MGT involves with two or more universe scores for the object of measurement at the same time, while UGT involves with only one universe score for the object of measurement at a time. In this sense, multivariate analysis of a specific dataset can be constructed based on multiple univariate analyses in a row. More importantly, multivariate analysis account for not only variance components like univariate analysis does, but also covariance components between the universe scores that univariate analysis cannot do. This powerful function of multivariate analysis allows us to investigate and design reliable observations both at each universe score level and composite score or profiles level.

However, each multivariate design can have a counterpart in a univariate design, which means, logically, that any data can be performed with univariate GT. The choice of multivariate GT over univariate GT depends on the complexity of the data and what kind of information that we want to derive. Brennan, (2001) recommended performing a full multivariate analysis if there is a fixed facet in the research design. In his book Generalizability Theory, Brennan discussed the problems of analyzing unbalanced data, where the sample sizes in each condition of a facet are not equal. Unbalanced data creates complexity when we want to decompose the variance components. One way to reduce the complexity of unbalanced data in univariate analysis is to analyze the data under the framework of multivariate design if possible (Shavelson & Webb, 1991a; Brennan, 2001a)..

Take the SSQ data as an example: it is reasonable that different sample sizes of items are distributed to four temperament scales in SSQ test with more items in one scale and fewer items in the other scale. Multivariate design avoids the problem of unbalanced data by analyzing four parallel univariate designs. In the end, each univariate design has balanced data under four levels of fixed facet temperament. In addition, Haertel, (2006) pointed out two disadvantages of using univariate analysis for data containing fixed facet. First, variance components are forced to be the same for observed scores for the levels of the fixed facet. Second, universe score represents an equally weighted composite of scores on the levels of the fixed facet, which is not always true. Consequently, some information about the scores cannot be derived from univariate analysis. In sum, the advantages of multivariate GT over univariate GT are: a) multivariate GT reduces the complexities and ambiguities in terms of unequal numbers of items within fixed facets if univariate analysis is used; b) estimations of variance and covariance components can be alienated in a multivariate analysis, but not in univariate analysis (Brennan, 2001a); (c) estimate observable correlations, or universe-score and error correlations for various D study designs (Brennan, 2001a); (d) estimate the reliability of profiles of scores using multiple regression of universe scores on the observed scores in the profile (Brennan, 2001a, Cronbach et al 1972); or (e) produce a composite of scores with maximum generalizability (Shavelson & Webb 1981).

Empirical Studies on Multivariate Generalizability Theory

Albert (1984) cited in Tunde (2015) applied multivariate GT to the assessment of the student achievement in art education. Twenty-five art students rated paintings of 60 fourth-grade students with regard to three criteria. The results indicated that Generalizability coefficient was low with respect to different raters and moderate with respect to different topics. Webb, Schlackman & Sugrue (2000) carried out a study on the dependability and interchangeability of assessment methods in science. Specifically, the study investigated the importance of occasion as a hidden source of error variance in the estimates of the dependability of science assessment scores and the interchangeability of science test formats. Six hundred and sixty two students were involved in the study. Univariate Generalizability analysis were conducted to examine the dependability of the assessment methods and the consistency of performance scores across tasks raters and occasion, while multivariate Generalizability analyses were conducted to examine the universe score correlations among testing methods used for measurement error. The findings of the study showed that ignoring occasion as a source of variation can seriously over estimate the dependability of achievement test scores, whether hands-on-performance test or paper-and-pencil test hence will lead to misleading conclusion regarding other sources of error in the measurement .The findings of the univariate Generalizability analyses confirm those found by McBee and Barnes (1998) and Shavelson, Riuz-primo & Wiley (1999). In the previous studies, when occasion was not considered as an explicit facet, task sampling was the major source of variation. Both the hands-on and paper-and pencil tests showed that the person and task interaction effect was quite large indicating that the relative standing of examinees was not consistent from one task to the other. On the other hand, differences between raters were very small. However, the importance of task sampling was reduced when occasion was included in the design. Also adding occasion as a source of variance in the multivariate Generalizability analysis influenced the interpretation of the observed correlation between hands-on and paper-and-pencil scores. Finally, estimates of the correlations from the multivariate Generalizability analyses were high and similar for both designs-.85 for persons x tasks x rater using the test from occasion 1 and .89 for persons x tasks x rater x occasion design using test from occasion 2.

Card, Myford, Dowing and Rachel (2007) investigated the quality control of an Objective Structure Critical Examination (OSCE) using Generalizability theory and many-faceted Rasch measurement for evaluating competencies. Communication skills were examined using OSCE with 79 residents from a Midwestern University in the United States. Each resident performed six examination tasks with Standardized Patients (SPs), who rated the performance of each resident using 5 category rating scale items. The ratings were analyzed with Generalizability and Many-Faceted Rasch Measurement (MFRM). It was revealed that the largest source of error variance besides the residual error variance was SPs/Cases. The MFRM. Study identified specific SPs/Cases and items that introduced measurement errors and suggested in their levels of severity/difficulty from the study two Sps gave inconsistent rating that suggested problems related to the ways, they proofread the case, their understanding of the rating scale, and/or the case content. It was also revealed that Sps/Cases interpreted two of the items inconsistently and the rating scales for two items did not function as 5-categories scales.

Yonyan, Shu and Shun (2007) examined the use of Generalizability Theory to evaluate the quality of an alternative assessment (journal writing) in mathematics, twenty-nine junior college student wrote journal task on the given topics and two raters marked the tasks using a scoring rubric, constituting a two-facet G-study design in which students were crossed with task and raters. The results showed that increasing the number of tasks had a lager effect on the G coefficient and index of dependability, than increasing the number of raters.

Dongmei and Brennan (2007) conducted a series of Generalizability analyses of a reading comprehension test for both groups, this study demonstrated the amount of discrepancy in coefficient and error variances when different facts are taken into account and the differential contribution of these facets to measurement error for ELLs and native English speakers. Youzhen (2007) applied the method of multivariate GT to assess the reliability of the student style questionnaire (SSQ). In particular, random effect variance and covariance components were estimated. The results indicated that the G coefficient were acceptable for the total scale and two of the subscales.

Webb, Shavelson and Hartel (2007) carried out a study on a Generalizability study of job performance measurements of Navy machinists' mates. The study involved 26 machinists' mates, 2 tasks, and 11 observers. The findings of the study showed that the estimated variance components from the G-study suggest that a single task would probably provide dependable ratings but that multiple observers are needed to represent job requirement. Averaging over 11 observers with a single task yields moderate estimated Generalizability coefficients and dependability index using two tasks has no appreciable effect on the results. A very large number of observers, of questionable feasibility, would be needed to obtain a reasonable level of Generalizabilty coefficient.

Huang (2008) in a study on “how accurate are ESL students’ holistic writing scores on large scale assessments”? A Generalizability theory approach examined both variability and reliability of ESL students’ writing in the provincial English examination in one province in Canada. The purpose of the study was to examine both the rating variability and reliability of a large-scale ESL students’ writing in the provincial English examinations in Canada. It was intended to find out if there was any difference between the rating variability and reliability of the writing scores assignment to ESL students and the NE students for the provincial English examination in that province across a 3 year period.

The existing data from the writing components of the 2002, 2003 and 2004 administrations of the provincial English examination were used for the G-theory analyses. According to the study, by using data for three consecutive years, it was possible to replicate the analyses and then check the stability of the results stressing that a variety of large-scale language performance assessment have used this strategy (Bachman, Lynch & Mason, 1995; Lee, Kantor and Mollaun 2002). The provincial English examination is administered five times a year in November, January, April June and August. In the examination, the students were asked to complete three separate writing tasks (paragraph format writing task of poetry, essay format, writing task for literary prose, and original composition) and each writing task received scores from two different raters.

The findings of the study showed that differences in score variation did exist between ESL and NE students when adjudicated scores were used. There was a large effect for both language group and person within language-by-task interaction. Previous research has indicated that there was little consistency among diverse tasks (Lee & Mollaun, 2002). The findings further showed that ESL and NE students had unequal performance across tasks. The result further showed that, the 95% confidence interval on the residual and person variance components did not overlap hence they were significantly different between ESL and NE students while the 95% C I on the remaining variance components overlapped, indicating that these variance components were not significantly different between ESL and NE students.

Knut (2008) studied the emotional intelligence, the Mayer, Salovey and Caruso emotional intelligence test (MSCEIT) from the perspective of Generalizability theory that the researcher used multi-facet measurement design. The results from Generalizability analyses of scores from 111 Norwegain executives responses measurement error were revealed. Generalizability coefficient for scores from perceiving emotions, facilitating, thought, understanding emotion and managing emotion were estimated to 0.71, 0.37 0.50 and 0.46 respective, which is substantially lower. The low estimated Generalizability coefficient suggests that the scores may not generalize well to intended domains and the validity of some of the scores may be questioned as opined by the researcher.

Keller, Clanser and Swanson, (2010) carried out study on using Generalizability theory to assess the effect of content stratification on the reliability of a performance assessment. This study explores the effect of representing and misrepresenting the multivariate and univariate Generalizability studies were reported. Results indicated that the proper specification on the analytic design is essential in order to yield proper information both on Generalizability of the assessment and the standard error of measurement. Lombardi, Seburn, Conley and Snow (2010) carried out a study on a Generalizability investigation of cognitive demand and Rigor Ratings of items and standards in an Alignment study. The study was to examine the Generalizability of ratings used to compute various alignment indices in the context of a broader alignment study between college admission and placement test items and a set of college readiness standards. The reliability of the cognitive demand and rigor ratings were investigated by conducting a Generalizability theory analysis with items crossed by raters (i x r) and standards crossed by raters (s x r) designs. Six English and six mathematics content area experts were recruited to participate in the alignment study. The findings of the study indicated that the six raters did reach an acceptable level of dependability for estimating mathematics and English standards' level of cognitive demand but for level of rigor. It shows that for both items and standards', six raters appear to be sufficient for cognitive demand, but insufficient for rigor. The findings of the study indicated stronger Generalizability across raters for mathematics items and standard ratings than for English items and standard ratings. Additionally, the results indicate stronger Generalizability across raters for cognitive demand ratings than for rigor ratings. The findings of this result also confirm the findings of Herman, Webb and Zuniga (2005). They reported that the ratings of cognitive demand were more reliable than were ratings of centrality (similar to rigor). The findings also show that there are greater differences in residual effects in cognitive demand than rigor, and there is greater difference in standards than items. These findings suggest larger interaction effects for cognitive demand and standards. (i.e rater rank-ordered items and standard differently on rigor and cognitive demands). The findings of the study also indicated that more raters are necessary to obtain sufficient reliability in rigor than cognitive demand.

Lisser, Brian and David (2010) assessed using multivariate GT to assess the effect of content stratification on the reliability of a performance assessment. The study further investigated the effect of representing and misrepresenting the stratification appropriately in estimation of reliability and standard error of measurement. The results indicated that proper specification of the analytic design is essential in yielding the proper information both about the Generalizability of the assessment and the standard error of measurement. John and Jeremy (2012) assessed a frame work for conceptualizing measurement error when using authentic assessment and investigates the extent to which student writing performance may generalized across multiple tasks. Results from a Generalizability study found that 77% of error variance may be attributable to differences within people across multiple writing assignments.

Guemin and In-Yong (2012) assessed a comparison of the approaches of Generalizability Theory (GT) and Item Response Theory (IRT) in estimating the reliability of test scores for testlet-composed tests, the study was designed to address issues related to the extent to which item-based estimation methods overestimate the reliability of test scores composed of testlet and to compare several estimation methods for different measurement models using simulation techniques. The results of the study revealed that reliability estimates from TSA were lower than those from INTA due to loss of information with IRT approaches. However, this could not be applied in GT.

Egbulefu (2013) carried out study on estimation of measurement error and score dependability in examination using GT. The population of the study comprised 25,230 senior secondary three (SS3) students in Rivers state. 2,553 SS3 students participated in the study. A Mathematics Achievement Test with items drawn from past WAEC and NECO SSCE questions was used for data collection. EduG version 6.0-e based on ANOVA and Generalizability theory was used to answer the four research questions. A 95% confidence interval was computed using the S E variance components to determine whether there was a significant difference in the contributions and effects of the facets and their interactions to measurement error and score dependability in examinations. The findings of the study revealed that some hidden sources of error were at play in the study. The residual made the highest contribution to measurement error. This was followed by the student factor. Similarly, the residual and the students variance components were significantly ($p < 0.05$) different in their contributions to measurement error in examination scores. Conversely, questions and invigilators were not significantly different in their contributions and effects on measurement error and score dependability in examinations ($p > 0.05$). The findings also revealed that an increase of invigilators to 90, increased the generalizability coefficient (EP^2) and index of dependability (Ø) which rank ordered students and classified them based on their performance, irrespective of the performance of other students.

Appraisal of Related Literature Review

James and Daniel, (1981) assessed Multivariate Predictive Model of Organizational Commitment; the researchers opined that, a highly significant proportion of the variation in commitment within combined heterogeneous sample. Subsequently analyses of the model's Generalizability indicated that certain nuisance variable did not indicates significantly change in functional structure of the model or alter its ability to predict levels of commitment. Albert, (1984) studied Multivariate Generalizability Theory to Assessment of Student's Achievement in Art Education; the results indicated Generalizability coefficient is low with respect to different raters and moderate with respect to different topics. Dongmei and Brennan, (2007) at the Centre for Advance Studies in Measurement and Assessment (CASMA) studied the Multi-group Generalizability Analysis of a Large Scale Reading Compression Test. The results indicated that, the amount of discrepancy in G coefficients and error variances when facets are taken into account, differential contribution of their facets to measurement error for ELLs and native English speakers.

Finally, there were only two local studies on Generalizability (G) theory. All efforts made by the researcher to see if there were more local literatures or empirical works on G-theory, did not yield any result. This goes to confirm that Generalizability (G) theory, though old; is a relatively new concept in the Nigerian educational, clinical, marketing and psychological literature. In view of diversified findings in the empirical studies on Generalizability Coefficient of test items, it is clear that more studies are still required. From the reviewed studies, most of the researchers did not work on multivariate Generalizabilty of objective test items in Electrical Installations and Maintenance Works, only Tunde, (2015) worked on similar study, but the results cannot be generalized due to low population size and scope used. Equally, the City and Guild of London Institute was the examination body handling the trade courses before the inauguration of WAEC (technical) and later the establishment of NECO in 1990. Though, most of the questions that time were essay and not objective items, with the introduction of entrepreneurship subjects into secondary schools curriculum. The NECO introduces some trade subjects to be offered by science-based students in 2014; Electrical Installations and Maintenance Works was inclusive. Also being an indigenous examination body, that, they administered low quality items. Thus this motivated the researcher to fill these gaps and carry out study on Multivariate Generalizability of 2015 National Examinations Council Senior School Certificate Examination Objective Test in Electrical Installations and Maintenance Works in Nigeria; to determine the variance estimate components and the Generalizability coefficient, so that inference can be drawn on the 2015 National Examinations Council Senior School Certificate Examination Objective Test in Electrical Installations and Maintenance Works in Nigeria.

CHAPTER THREE

RESEARCH METHODOLOGY

This chapter deals with the procedures that were used to carried out the study under the following sub-headings: -

a. Research Design

b. Population, Sample and Sampling Techniques

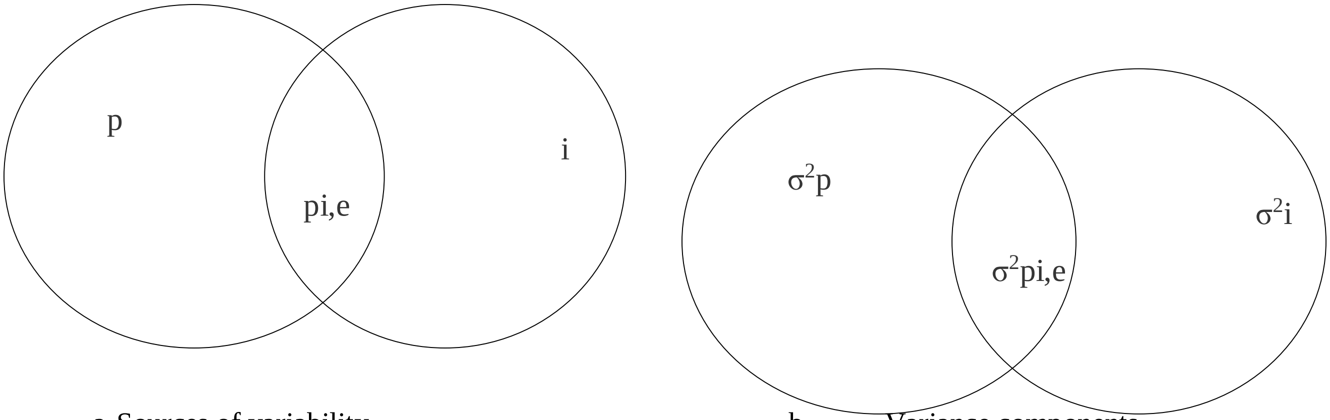
c. Instrumentation

d. Procedure for data Collection

e. Data analysis Techniques

Research Design

One facet persons by items ($p^{\bullet}$ x i°) crossed G study design was used for this study. Generalizability theory provides a framework to conceptualize and disentangle multiple sources of error. For administering Electrical Installations and Maintenance Works tests, with persons (p) as the object of measurement, one facet contributes to the person score of variability, i.e. items (i). It is usually the case that there are items that are intended for each one of the domain processes, and all persons would be administered with the same sets of items (Brennan, 2001a).



a. Sources of variability b. Variance components

Figure 2: Venn diagrams for a One-Facet, crossed (p x i) Design Shavelson &Webb, (1991)

Population, Sample and Sampling Techniques

The population for this study was all Public Senior Secondary School Students in Nigeria. The public schools were used because they have many things in common. They are owned and financed by the state governments, they are also comparable in terms of administrative structure, admission policy and selection process and management resources, recommended textbooks, materials for teaching and learning, also they use the same syllabus and scheme of work for the preparation of students for Senior School Certificate Examinations. The target population for the study consisted of 1,735 Senior Secondary School three (SSS3) Students offering Electrical Installations and Maintenance Works in all Senior Secondary Schools in Nigeria. These students were chosen because they are expected to have covered major parts of the SSCE syllabus. As well they are suitably qualified to write the tests adopted for this study because they are preparing for their Senior Secondary School Certificate Examinations. A total sum of one thousand one hundred and ninety-eighty (1,198) out of 3,488 students that registered for the course in the final Senior School Certificate Examinations in 76 public senior secondary schools participated in the study.

The purposive sampling technique was used for the selection of both schools and the students that participated in this study. A total of one thousand one hundred and ninety-eight (1,198) Senior Secondary Schools students offering Electrical installations and Maintenance Works in Nigeria were selected out of three thousand four hundred forty-eight (3,448) students of which was about ratio 1:3 of the total population that were involved in this study; which gives a fair representation of the sampled population.

Instrumentation

The (June/July) 2015 NECO Senior School Certificate Examination Objective Test in Electrical Installation and maintenance works was adopted and used as an instrument for this study. It consists of 40 items; this instrument is a standardized achievement test developed by National Examination Council an indigenous public examination body in Nigeria, and the researcher is of opinion that both the validity and reliability of this test might have been determined by the relevant unit of the said examination body before administration, hence the issue of validity and reliability estimation of this test items have been taken care of. The instrument was tagged National Examination Council Adopted Electrical Installation and Maintenance Achievement Test (NECOAEIMAT). The instrument was in one section (1) which consists of 40 Electrical Installations and Maintenance Works items. The key to instrument was collected from the National Examination Council.

Procedure for Data Collection

In the course of administering the instrument, the researcher visited each of the selected Senior Secondary Schools to seek permission from the school authority. Dates and time of the administration of the test were fixed in order not to affect school activities, prep time was suggested. The researcher administered the instrument to the S.S.3 students in each of the selected schools on the scheduled dates with the help of trained Research Assistants. The participants were guided to respond to the instrument independently. Duration of 1 hour was allocated for the instrument.

Ethical Consideration: -The participants were allowed to participate voluntarily; because involuntary human participation in a research study is unethical. The researcher refrained from using deception to gain the participation of subjects in this study. Being an Electrical Installations and Maintenance Works objective test, it is cognitive based; therefore, the instrument did not cause any harm to the participants, but instead it enable the participants be more inclined in Electrical Installations and Maintenance Works as a subject at senior secondary school level. The participants were allowed to give their consent to be involved in the study based on thorough knowledge of the procedures and obtained in the form of written and witnessed documentation from the authority of the sampled schools. Participants' consent and study results were kept absolutely confidential. When reporting the facts and data collected from the subjects their identities were not disclosed; all details about the participants were under the custody of the researcher and would not be revealed; all information gathered during the course of this study were handled with topmost confidentiality.

Data Analysis Techniques

The data collected for this study was subjected to analysis with due consideration to the five research questions answered in this study. Thus, after the administration of the instruments and scoring of the responses dichotomously, the data generated were analysed using Variance Component (VARCOMP) to appraise the estimate of variance components for persons, items, persons by items, Generalizability and dependability coefficients.

CHAPTER FOUR

DATA ANALYSIS AND RESULTS

This chapter presents the results of analysis of data collected for the study. The results are presented according to the five research questions generated.

To answer the research questions, each item of 2015 NECO Electrical Installations and Maintenance Works objective test were scored dichotomously and entered into the SPSS version IBM 21 and subjected to syntax analysis to obtain the three variance components. Table 2 reveals the results obtained.

Table 2: Estimated Variance Components and there corresponding total variance percentage for 2015 NECO Electrical Installations and Maintenance Works objective test

Sources	**Variance Comp.**	**Estimated Variance**	**% of Total**
Variance			
		Component	**Variance**
Persons (p)	σ^2p	0.02	8%
Items (i)	σ^2 i	0.03	12%
Residual (pi,e)	σ^2pi, e	0.20	80%
Total		**0.25**	**100%**

Research Question 1: **What is the variance component due to persons (testees) in the 2015 NECO SSCE objective test in Electrical Installations and Maintenance Works?**

To answer this research question, the data collected were entered into the SPSS (IBM version 21) data view, after which the data were subjected to syntax analysis to obtain the estimate variance component for persons. The result obtained, is extracted from Table 2 presented in Table 3.

Table 3: Estimated Variance Component for Persons in 2015 NECO Electrical Installations and Maintenance Work objective test

Sources Variance	Variance Comp.	Estimate of Variance Component	% of Total Variance
Persons (p)	σ^2p	0.02	8%

The research question was aimed to show how much variance component for persons in the observed scores is due to differences among persons characteristics. From the table 3 it was revealed that the estimated variance component for persons (σ^2p) is 0.02 which account for 8% of the total variance in the 2015 NECO SSCE Electrical Installations and Maintenance Works objective test. The estimated variance component for persons (0.02) is the lowest as it accounted 8% of the total variance in the 2015 objective test in NECO SSCE Electrical Installations and Maintenance Works. See appendix III on page 144 illustrates how it was obtained.

Research Question 2: **What is the variance component due to items used in the 2015 NECO SSCE objective test in Electrical Installations and Maintenance Works?**

To answer this research question, the data collected were entered into the SPSS (IBM version 21) data view after which the data were subjected to syntax analysis to obtain the estimate variance component for items. The result obtained; which is extracted from Table 2 is presented in Table 4.

Table 4: Estimated Variance Component for Items in 2015 NECO Electrical Installations and Maintenance Works objective test

Sources Variance	**Variance Comp.**	**Estimate of Variance Component**	**% of Total Variance**
Items (i)	σ^2 i	0.03	
12%			

The research question sought to show how much variance component for items in the observed scores has effects among items. The estimated variance component for items (σ^2 i) is 0.03 and it accounts for 12% of the total variance. See appendix III page 144 shows it was obtained.

Research Question 3: **What is the variance component due to the interactions of persons by items in the 2015 SSCE objective test in Electrical Installations and Maintenance Works?**

To answer this research question, the data collected were entered into the SPSS (IBM version 21) data view after which the data were subjected to syntax analysis to obtain the estimate variance component for persons by items (residual). The result obtained, extracted from Table 2, is presented in Table 5.

Table 5: Estimated Variance Component for persons by items in 2015 NECO Electrical Installations and Maintenance Works objective test

Sources	Variance Comp.	Estimate of Variance	% of Total
Variance			
		Component	Variance
Residual (pi,e)	σ^2pi, e	0.20	80%

This research question sought to show the effects of the persons by items and their interactions to Generalizability and dependability coefficients of the 2015 objective test in Electrical Installations and Maintenance Works. The estimated variance for persons by items (σ^2pi, e) is 0.20 which accounts for 80% of the total variance component in the 2015 Electrical Installation and Maintenance Works objective test and is largest estimated component as it accounted for 80% of the total variance components. See appendix III page 144 on how it was obtained.

Research Question Four: What is the Generalizability coefficient of the 2015 Senior NECO School Certificate Examination objective test in Electrical Installations and Maintenance Works?

Relative error variance and Generalizability coefficient formulae were used to answer this research question. The estimated variance component for persons (σ^2p) and persons by items (residual) σ^2pi,e were used to determine the relative error variance while the estimate variance component for persons and relative error variance were also used to obtain Generalizability coefficient. The results are presented in Table 6. See appendix iv page 145 on how it was obtained.

Table 6: Relative error variance and generalizability coefficient for 2015 NECO Electrical Installations and Maintenance Works objective test

Relative error variance	Generalizability coefficient
0.0051	0.8000

Table 6 shows the estimated relative error variance of 0.0051 and generalizability coefficient of 2015 NECO objective test in Electrical Installations and Maintenance Works objective test is 0.80. Therefore, the generalizability coefficient obtained is high or acceptable. Since the value obtain is not less than the acceptable value of 0.70

Research Question Five: What is the Dependability coefficient of the 2015 Senior School Certificate Examination objective test in Electrical Installations and Maintenance Works?

The D study or dependability coefficient for the 2015 NECO objective test in Electrical Installations and Maintenance Works objective test is obtain using absolute error variance and dependability equations respectively; Table 7 revealed the results obtained. See appendix iv page 146 on how it was obtained.

Table 7: Relative error variance and dependability coefficient for 2015 NECO Electrical Installations and Maintenance Works objective test

Absolute error variance	Dependability coefficient
0.0030	0.7800

Table 7 revealed that the absolute error variance is 0.0030 while the D study or dependability coefficient of the 2015 NECO Electrical Installations and Maintenance Work objective test is 0.780. This shows that dependability of the 2015 NECO Electrical Installations and Maintenance Work objective test is acceptable since the obtained value does not below the acceptable value of 0.70. From the results of this study, it is was revealed that 0.80 Generalizability coefficient and 0.79 D study were obtained; this shows that there is no need to increase the number of item in 2015 NECO Electrical Installations and Maintenance Works objective test. Further D study could have been carried out assuming the Generalizability and dependability coefficients are not above the acceptable value of 0.7.

Summary of the Findings

The summaries of answers to the five research questions posed for this research work are as follows:

1. The estimate variance due to presons (testees) is 0.02 which account for 8% 0f the total variance in the 2015 NECO SSCE Electrical Installations and Maintenance Works objective test.
2. The estimate variance in the 2015 NECO SSCE Electrical Installations and Maintenance Works objective test due to items used in the test is 0.03 that account for 12% and was the second largest variance.
3. The estimate variance in the 2015 NECO SSCE objective test in Electrical Installations and Maintenance Works due to the interaction of presons by items is 0.02 which account for 80% of the total variance and is the largest variance.
4. The Generalizability coefficient of the 2015 NECO SSCE Electrical Installations and Maintenance Works objective test is 0.80. This means that Generalizability coefficient is high or acceptable.
5. The Dependability coefficient of the 2015 NECO SSCE Electrical Installations and Maintenance Works objective test is 0.80. This shows that Dependability coefficient of the 2015 NECO SSCE Electrical Installations and Maintenance Works objective test is high or acceptable.

CHAPTER FIVE

DISCUSSIONS, CONCLUSION AND RECOMMENDATIONS

This chapter focuses on the discussions of findings, conclusions and recommendations that were proffered; educational implications to bring an improvement to the present situation and for further research.

Discussion of Findings

The highest contributions to measurement error in the test was the residual (ópi,e) 0.20 accounting for 80% of the total variance. This showed that a proportion of the variance was due to the interaction of persons by items and other systematic or unsystematic source of variance that were not measured in the study. The second largest source of variation to measurement error was due to differences among items with variance component of 0.03 accounting for 12% of the total variance. This indicates that the persons somehow distinguish among items.

This study was in consonance with the study of Shavelson, Pine, Goldman, Baxter and Hine (1989), who reported that One-Facet Crossed design was used and revealed that the highest estimate variance component was residual 0.2103 which account for 84% of the total variance, followed by variance component for persons 0.0305 accounting for 12% and variance component of items 0.0093 which account for 4% of total variance respectively. But not in line with the Generalizability Coefficient which was found to be 0.54 relatively low. The study was also in line with Tunde (2015) who reported that the residual has the major source of measurement error followed by persons' variance component and lastly items variance component. The findings of this study was equally supported in Shavelson and Webb (1991), whose study found that the residual as the largest contributor to measurement error. The outcome of this study was supported by the findings of Hintze and Peltite (2001) on performance based assessment.

Findings in this study revealed that the Generalizability coefficient of the 2015 NECO Electrical Installations and Maintenance Works objective test was 0.80. This indicated that the estimated relative error variance (Ó$^{2}_{\mathrm{Rel}}$) which is the difference between persons observed deviation scores and their universe deviation scores (Brennan, 2001a) was obtained through the variance components that contributed to observed scores alone. The estimated relative error variance and estimated variance components that were used to obtain the Generalizability coefficient of the 2015 NECO Electrical installations and maintenance Works objective test were high or acceptable. Therefore, if Generalizability coefficient of 0.80 is to be Classical Test Theory, it will then be regarded as a high reliability. The researcher also find out the dependability coefficient of the 2015 NECO Electrical Installations and Maintenance Works objective test from the finding it was revealed that the estimated absolute error variance (Ó$^{2}_{\mathrm{Abs}}$) for the design of this study is type difference between person’s observed score and their universe score variance (Brennan, 2000). Two out of the three variance components in this study contributed to absolute variance error, it was only variance component for persons that did not contributed to absolute variance error. The obtained D study or dependability coefficient was high 0.79 considering the 0.70 level of acceptability value. Therefore, the dependability of the 2015 NECO Electrical Installations and Maintenance Works is high as the dependability coefficient is the parameter used to determine the dependability level of an instrument (Nie, et al 2007). The high dependability level of the 2015 NECO Electrical Installations and Maintenance work objective test may be due to the contributions of the three sources of measurement errors and to the differences in the persons scores and high level of achieving content validity and quality of the teachers teaching the subject. The interpretations given to the variance components are in line with Shalveson and Webb's (1991b) they opined that when the estimated variance components are added together, the outcome reveals that total variance components and the variance component that has the largest percentage contributed the largest part of the total variation while the variance component with lowest percentage contributed the least part of the total variation.

This finding is not in line with that of Turner's et al (2010), they carried out study on the effects of ventilation on segmental bio- impedance spectroscopy measures using Generalizability theory and found that the largest source of variation was persons (P) variance components but this study has persons by item (pi,e) as the largest source of variation. The findings of this study are not in line with findings of Turner's (2010) study. It may be due to the fact that the study was carried out in North America and this study was carried out in Nigeria; this could be an indication of location difference in researching. Equally, 100 participants were used in Turner's study while, 1,198 participated in this study; these factors could be responsible for the differences.

Conclusion

It could be concluded from the findings of this study that the value of generalizability and dependability coefficients of 2015 NECO Senior School Certificate Examinations

(SSCE) in Electrical Installations and maintenance Works objective test is high or acceptable. This means that the quality of items administered in Electrical Installations and Maintenance Works was moderately okay. This could be due to the contributions of the three sources of measurement errors to the differences in the 2015 NECO SSCE Electrical Installations and maintenance Works and the number of items used for the study; on this note, the measures of the 2015 NECO Electrical Installations and Maintenance Works objective test is standard.

Educational Implications of the Study

The implications of the findings of this study are that it exposes:

1. the need for educational researchers and evaluators to use Generalizability and dependability to determine the quality of items and other measurement or parameters behaviours rather than making use of reliability coefficients alone.
2. the examination bodies operating at these levels the procedures of obtaining items that are of acceptable (0.7) level.
3. the society that the objective tests used by NECO in 2015 in Electrical Installation and Maintenance Works is of high standard
4. educational researchers and evaluators that this can also be obtained in both the theory and practical items

Recommendations

Based on the findings of this study, the following recommendations were proffered:

1. The quality of the items in Electrical Installations and Maintenance Works objective test should be maintained and extended to other subjects.
2. The quality of the items should also be maintained in the theory and practical examinations.
3. The Generalizability and dependability coefficients were high or acceptable meaning that the content validity was achieved. Therefore Educational Administrators and Teachers should attainment of content validity always.
4. Educational researchers and evaluators should follow the procedures for estimating high dependability level of Electrical Installations and maintenance Works objective test and other measurement behaviours especially when the Generalizability and Dependability coefficient obtained are low.
5. The Educational Evaluators should endeavour to carry out similar studies in other areas of vocational tests so that inference can also be made on the quality of items.

Suggestions for Further Studies

This study was carried out to analyse the multivariate Generalizability of 2015 senior school certificate examination objective test in electrical installation and maintenance works in Nigeria. The sample size could be increased; also other subjects could be investigated by other researchers. Also other forms of designs could be used such as One facet nested Random, Two facet crossed Random, Two facet nested Random, One facet crossed Fixed, One facet nested Fixed, Two facet crossed Fixed and Two facet nested Fixed, Two Facet Crossed, Partial Nested with one Facet could be used by other scholars. Further studies are also encouraged for replication of study in other vocational course.

REFERENCES

Abiri, J.O.O (2007). *Element of Evaluation Measurement and Statistical Techniques in Education, Ilorin:* Library and Publication Committee, University of Ilorin.

Abodunrin, I.O. (1999). *Evaluating Teaching/Learning Effectiveness; A basic Course in Educational Test, Measurement, and Evaluation.* Oro; Cardinal Information on Technology.

Adewale, A. E. (2005). Comparative Analysis of the Psychometric Properties of WAEC, NECO and NABTEB Mathematics Achievement Tests; A Ph.D. Thesis submitted to the Department of Guidance and Counselling, Faculty of Education, University of Ado-Ekiti, Nigeria

Adewuni, D. A. (2016). Hierarchical Cluster Analysis of the Dimensionality of Senior School Certificate Objective Test in Government; A Ph. D. Thesis presented to the Department of Social Sciences Education, Faculty of Education, University of Ilorin; Ilorin, Nigeria.

Albert, N. (1984). Multivariate Generalizability Theory in Educational Measurement: An Empirical Study, *Psychological Measurement Inc.* 8, 2, 219-230.

AERA, APA, & NCME (1985). *Standards for educational and psychological testing.* Washington DC: American Psychological Association.

Aggarwal, J.C. (1997). Essentials of Examination System. New Delhi Vikas Publishing House PVT Ltd

Airasian, P. (1994). "*Classroom Assessment" 2nd Edition,* New York: Mc Graw-Hill.

Akeju, S.A. (1972). The reliability of general certificate in education examination, english composition papers in West Africa. *Journal of Educational Measurement*. Seminar 1972. Retrieved March 10th 2011 from http://www.Jstor.org/pss/1434162

Allen, M.J. & Jen, W.M. (1979). *Introduction to measurement theory.* California: Brooks/Cole.

Allen, M.J. & Yen, W.M. (2002). *Introduction to Measurement Theory*. Long Grove,IL: Waveland Press.

Ayodele, Z.O. (1999). The meaning of Vocational Technical Education, Readings in Vocational Technical Education, Yaba.

Ary, D., Jacobs, L. & Razavieb, A. (1996). *Introduction to research in education.* (5th Eds.). US: Harcourt Brace & Co.

Bachman, L. (2004). *Statistical analyses for language assessment*, Cambridge: Cambridge University Press.

Bachman, L.E., Lynch, B.K. & Mason, M. (1995). Investigating variability in tasks and rate judgment as a performance. *Test of Foreign Speaking Language Testing*, **12**, 239-257.

Bamidele, S.O. (2004). Educational Research in perspective. Ibadan Niyi Commercial and Printing Ventures.

Bolus, R. E., Hinofotis, F. B. & Bailey, K. M. (2006). An introduction to generalizability theory in second language research. *Language Learning, 32(2), 245-258*

Breland, H., Bridgeman, B. & Fowles, M.E. (1999). *Writing education: Review and Framework* (ETSRR-99-3). Princeton, NJ: ETS.

Brennan, R.L. (1984). *Estimating the dependability of scores*. In K.A. Berk (ed.), A guide to criterion-referenced test construction 292-334. Baltimore: Johns Hopkins University Press.

Brennan, R. L. (1998). Raw-score conditional standard errors of measurement in generalizability theory. *Applied Psychological Measurement*, 22(4), 307-331.

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

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

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

Brennan, R. L., Yin, P., & Kane, M. T. (2003). Methodology for Examining the Reliability of Group Mean Difference Scores. *Journal of Educational Measurement*. 40, 3, 207-230.

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

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

Briesch, A. M.,Swaminathan, H., Welsh, M. & Chafouleas, S. M. (2014). Generalizability theory: A practical guide to study design, implementation and interpretation. Journal of School Psychology, 52(1), 13-35

Burns, N. & Grove, S. (1997). *The Practice of nursing research; conduct, critique and utilization.* 3rd edu. WB. Saunders Company Philadelphia

Carroll, J.B. (1990). *Future Developments in educational measurement*. In JP. Keeves (Ed.). Educational research methodology and measurement (2nd ed. 247-53). Sudney. Pergamon Press.

Chen, M.J. & Fan, X. (1998). The relationship between variance components and mean difference effect size. Current psychology: Issues December 1998. New York: *Springer* **5** , 301-312.

Clark, L. (2008). Assessment is for learning. Formative assessment and positive learning interactions Florida journals of educational Administration and policy, 2 (1):1-16.setting results within a generalizability theor framework. *Journal of Educational Measurement in Education, 12(3), 281-299.*

Cluaster, J. C., Margolis, M. J. & Cluaster, B. E. (2014). An examination of the replicability of angoff standard

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich.

Cronbach, L. J., Gleser, G. C., Nandam, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability scores and profiles*. New York: John Wiley.

Cuiyun, M. Yuanhang, H., Qiang, M. & Jianyi, L. (2010). Decision of entrepreneurs' inner psychological qualities evaluation based on multivariate generalizability theory. Wuhan, China: International Conference on Management and Service Science.

Dongmei, Li & Robbert, L.B. (2007). A Multi-group Generalizability Analysis of *A Large-Scale Reading Comprehension Test; Centre for Advance Studies in Measurement and Assessment Research Report* No 25.

Eason ,S. (1991). Why generalizability theory yields better result from classical test theory. A primer with concrete examples. In B. Thompson (Ed.). Advances in Educational research: Substantive findings, methodological development **1,** (83-98). Greenwich, CT: JAL.

Eason, S. (1989). *Why Generalizability theory yields better results than classical test theory*. Paper presented at the Annual Meeting of the Mid South Educational Research Association (little Rock, AR, November 8-10, 1989). (ERIC Document Reproduction Service No ED. 314434).

Ebel, R.L. & Fribie, D.A. (1986). *Essential of educAtional measurement* (4th edition.). Englewood Cliffs, NJ: Prentice-Hall.

Egbulefu, C.A. (2013). Estimating Measurement Error and Score Dependability in Examination Using Generalizabilty Theory. A Ph. D. Thesis presented to the Department of Sciences Education, Faculty of Education, University of Nigeria; Nsukka, Nigeria.

Emaikwu,S.O.(2011). *Issues in test items bias in public Examination in Nigeria and implications for testing. International Journal of Academics Research in Progressive and Education and development, 1 (i) pages 175-187.*

Fafunwa, A. Babs, (1967). *New Perspective in African Education.* Macmillan & Co. Ltd. Ibadan, Nigeria.

Fafunwa, A. Babs, (1991). Up and On: *a Nigerian Teachers*, West Africa Book Publisher Ltd Lagos.

Feldt, L. & Brennan, R. (1989). ‘Reliability’. In R.L. Linn (Ed.). Measurement (3rd Edition, R. Linn Ed.), 105-146. New York: Macmillan.

Frisbie, D.A. (1988). Reliability of scores from teacher-made tests. Educational measurement: *Issues and Practice* **7**, 25-35.

Follesdal, H. & Hagtvet, K.A. (2009). Emotional intelligence: The MSCEIT from the perspective of generalizabilty theory, intelligence, 37, 94-105

Gao, X. & Brennan, R. L. (2001). Variability of estimated variance component and related statistics in a performance assessment. *Applied Measurement in Education*, 14(2), 191-203.

Gebril, A. (2013). Generalizabilty theory in language testing. The Encyclopedia of applied linguistics, 1-7

Gill, Flitman & Dar, (2000). Vocational Education and Training Reform. Matching Skills to Makes and Budgets; Oxford University Press, New York

Gronlund, N. (1993). How to make an Assessment 5th Edition, NY: Allyn and Bacon.

Gronlund, N.E. & Linn, R.L. (1990). *Measurement and Evaluation in Teaching;* N.Y. Allyn and Bacon.

Hassan, T (1991). *Assessment Tools in Counselling Practicum.* In S.A. Gesinde (ed) Reading in Counselling Practicum Ibadan; Vantage

Hagtvet, K. A. & Hanin, Y. L. (2007). Consistency of performance-related in elite athletes: Generalizability theory applied to the IZOF model, *Psychology of sport and Exercise*, *8, 47-72*

Herman, J.L., Webb, N.M. & Zuniga, S.A. (2007). Measurement issues in the alignment of standards and assessments: A Case study. A*pplied Measurement in Education*, **20**, 101-126.

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

Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments? A generalizability theory approach. Retrieved 6th 2010 from http:/www.niagara.edu/assets/assets/ncetelstadards/3/3.6f.DnHuang.

Isu, C. (2000). Innovation in Business Education Curriculum: A paper Presented at 1st Annual National Conference of the Nigeria Association of Contemporary Nigerian Business Environment from 24th-27th October 2000.

Iyewarun, S.A. (1984). The Teaching of Social Studies. Ilorin, Woye Press

James, H.M. & Daniel, J.S. (1981). *Generalizability of an Organisational Commitment Model; Academy of Management Journal*, Vol.24, No.3, pp 512-526

Johann, L. W & Fans, s. (2008). *The Validation of Language Tests. Stellenbosched papers in linguistic, vol. 38, pg. 191-240*

John, D.H. & Jeremy, D.P. (2012). Generalizability of Student Writing across multiple Tasks: A Challenge for Authentic Assessment; *Research and Practical in Educational Assessment* Vol.07

Johnson, S., Dulanay, C. & Banks, K. (2000 February). Measurement error. Retrieved April 3, 2009 from http://www.wcpss.net/evaluation.research/reports/2000/mment_error. pdf.

Kane, M.T. (1993). Review of the book Generalizability theory. *A Primer Journal of Educational Measurement*, **30**, 269-272.

Kane, M.T. (2008). Errors of measurement theory and public policy. Retrieved from http.//ets.org/media/Research/pdf/P.CANG12.PDF.

Keeves, J.B. (1990). *Educational research methodology and measurement.* An International Handbook. Sydney. Pergamon Press.

Keller, L. S., Clauser, B. E. & Swanson, D. B. (2010). Using Multivariate generalizability theory to assessed the effect of content strafication on the reliability of a performance assessment. *Advance in Health Sciences Education.*

Kieffer, K. M. (1999). Why generalizability theory is essential and classical test theory is often inadequate. In B. Thompson (Ed.), *Advances in socialscience methodology* (Vol. 5, pp. 149-170). Stamford, CT: JAI.

Kolawole, E.B. (2001). Test and Measurement, Yemi Prints and Publishing Services, Ado-Ekiti.

Kozaki, Y. (2004). Using GENOVA and FACETS to set multiple standards on performance assessment for certification in medical translation from Japanese into English Language Testing, 21, 1-27.

Kuder, G.F. & Richardson, M.W. (1937). The theory of the estimation of test reliability. *Psychometrika*. **2**, 151-160.

Li, M. N. F., & Lautenschlager, G. (1997). Generalizability theory applied to categorical data. *Educational and Psychological Measurement, 57(5), 813-822*

Lin, C. K. & Zhang, J. (2014). Investigating corresponding between language proficiency standards and academic content standards – A generalizability theory study. *Language Testing, 31 (4), 413-431*

Lisser, A. K., Brian, E. C. & David B. S. (2010). Using Multivariate Generalizabilty Theory to assess the effect of content stratification on the reliability of a performance assessment; University of Massachusetts Amherst, Amherst, MA USA.

Lombardi, A., Seburn, M., Conley, D. & Snow, E. (2010). *A Generalizability of cognitive demand and rigor ratings of items and standards in an alignment study*. Presented at the Annual Conference of the American Educational Research Association Denver, Co. Educational Policy Improvement Center 720E. 13th Ave; Suite 202.

LoPliato, A. C., Carter, N. T. & Wang, M. (2014). Updating generalizability theory in management research: Bayesian estimation of variance components. *Journal of management, 41(2), 1-26*

Lord. F., Novick, T. (1968). *Statistical theories of mental test scores*. Reading, M: Addison Wesley.

Macmilian, P.D. (2000). 'Classical; generalizability and multifaceted rasch detection of interrater variability in large, sparse data sets', *Journal of Experimental Education,* **66**, 167-190.

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

Meadows, M. & Billington, C. (2000). A review of the literature as making reliability, report to NNA. Retrieved from www.naa.ong.uk/library/assets/media/Review_of_the_literature_on_making_reliability pdf.

Messick, S. (1989). validity, In R.I.L.Lin(Ed),Educational Measuring (3rded ,13-104). Newyork:American council on Education and Macmillan.

McGrew, K.S., Johnson, D.R., Cosio, A. & Evans, J. (2003). Increasing the chance of no child being left behind: Beyond cognitive and achievement abilities. Unpublished manuscript, University o Minnesota.

Mitchell, S.K. (1979). *Inter observer agreement, Reliability and Generalizability of Data collected in observational studies; psychological* Bulletin, 2, 376-390.

National Policy on Education Revised Edition (2013).

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

Neworgel, B.G. (1992). Educational Measurement and Evaluation; theory and practice Awka;Hallman publisher.

Nworgu, B.G. (1992). Educational Measurement and Evaluation; Theory and Practice. Awka Hallman Publishers

Obinne, A.D.E. (20011). A Psychometric Analysis of Two Major Examinations in Nigeria: Standard Error of Measurement. International Journal of Education Science, 3 (2): 144-147

Osuala, E.C. (1985). The Development of Business Education With of Implication for Meeting the Manpower Needs of Nigeria: Vocational Education Journals

Rivera, J.E (2007). Test Item Construction and Validation: Developing A Statewide Assessment for Agricultural Science Education, A Ph.D Dissertation Presented to the Faculty of Graduate School, Cornell University.

Shavelson, R.J., Webb, N.M. & Rowley, G. (1989). *Generalizability theory. American Psychologist,* **49**, 922-932.

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

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

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

Shavelson, R.J., Ruiz-Primo, M.A. & Wiley, E.W. (1999). Note in sources of sampling variability in science performance assessments. *Journal of Educational Measurement,* **36**, 61-71.

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

Solano-Flores, G. & Li, M. (2006). The use of generalizability theory in testing of linguistic minorities. *Journal of Educational Measurement: issues and practice*

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, **15**, 72-101.

Spearman, C. (1904b). In classical test theory and the measurement of reliability. Retrieved March, 4, 2010. from http://ww.personality.project.org/R/book/ chapter7pdf.

Stora, B., Hagtvet, K. A . & Heyerdahl, S. (2013). Reliability of observers' subjective impression of families: A generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. Assessing Writing. 9, 239-261

Taylor, M. A. & Pastor, D. A. (2013). An application of generalizabilty theory to evaluate the technical quality of an alternate assessment. *Applied Measurement in Education, 26, 279-297*

Thakur & Ezenne (1980). A Short Story of Education in Nigeria, Agoro Publicity Co. Ibadan.

Thompson, B. & Vacha-Haase, T. (2000). Psychometrics is data metrics: The test is not reliable. *Educational and Psychological Measurement*, **60**, 174-195.

Thompson, B. (1991). Review of the book Generalizability theory: *A nume Educational and Psychological Measurement*, **51**, 1069-1075.

Thompson, B. (1992). Two and one half decades of leadership in measurement and evaluation. *Journal of Counseling and Development*, **70**, 434-448.

Thorndike, R.L. (1990a). Reliability. In E.F. Lindquist (Ed.). Educational measurement (560-620) Washington, DC: American Council on Education.

Thorndike, R.L. (1990). Reliability. In J.P. Keeves (ed.). Educational resaerch methodology and measurement (2nd ed. 330-31). Sydney. Pergamon Press.

Traub, R.E. & Rowlay, G.L. (1991). Understanding reliability. Educational Measurement. *Issues and Practice*, **10**, 37-45.

Trochim, W.M.K. (2006). Introduction to Validity, Social Research Methods, Retrieve from www.socialresearchmethods.net/kb/Introval.php.September 9, 2010

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

Turner, A.A., Lozano-Neito, A. & Bouffard, M. (2010). Effects of Ventilation on Segmental Bio-impedance Spectroscopy measure using Genralizability Theory.*Measurement in Physical Education and Exercise Science, 14(2), 16-129*

Wallace, S. (2009). Oxford Dictionary of Education: Oxford University Press, London.

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

Watkin, J.M., Lee, H.B. & Erlich, O. (1980). *Toward a dependability of reflection-impulsivity: An application of generalizability theory. Journal of Behaviour Assessment, 2(1), 1-6.*

Webb, N. M. & Shavelson, R. J. (1981). Multivariate Generalizability of General Educational Development Ratings. *Journal of Educational Measurement*. Vol.18. NO. 1, pp13-22.

Webb, N. M. & Shavelson,R. J. & Haertel, E. H. (2007). Reliability coefficient and generalizability theory. Handbook of Statistics, 26 (4), 81-125.

Webb, N.M., Rowley, G.& Shavelson, R.J. (1988). Using generalizability theory in counseling and development. *Measurement and Evaluation in Counselling and Development*, **21**, 81-90.

Webb, N.M., Schlackman, J., Sugrue, B. (2000). The dependability and interchangeability of Assessment methods in science. *Applied Measurement in Education,* **13**, 277-301.

Webb, N.M, Shavelson, R.J., & Hartel, E.H. (2007). "Reliability coefficient and generalizability theory' Handbooks of statistics 26: Psychometrics (C. Rao and S. Sinharay Eds.). 81-124, the Netherlands: Elsevier.

Worther, B.R, Borg. W.R. & white, K.R (1993). Measurement and Evaluation in the Schools. White Plains. NY. Longman

Youyan, N. Shu, M.Y. & Shun, L (2006). Application of Generalizability Theory in the investigation of the Quality of Journal Writing in Mathematics: Studies in Educational Evaluation (33) 371-383 Elsevier Ltd Singapore.

Youzhen, Z. (2007). A Multivariate Generalizability Analysis of Student Style Questionnaire: Thesis presented to the Graduate School of the University of Florida for the award of Master of Arts Education.

APPENDIX 1

SOURCES OF VARIABILITY IN ONE-FACET CROSSED DESIGN

Sources of Variability in One-Facet Crossed Design Measurement

Source of variability	**Type of Variability**	**Variance Notation**
Person (p)	Universe score	$\sigma^2 p$
Items (i)	Conditions	$\sigma^2 i$
Person by item Interaction	Residual	$\sigma^2 pi,e$

The sources of error in a One-Facet Crossed Design as observed in above is three i.e. persons, items and the residual which represents both the interactions between persons and items and random error. Brennan (2010a) supported this view that, the interactions between persons by items and random error be combined.

APPENDIX II

ESTIMATED VARIANCE COMPONENTS FOR ONE FACET CROSSED DESIGN

Estimated Variance Components from Mean Square for One Facet Crossed Design

Sources	**Mean Square**	**Expected Mean**	**Estimated**
Variance		**Square**	**Component**
Persons (p)	MSp	$\sigma^2 pi,e + ni\sigma^2 p$	$\sigma^2 p=(MSp-\sigma^2 pi,e)/ni$
Items (i)	MSi	$\sigma^2 pi,e + np\sigma^2\ i$	$\sigma^2 i=(MSi - \sigma^2 pi,e)/np$
Residual (pi,e)	MSpi,e	$\sigma^2 pi,\ e$	$\sigma^2 p= MSpi,e$

This shows the estimated variance components from mean square for One Facet Full Crossed Design. The estimated mean square are meant for the population, while mean squares are sample and this is used to derive the variance component for each sources of variability.

APPENDIX III

VAR IANCE COMPONENTS ESTIMATION FOR THE DESIGN OF THIS STUDY

One Facet Full Crossed Design is the design for this study

The variance components were estimated through VARCOMP of SPSS, IBM version 21. The outputs were stated as follows:

```
VARCOMP
Person_id item_id
/RANDOM=person_id item_id
/METHOD=REML
/DESIGN=person_id item_id
/INTERCEPT=INCLUDE
```

Estimate Variance Component From SPSS

Component	Estimate
Var(Persons_id)	0.0200
Var(Items_id)	0.0300
Var(Error)	0.2000

Dependent Variable: person
Method: Restricted Maximum Likelihood Estimation

Estimation Variance Components

Sources Variance Component	Variance Comp.	Estimated Variance
Persons (p)	$\sigma^2 p$	0.02
Items (i)	$\sigma^2 i$	0.03

Residual (pi,e)	σ^2pi, e	0.20
Total		**0.25**

The percentage of variance was calculated by adding all the variance components together and divided by each of the variance.

Thus 0.02 + 0.03 + 0.2 = 0.25

Percentage of total variance for each estimated variance components are calculated below:

Percentage of total variance component for persons σ^2p = $\frac{0.02}{0.25}$ x 100 = 8%

Percentage of total variance for each estimated variance components are calculated below:

Percentage of total variance component for items σ^2i = $\frac{0.03}{0.25}$ x 100 = 12%

Percentage of total variance for each estimated variance components are calculated below:

Percentage of total variance component for residual σ^2pi,e = $\frac{0.20}{0.25}$ x 100 = 80%

Estimation Variance Components and their Equivalent Percentage

Sources Variance	**Variance Comp.**	**Estimated Variance Component**	**% of Total Variance**
Persons (p)	σ^2p	0.02	8%
Items (i)	σ^2 i	0.03	12%
Residual (pi,e)	σ^2pi,e	0.20	80%
Total		**0.25**	**100%**

APPENDIX IV

ESTIMATION OF RELATIVE ERROR VARIANCE AND GENERALIZABILITY COEFFICIENT

The estimation was carried out using the formulae for both relative error variance and Generalizability coefficient.

$$\sigma^2 Rel = \frac{\sigma^2 p}{ni} + \frac{\sigma^2 pi,e}{ni}$$

Where $\sigma^2 p = 0.02$, $ni = 40$ and $\sigma^2 pi,e = 0.20$

$$\sigma^2 Rel = \frac{0.02}{40} + \frac{0.20}{40} = 0.0051$$

$$\sigma^2 Rel = 0.0051$$

Generalizabilty coefficient $p^2 = \frac{\sigma^2 p}{\sigma^2 p + \sigma^2 Rel}$

Where $\sigma^2 p = 0.02$ and $\sigma^2 Rel = 0.0051$

$$p^2 = \frac{0.02}{0.02 + 0.0051} = 0.80$$

$$p^2 = 0.80$$

APPENDIX V

ESTIMATE OF ABSOLUTE ERROR VARIANCE AND DEPENDABILTY (Ø) COEFFICIENT

The estimation was done through the formulae for both absolute error variance and Dependability coefficient.

Given that Absolute Variance $\sigma^2\text{Abs} = \frac{\sigma^2 i}{ni} + \frac{\sigma^2 pi,e}{ni}$

Where $\sigma^2 i = 0.03$, $ni = 40$ and $\sigma^2 pi,e = 0.20$

$$\sigma^2\text{Abs} = \frac{0.03}{40} + \frac{0.20}{40} = 0.0030$$

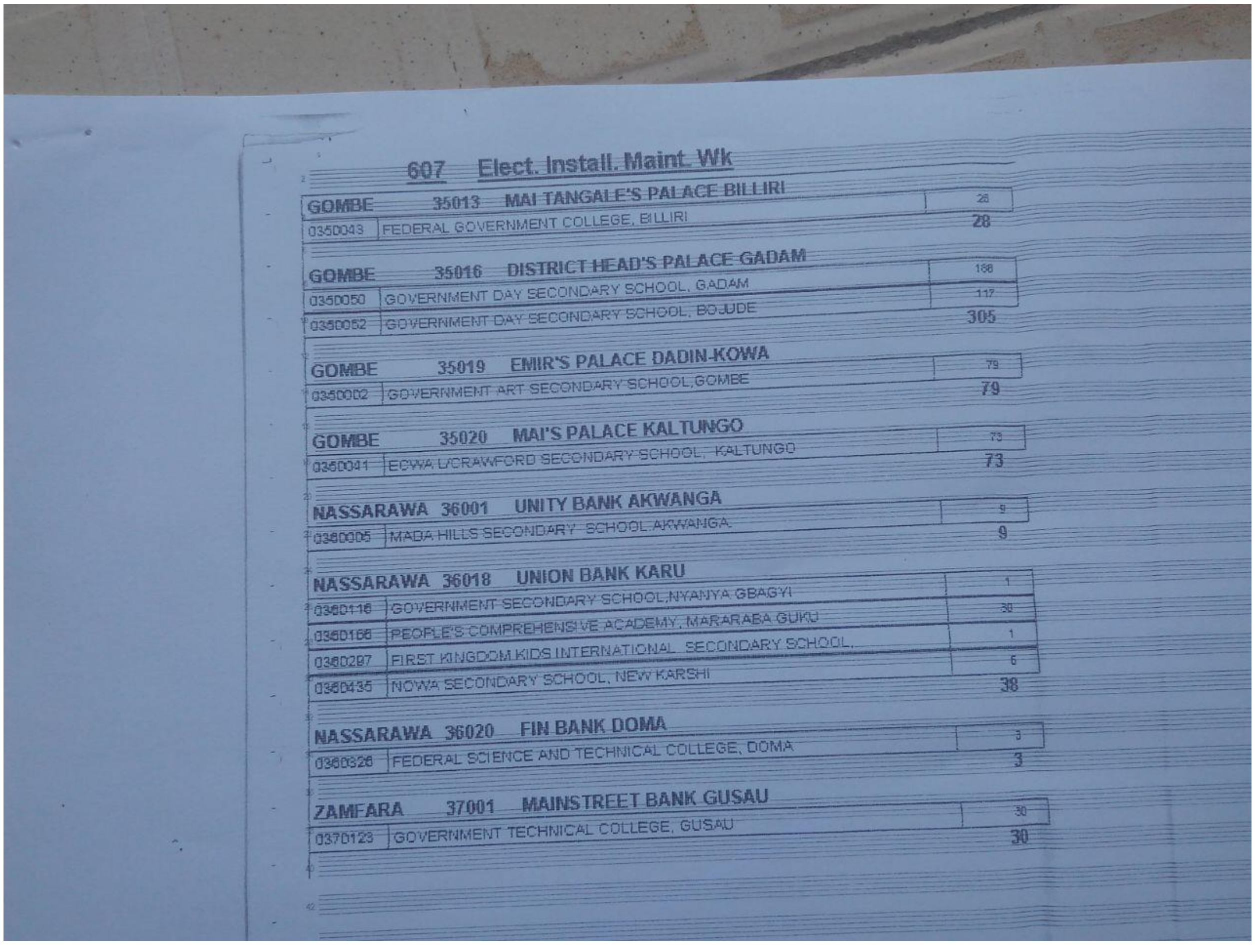
Therefore Dependability coefficient (phi) = $\frac{\sigma^2 p}{\sigma^2 p + \sigma^2\text{Abs}}$

Where $\sigma^2 p = 0.02$ and $\sigma^2\text{Abs} = 0.0030$

$$\text{Then phi} = \frac{0.02}{0.02 + 0.0030} = 0.7800$$

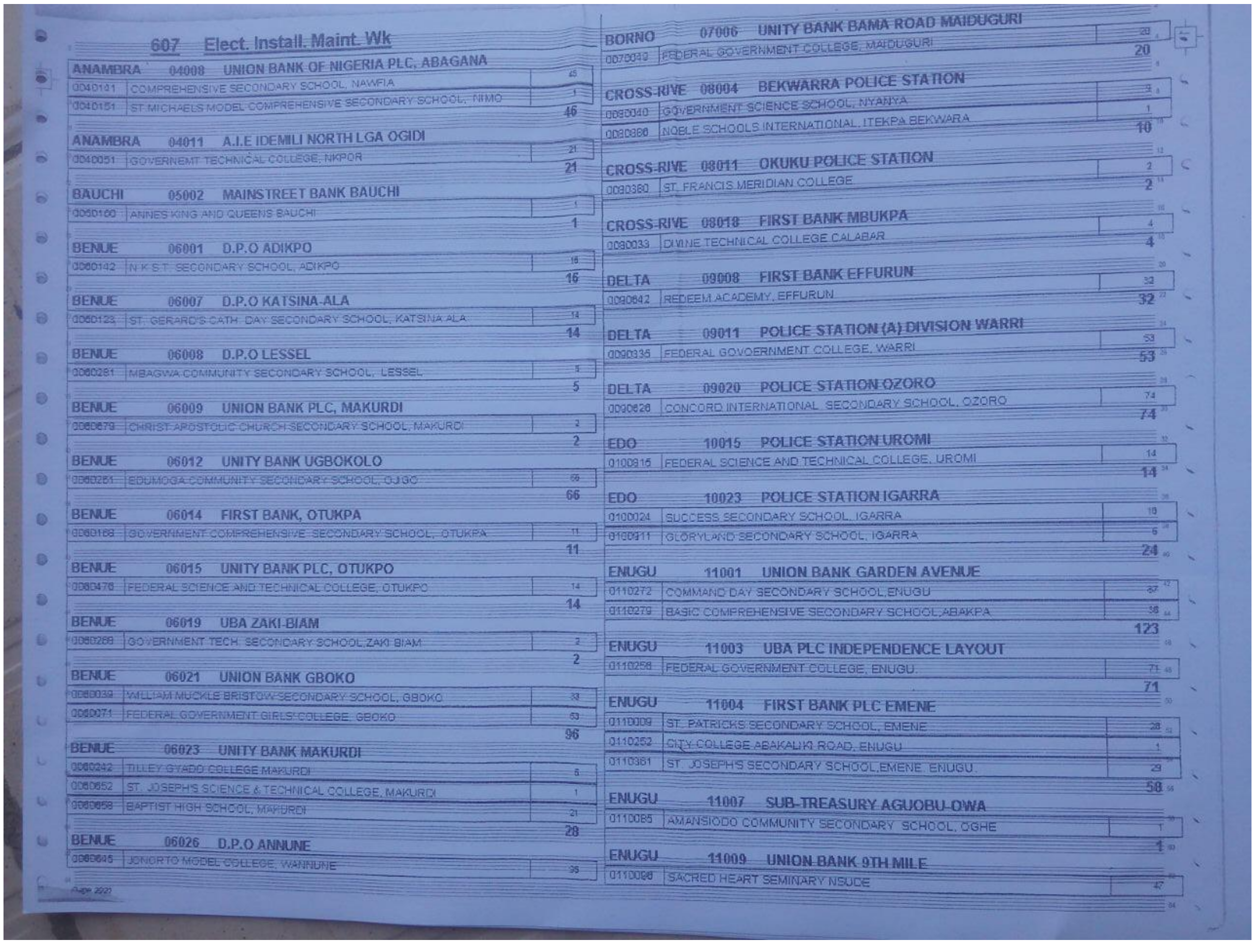
Phi = 0.78

APPENDIX VI
LIST OF THE 1,735 SENIOR SCECONDARY SCHOOLS REGISTRED FOR 2015 NECO ELECTRICAL INSTALLATIONS AND MAINTENANCE WORKS IN NIGERIA



607 Elect. Install. Maint. Wk

GOMBE	35013 MAI TANGALE'S PALACE BILLIRI	28	
0350043	FEDERAL GOVERNMENT COLLEGE, BILLIRI		28
GOMBE	35016 DISTRICT HEAD'S PALACE GADAM	188	
0350050	GOVERNMENT DAY SECONDARY SCHOOL, GADAM	117	
0350052	GOVERNMENT DAY SECONDARY SCHOOL, BOJUDE		305
GOMBE	35019 EMIR'S PALACE DADIN-KOWA	79	
0350002	GOVERNMENT ART SECONDARY SCHOOL,GOMBE		79
GOMBE	35020 MAI'S PALACE KALTUNGO	73	
0350041	ECWA L/CRAWFORD SECONDARY SCHOOL, KALTUNGO		73
NASSARAWA	36001 UNITY BANK AKWANGA	9	
0360005	MADA HILLS SECONDARY SCHOOL AKWANGA		9
NASSARAWA	36018 UNION BANK KARU	1	
0360116	GOVERNMENT SECONDARY SCHOOL,NYANYA GBAGYI	30	
0360166	PEOPLE'S COMPREHENSIVE ACADEMY, MARARABA GUKU	1	
0360297	FIRST KINGDOM KIDS INTERNATIONAL SECONDARY SCHOOL,	6	
0360435	NOWA SECONDARY SCHOOL, NEW KARSHI		38
NASSARAWA	36020 FIN BANK DOMA	3	
0360326	FEDERAL SCIENCE AND TECHNICAL COLLEGE, DOMA		3
ZAMFARA	37001 MAINSTREET BANK GUSAU	30	
0370123	GOVERNMENT TECHNICAL COLLEGE, GUSAU		30



607 Elect. Install. Maint. Wk

ANAMBRA 04008 UNION BANK OF NIGERIA PLC, ABAGANA

0040141	COMPREHENSIVE SECONDARY SCHOOL, NAWFIA	45
0040151	ST MICHAELS MODEL COMPREHENSIVE SECONDARY SCHOOL, NIMO	1
		46

ANAMBRA 04011 A.I.E IDEMILI NORTH LGA OGIDI

0040051	GOVERNEMT TECHNICAL COLLEGE, NKPOR	21
		21

BAUCHI 05002 MAINSTREET BANK BAUCHI

0050160	ANNES KING AND QUEENS BAUCHI	1
		1

BENUE 06001 D.P.O ADIKPO

0060142	N K S T. SECONDARY SCHOOL, ADIKPO	16
		16

BENUE 06007 D.P.O KATSINA-ALA

0060123	ST. GERARD'S CATH. DAY SECONDARY SCHOOL, KATSINA ALA	14
		14

BENUE 06008 D.P.O LESSEL

0060281	MBAGWA COMMUNITY SECONDARY SCHOOL, LESSEL	5
		5

BENUE 06009 UNION BANK PLC, MAKURDI

0060679	CHRIST APOSTOLIC CHURCH SECONDARY SCHOOL, MAKURDI	2
		2

BENUE 06012 UNITY BANK UGBOKOLO

0060261	EDUMOGA COMMUNITY SECONDARY SCHOOL, OJIGO	66
		66

BENUE 06014 FIRST BANK, OTUKPA

0060168	GOVERNMENT COMPREHENSIVE SECONDARY SCHOOL, OTUKPA	11
		11

BENUE 06015 UNITY BANK PLC, OTUKPO

0060476	FEDERAL SCIENCE AND TECHNICAL COLLEGE, OTUKPO	14
		14

BENUE 06019 UBA ZAKI-BIAM

0060268	GOVERNMENT TECH. SECONDARY SCHOOL,ZAKI BIAM	2
		2

BENUE 06021 UNION BANK GBOKO

0060039	WILLIAM MUCKLE BRISTOW SECONDARY SCHOOL, GBOKO	33
0060071	FEDERAL GOVERNMENT GIRLS' COLLEGE, GBOKO	63
		96

BENUE 06023 UNITY BANK MAKURDI

0060242	TILLEY GYADO COLLEGE MAKURDI	6
0060652	ST. JOSEPH'S SCIENCE & TECHNICAL COLLEGE, MAKURDI	1
0060658	BAPTIST HIGH SCHOOL, MAKURDI	21
		28

BENUE 06026 D.P.O ANNUNE

0060645	JONORTO MODEL COLLEGE, WANNUNE	36

Page 2020

BORNO 07006 UNITY BANK BAMA ROAD MAIDUGURI

0070049	FEDERAL GOVERNMENT COLLEGE, MAIDUGURI	20
		20

CROSS-RIVE 08004 BEKWARRA POLICE STATION

0080040	GOVERNMENT SCIENCE SCHOOL, NYANYA	9
0080866	NOBLE SCHOOLS INTERNATIONAL, ITEKPA BEKWARA	1
		10

CROSS-RIVE 08011 OKUKU POLICE STATION

0080380	ST. FRANCIS MERIDIAN COLLEGE	2
		2

CROSS-RIVE 08018 FIRST BANK MBUKPA

0080033	DIVINE TECHNICAL COLLEGE CALABAR	4
		4

DELTA 09008 FIRST BANK EFFURUN

0090642	REDEEM ACADEMY, EFFURUN	32
		32

DELTA 09011 POLICE STATION (A) DIVISION WARRI

0090335	FEDERAL GOVOERNMENT COLLEGE, WARRI	53
		53

DELTA 09020 POLICE STATION OZORO

0090626	CONCORD INTERNATIONAL SECONDARY SCHOOL, OZORO	74
		74

EDO 10015 POLICE STATION UROMI

0100916	FEDERAL SCIENCE AND TECHNICAL COLLEGE, UROMI	14
		14

EDO 10023 POLICE STATION IGARRA

0100024	SUCCESS SECONDARY SCHOOL, IGARRA	18
0100911	GLORYLAND SECONDARY SCHOOL, IGARRA	6
		24

ENUGU 11001 UNION BANK GARDEN AVENUE

0110272	COMMAND DAY SECONDARY SCHOOL,ENUGU	87
0110279	BASIC COMPREHENSIVE SECONDARY SCHOOL,ABAKPA	36
		123

ENUGU 11003 UBA PLC INDEPENDENCE LAYOUT

0110256	FEDERAL GOVERNMENT COLLEGE, ENUGU.	71
		71

ENUGU 11004 FIRST BANK PLC EMENE

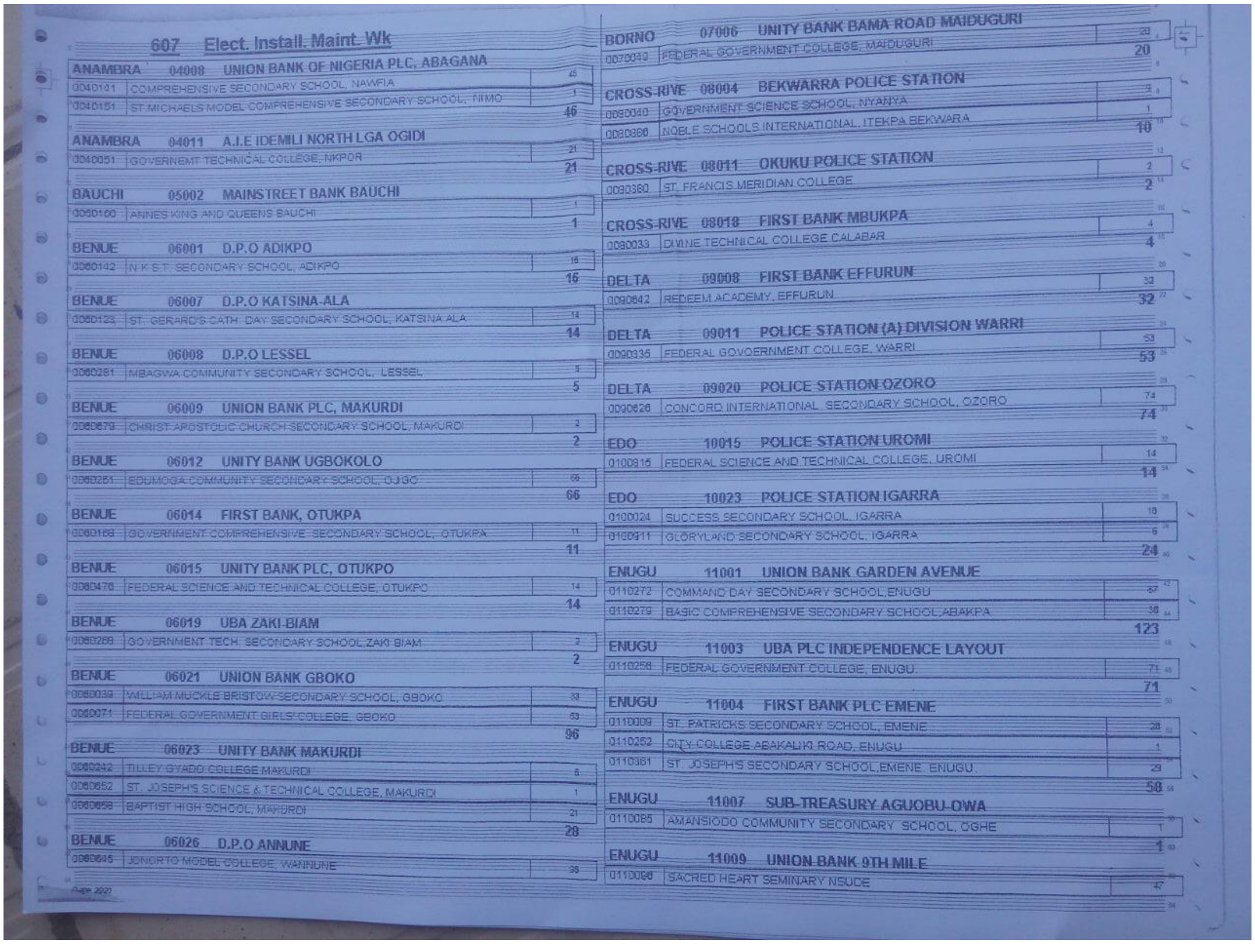
0110009	ST. PATRICKS SECONDARY SCHOOL, EMENE	28
0110252	CITY COLLEGE ABAKALIKI ROAD, ENUGU	1
0110361	ST. JOSEPH'S SECONDARY SCHOOL,EMENE ENUGU.	29
		58

ENUGU 11007 SUB-TREASURY AGUOBU-OWA

0110085	AMANSIODO COMMUNITY SECONDARY SCHOOL, OGHE	1
		1

ENUGU 11009 UNION BANK 9TH MILE

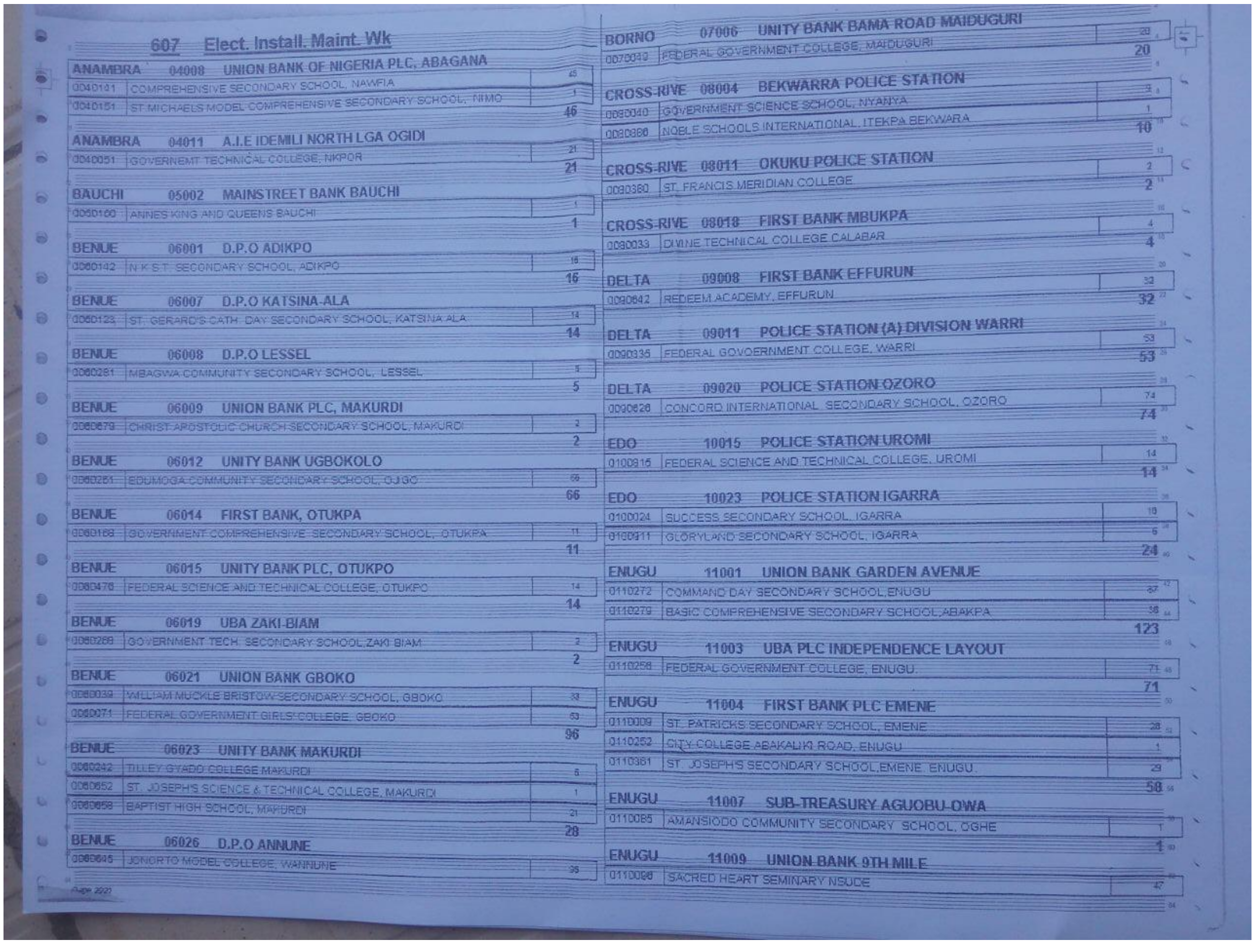
0110096	SACRED HEART SEMINARY NSUDE	47



607 Elect. Install. Maint. Wk

ANAMBRA	04008 UNION BANK OF NIGERIA PLC, ABAGANA	
0040101	COMPREHENSIVE SECONDARY SCHOOL, NAWFIA	45
0040151	ST MICHAELS MODEL COMPREHENSIVE SECONDARY SCHOOL, NIMO	1
		46
ANAMBRA	04011 A.I.E IDEMILI NORTH LGA OGIDI	
0040051	GOVERNEMT TECHNICAL COLLEGE, NKPOR	21
		21
BAUCHI	05002 MAINSTREET BANK BAUCHI	
0050160	ANNES KING AND QUEENS BAUCHI	1
		1
BENUE	06001 D.P.O ADIKPO	
0060142	N K S T SECONDARY SCHOOL, ADIKPO	16
		16
BENUE	06007 D.P.O KATSINA-ALA	
0060123	ST. GERARD'S CATH. DAY SECONDARY SCHOOL, KATSINA ALA	14
		14
BENUE	06008 D.P.O LESSEL	
0060281	MBAGWA COMMUNITY SECONDARY SCHOOL, LESSEL	5
		5
BENUE	06009 UNION BANK PLC, MAKURDI	
0060679	CHRIST APOSTOLIC CHURCH SECONDARY SCHOOL, MAKURDI	2
		2
BENUE	06012 UNITY BANK UGBOKOLO	
0060261	EDUMOGA COMMUNITY SECONDARY SCHOOL, OJGO	66
		66
BENUE	06014 FIRST BANK, OTUKPA	
0060168	GOVERNMENT COMPREHENSIVE SECONDARY SCHOOL, OTUKPA	11
		11
BENUE	06015 UNITY BANK PLC, OTUKPO	
0060476	FEDERAL SCIENCE AND TECHNICAL COLLEGE, OTUKPO	14
		14
BENUE	06019 UBA ZAKI-BIAM	
0060288	GOVERNMENT TECH. SECONDARY SCHOOL,ZAKI BIAM	2
		2
BENUE	06021 UNION BANK GBOKO	
0060039	WILLIAM MUCKLE BRISTOW SECONDARY SCHOOL, GBOKO	33
0060071	FEDERAL GOVERNMENT GIRLS' COLLEGE, GBOKO	63
		96
BENUE	06023 UNITY BANK MAKURDI	
0060242	TILLEY GYADO COLLEGE MAKURDI	6
0060652	ST. JOSEPH'S SCIENCE & TECHNICAL COLLEGE, MAKURDI	1
0060658	BAPTIST HIGH SCHOOL, MAKURDI	21
		28
BENUE	06026 D.P.O ANNUNE	
0060645	JONORTO MODEL COLLEGE, WANNUNE	35
BORNO	07006 UNITY BANK BAMA ROAD MAIDUGURI	
0070049	FEDERAL GOVERNMENT COLLEGE, MAIDUGURI	20
		20
CROSS-RIVE	08004 BEKWARRA POLICE STATION	
0080040	GOVERNMENT SCIENCE SCHOOL, NYANYA	9
0080886	NOBLE SCHOOLS INTERNATIONAL, ITEKPA BEKWARA	1
		10
CROSS-RIVE	08011 OKUKU POLICE STATION	
0080380	ST. FRANCIS MERIDIAN COLLEGE	2
		2
CROSS-RIVE	08018 FIRST BANK MBUKPA	
0080033	DIVINE TECHNICAL COLLEGE CALABAR	4
		4
DELTA	09008 FIRST BANK EFFURUN	
0090642	REDEEM ACADEMY, EFFURUN	32
		32
DELTA	09011 POLICE STATION (A) DIVISION WARRI	
0090335	FEDERAL GOVOERNMENT COLLEGE, WARRI	53
		53
DELTA	09020 POLICE STATION OZORO	
0090626	CONCORD INTERNATIONAL SECONDARY SCHOOL, OZORO	74
		74
EDO	10015 POLICE STATION UROMI	
0100816	FEDERAL SCIENCE AND TECHNICAL COLLEGE, UROMI	14
		14
EDO	10023 POLICE STATION IGARRA	
0100024	SUCCESS SECONDARY SCHOOL, IGARRA	18
0100911	GLORYLAND SECONDARY SCHOOL, IGARRA	6
		24
ENUGU	11001 UNION BANK GARDEN AVENUE	
0110272	COMMAND DAY SECONDARY SCHOOL,ENUGU	87
0110279	BASIC COMPREHENSIVE SECONDARY SCHOOL,ABAKPA	36
		123
ENUGU	11003 UBA PLC INDEPENDENCE LAYOUT	
0110258	FEDERAL GOVERNMENT COLLEGE, ENUGU.	71
		71
ENUGU	11004 FIRST BANK PLC EMENE	
0110009	ST. PATRICKS SECONDARY SCHOOL, EMENE	28
0110252	CITY COLLEGE ABAKALIKI ROAD, ENUGU	1
0110361	ST. JOSEPH'S SECONDARY SCHOOL,EMENE. ENUGU.	29
		58
ENUGU	11007 SUB-TREASURY AGUOBU-OWA	
0110085	AMANSIODO COMMUNITY SECONDARY SCHOOL, OGHE	1
		1
ENUGU	11009 UNION BANK 9TH MILE	
0110096	SACRED HEART SEMINARY NSUDE	47

Page 2020

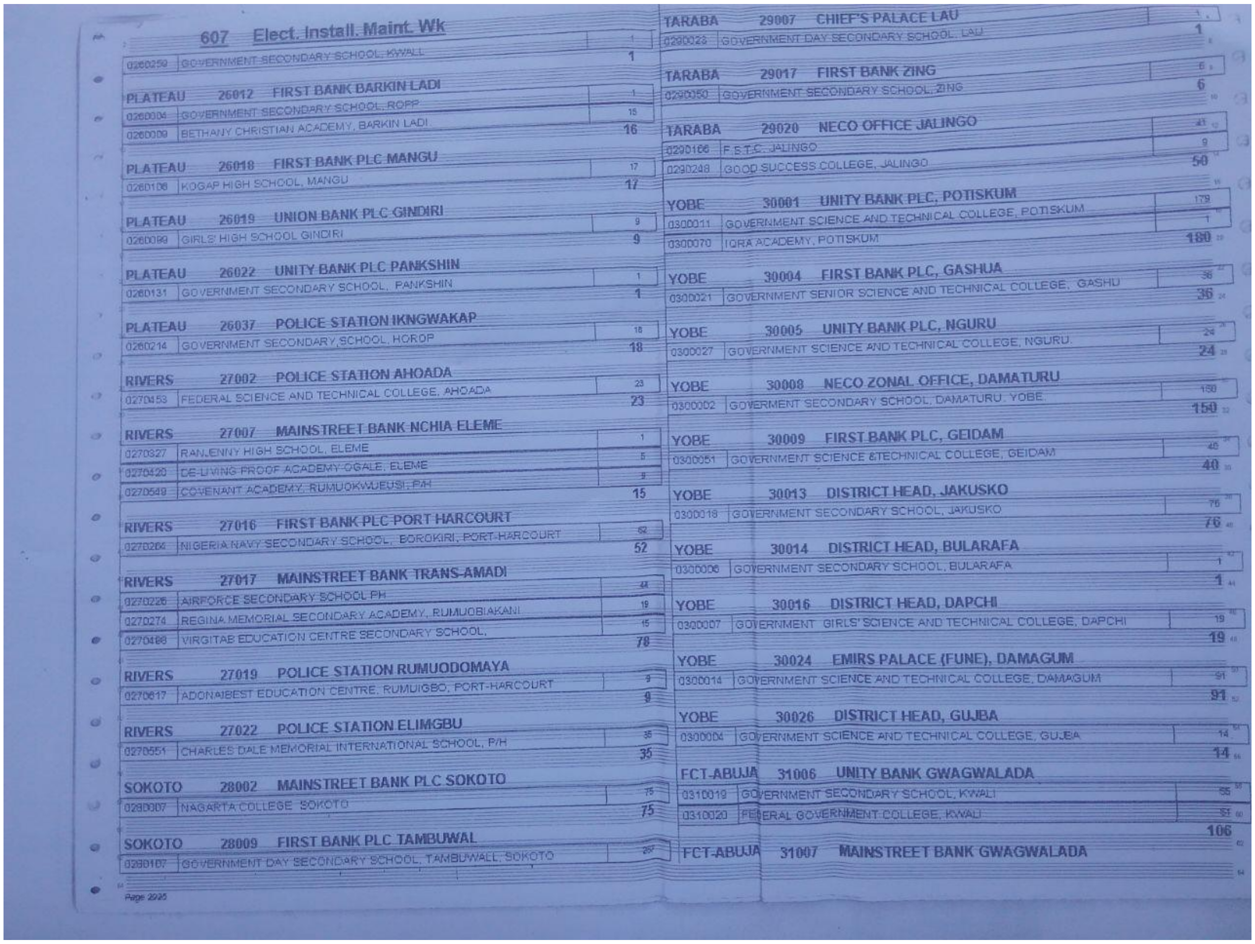


607 Elect. Install. Maint. Wk

ANAMBRA	04008	UNION BANK OF NIGERIA PLC, ABAGANA	
0040141	COMPREHENSIVE SECONDARY SCHOOL, NAWFIA	45	
0040151	ST MICHAELS MODEL COMPREHENSIVE SECONDARY SCHOOL, NIMO	1	
			46
ANAMBRA	04011	A.I.E IDEMILI NORTH LGA OGIDI	
0040051	GOVERNEMT TECHNICAL COLLEGE, NKPOR	21	
			21
BAUCHI	05002	MAINSTREET BANK BAUCHI	
0050160	ANNES KING AND QUEENS BAUCHI	1	
			1
BENUE	06001	D.P.O ADIKPO	
0060142	N K S T. SECONDARY SCHOOL, ADIKPO	16	
			16
BENUE	06007	D.P.O KATSINA-ALA	
0060123	ST. GERARD'S CATH. DAY SECONDARY SCHOOL, KATSINA ALA	14	
			14
BENUE	06008	D.P.O LESSEL	
0060281	MBAGWA COMMUNITY SECONDARY SCHOOL, LESSEL	5	
			5
BENUE	06009	UNION BANK PLC, MAKURDI	
0060679	CHRIST APOSTOLIC CHURCH SECONDARY SCHOOL, MAKURDI	2	
			2
BENUE	06012	UNITY BANK UGBOKOLO	
0060261	EDUMOGA COMMUNITY SECONDARY SCHOOL, OJIGO	66	
			66
BENUE	06014	FIRST BANK, OTUKPA	
0060168	GOVERNMENT COMPREHENSIVE SECONDARY SCHOOL, OTUKPA	11	
			11
BENUE	06015	UNITY BANK PLC, OTUKPO	
0060476	FEDERAL SCIENCE AND TECHNICAL COLLEGE, OTUKPO	14	
			14
BENUE	06019	UBA ZAKI-BIAM	
0060268	GOVERNMENT TECH. SECONDARY SCHOOL,ZAKI BIAM	2	
			2
BENUE	06021	UNION BANK GBOKO	
0060039	WILLIAM MUCKLE BRISTOW SECONDARY SCHOOL, GBOKO	33	
0060071	FEDERAL GOVERNMENT GIRLS' COLLEGE, GBOKO	63	
			96
BENUE	06023	UNITY BANK MAKURDI	
0060242	TILLEY GYADO COLLEGE MAKURDI	6	
0060652	ST. JOSEPH'S SCIENCE & TECHNICAL COLLEGE, MAKURDI	1	
0060658	BAPTIST HIGH SCHOOL, MAKURDI	21	
			28
BENUE	06026	D.P.O ANNUNE	
0060645	JONORTO MODEL COLLEGE, WANNUNE	35	

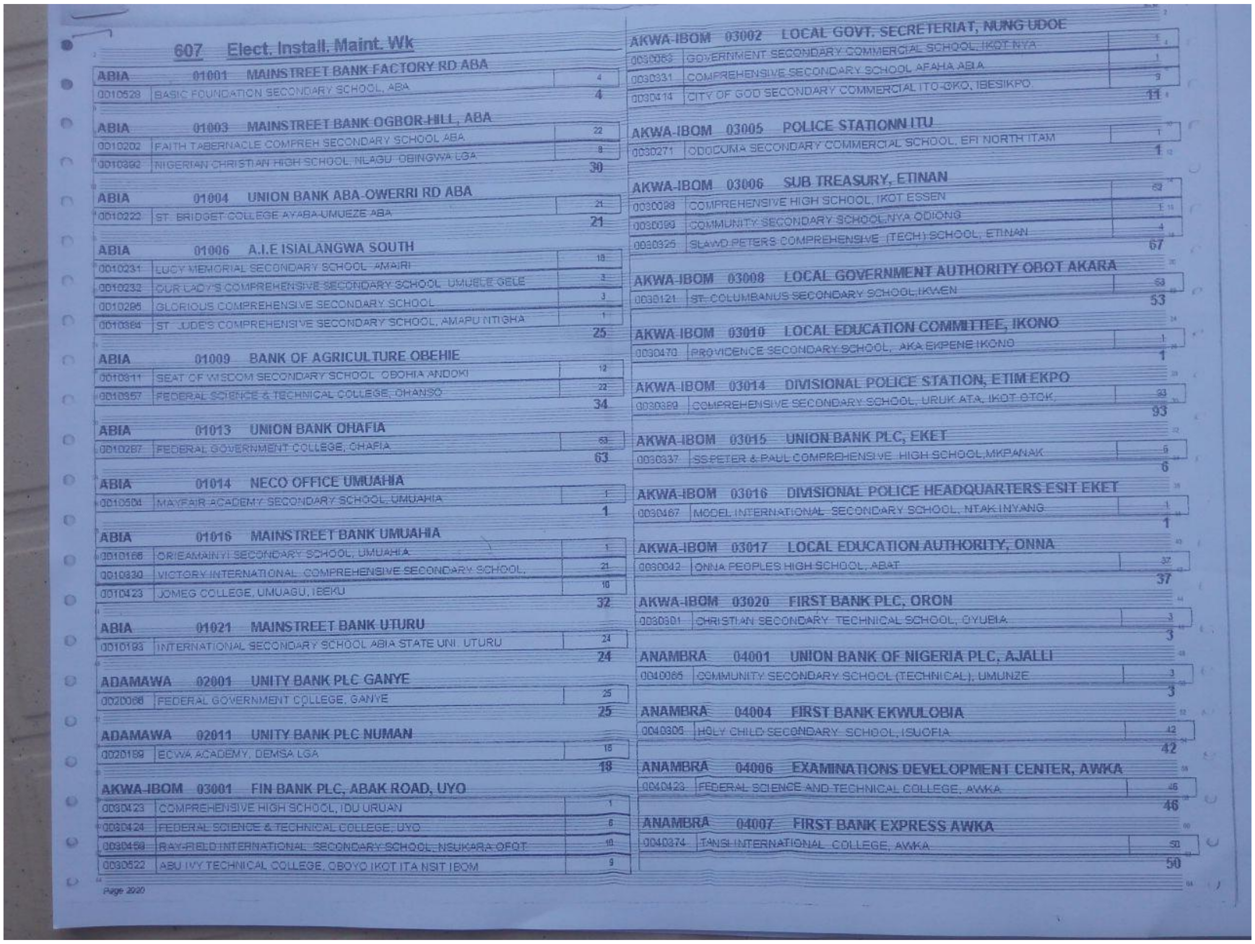
Page 2021

BORNO	07006	UNITY BANK BAMA ROAD MAIDUGURI	
0070049	FEDERAL GOVERNMENT COLLEGE, MAIDUGURI	20	
			20
CROSS-RIVE	08004	BEKWARRA POLICE STATION	
0080040	GOVERNMENT SCIENCE SCHOOL, NYANYA	9	
0080886	NOBLE SCHOOLS INTERNATIONAL, ITEKPA BEKWARA	1	
			10
CROSS-RIVE	08011	OKUKU POLICE STATION	
0080380	ST. FRANCIS MERIDIAN COLLEGE	2	
			2
CROSS-RIVE	08018	FIRST BANK MBUKPA	
0080033	DIVINE TECHNICAL COLLEGE CALABAR	4	
			4
DELTA	09008	FIRST BANK EFFURUN	
0090642	REDEEM ACADEMY, EFFURUN	32	
			32
DELTA	09011	POLICE STATION (A) DIVISION WARRI	
0090335	FEDERAL GOVOERNMENT COLLEGE, WARRI	53	
			53
DELTA	09020	POLICE STATION OZORO	
0090626	CONCORD INTERNATIONAL SECONDARY SCHOOL, OZORO	74	
			74
EDO	10015	POLICE STATION UROMI	
0100916	FEDERAL SCIENCE AND TECHNICAL COLLEGE, UROMI	14	
			14
EDO	10023	POLICE STATION IGARRA	
0100024	SUCCESS SECONDARY SCHOOL, IGARRA	18	
0100911	GLORYLAND SECONDARY SCHOOL, IGARRA	6	
			24
ENUGU	11001	UNION BANK GARDEN AVENUE	
0110272	COMMAND DAY SECONDARY SCHOOL,ENUGU	87	
0110279	BASIC COMPREHENSIVE SECONDARY SCHOOL,ABAKPA	36	
			123
ENUGU	11003	UBA PLC INDEPENDENCE LAYOUT	
0110258	FEDERAL GOVERNMENT COLLEGE, ENUGU.	71	
			71
ENUGU	11004	FIRST BANK PLC EMENE	
0110009	ST. PATRICKS SECONDARY SCHOOL, EMENE	28	
0110252	CITY COLLEGE ABAKALIKI ROAD, ENUGU	1	
0110361	ST. JOSEPH'S SECONDARY SCHOOL,EMENE. ENUGU.	29	
			58
ENUGU	11007	SUB-TREASURY AGUOBU-OWA	
0110085	AMANSIODO COMMUNITY SECONDARY SCHOOL, OGHE	1	
			1
ENUGU	11009	UNION BANK 9TH MILE	
0110098	SACRED HEART SEMINARY NSUDE	47	



607	Elect. Install. Maint. Wk	
0260259	GOVERNMENT SECONDARY SCHOOL, KWALL	1
		1
PLATEAU 26012	FIRST BANK BARKIN LADI	
0260004	GOVERNMENT SECONDARY SCHOOL, ROPP	1
0260009	BETHANY CHRISTIAN ACADEMY, BARKIN LADI	15
		16
PLATEAU 26018	FIRST BANK PLC MANGU	
0260106	KOGAP HIGH SCHOOL, MANGU	17
		17
PLATEAU 26019	UNION BANK PLC GINDIRI	
0260099	GIRLS' HIGH SCHOOL GINDIRI	9
		9
PLATEAU 26022	UNITY BANK PLC PANKSHIN	
0260131	GOVERNMENT SECONDARY SCHOOL, PANKSHIN	1
		1
PLATEAU 26037	POLICE STATION IKNGWAKAP	
0260214	GOVERNMENT SECONDARY,SCHOOL, HOROP	18
		18
RIVERS 27002	POLICE STATION AHOADA	
0270453	FEDERAL SCIENCE AND TECHNICAL COLLEGE, AHOADA	23
		23
RIVERS 27007	MAINSTREET BANK NCHIA ELEME	
0270327	RANJENNY HIGH SCHOOL, ELEME	1
0270420	DE-LIVING PROOF ACADEMY OGALE, ELEME	5
0270549	COVENANT ACADEMY, RUMUOKWUEUSI, P/H	9
		15
RIVERS 27016	FIRST BANK PLC PORT HARCOURT	
0270264	NIGERIA NAVY SECONDARY SCHOOL, BOROKIRI, PORT-HARCOURT	52
		52
RIVERS 27017	MAINSTREET BANK TRANS-AMADI	
0270226	AIRFORCE SECONDARY SCHOOL PH	44
0270274	REGINA MEMORIAL SECONDARY ACADEMY, RUMUOBIAKANI	19
0270488	VIRGITAB EDUCATION CENTRE SECONDARY SCHOOL,	15
		78
RIVERS 27019	POLICE STATION RUMUODOMAYA	
0270617	ADONAIBEST EDUCATION CENTRE, RUMUIGBO, PORT-HARCOURT	9
		9
RIVERS 27022	POLICE STATION ELIMGBU	
0270551	CHARLES DALE MEMORIAL INTERNATIONAL SCHOOL, P/H	35
		35
SOKOTO 28002	MAINSTREET BANK PLC SOKOTO	
0280007	NAGARTA COLLEGE SOKOTO	75
		75
SOKOTO 28009	FIRST BANK PLC TAMBUWAL	
0280107	GOVERNMENT DAY SECONDARY SCHOOL, TAMBUWALL, SOKOTO	257
TARABA 29007	CHIEF'S PALACE LAU	
0290023	GOVERNMENT DAY SECONDARY SCHOOL, LAU	1
		1
TARABA 29017	FIRST BANK ZING	
0290050	GOVERNMENT SECONDARY SCHOOL, ZING	6
		6
TARABA 29020	NECO OFFICE JALINGO	
0290166	F.S.T.C. JALINGO	41
0290248	GOOD SUCCESS COLLEGE, JALINGO	9
		50
YOBE 30001	UNITY BANK PLC, POTISKUM	
0300011	GOVERNMENT SCIENCE AND TECHNICAL COLLEGE, POTISKUM	179
0300070	IQRA ACADEMY, POTISKUM	1
		180
YOBE 30004	FIRST BANK PLC, GASHUA	
0300021	GOVERNMENT SENIOR SCIENCE AND TECHNICAL COLLEGE, GASHU	36
		36
YOBE 30005	UNITY BANK PLC, NGURU	
0300027	GOVERNMENT SCIENCE AND TECHNICAL COLLEGE, NGURU.	24
		24
YOBE 30008	NECO ZONAL OFFICE, DAMATURU	
0300002	GOVERMENT SECONDARY SCHOOL, DAMATURU. YOBE	150
		150
YOBE 30009	FIRST BANK PLC, GEIDAM	
0300051	GOVERNMENT SCIENCE &TECHNICAL COLLEGE, GEIDAM	40
		40
YOBE 30013	DISTRICT HEAD, JAKUSKO	
0300018	GOVERNMENT SECONDARY SCHOOL, JAKUSKO	76
		76
YOBE 30014	DISTRICT HEAD, BULARAFA	
0300006	GOVERNMENT SECONDARY SCHOOL, BULARAFA	1
		1
YOBE 30016	DISTRICT HEAD, DAPCHI	
0300007	GOVERNMENT GIRLS' SCIENCE AND TECHNICAL COLLEGE, DAPCHI	19
		19
YOBE 30024	EMIRS PALACE (FUNE), DAMAGUM	
0300014	GOVERNMENT SCIENCE AND TECHNICAL COLLEGE, DAMAGUM	91
		91
YOBE 30026	DISTRICT HEAD, GUJBA	
0300004	GOVERNMENT SCIENCE AND TECHNICAL COLLEGE, GUJBA	14
		14
FCT-ABUJA 31006	UNITY BANK GWAGWALADA	
0310019	GOVERNMENT SECONDARY SCHOOL, KWALI	55
0310020	FEDERAL GOVERNMENT COLLEGE, KWALI	51
		106
FCT-ABUJA 31007	MAINSTREET BANK GWAGWALADA	

Page 2926



607 Elect. Install. Maint. Wk

ABIA 01001 MAINSTREET BANK FACTORY RD ABA

0010528	BASIC FOUNDATION SECONDARY SCHOOL, ABA	4
		4

ABIA 01003 MAINSTREET BANK OGBOR-HILL, ABA

0010202	FAITH TABERNACLE COMPREH SECONDARY SCHOOL ABA	22
0010392	NIGERIAN CHRISTIAN HIGH SCHOOL, NLAGU OBINGWA LGA	8
		30

ABIA 01004 UNION BANK ABA-OWERRI RD ABA

0010222	ST. BRIDGET COLLEGE AYABA-UMUEZE ABA	21
		21

ABIA 01006 A.I.E ISIALANGWA SOUTH

0010231	LUCY MEMORIAL SECONDARY SCHOOL AMAIRI	18
0010232	OUR LADY'S COMPREHENSIVE SECONDARY SCHOOL UMUELE GELE	3
0010285	GLORIOUS COMPREHENSIVE SECONDARY SCHOOL	3
0010384	ST. JUDE'S COMPREHENSIVE SECONDARY SCHOOL, AMAPU NTIGHA	1
		25

ABIA 01009 BANK OF AGRICULTURE OBEHIE

0010311	SEAT OF WISDOM SECONDARY SCHOOL OBOHIA ANDOKI	12
0010357	FEDERAL SCIENCE & TECHNICAL COLLEGE, OHANSO	22
		34

ABIA 01013 UNION BANK OHAFIA

0010287	FEDERAL GOVERNMENT COLLEGE, OHAFIA	63
		63

ABIA 01014 NECO OFFICE UMUAHIA

0010504	MAYFAIR ACADEMY SECONDARY SCHOOL, UMUAHIA	1
		1

ABIA 01016 MAINSTREET BANK UMUAHIA

0010166	ORIEAMAINYI SECONDARY SCHOOL, UMUAHIA	1
0010330	VICTORY INTERNATIONAL COMPREHENSIVE SECONDARY SCHOOL,	21
0010423	JOMEG COLLEGE, UMUAGU, IBEKU	10
		32

ABIA 01021 MAINSTREET BANK UTURU

0010193	INTERNATIONAL SECONDARY SCHOOL ABIA STATE UNI. UTURU	24
		24

ADAMAWA 02001 UNITY BANK PLC GANYE

0020068	FEDERAL GOVERNMENT COLLEGE, GANYE	25
		25

ADAMAWA 02011 UNITY BANK PLC NUMAN

0020159	ECWA ACADEMY, DEMSA LGA	18
		18

AKWA-IBOM 03001 FIN BANK PLC, ABAK ROAD, UYO

0030423	COMPREHENSIVE HIGH SCHOOL, IDU URUAN	1
0030424	FEDERAL SCIENCE & TECHNICAL COLLEGE, UYO	6
0030458	RAY-FIELD INTERNATIONAL SECONDARY SCHOOL, NSUKARA OFOT	10
0030522	ABU IVY TECHNICAL COLLEGE, OBOYO IKOT ITA NSIT IBOM	9

Page 2020

AKWA-IBOM 03002 LOCAL GOVT. SECRETERIAT, NUNG UDOE

0030063	GOVERNMENT SECONDARY COMMERCIAL SCHOOL, IKOT NYA	1
0030331	COMPREHENSIVE SECONDARY SCHOOL AFAHA ABIA	1
0030414	CITY OF GOD SECONDARY COMMERCIAL ITO-OKO, IBESIKPO.	9
		11

AKWA-IBOM 03005 POLICE STATIONN ITU

0030271	ODODUMA SECONDARY COMMERCIAL SCHOOL, EFI NORTH ITAM	1
		1

AKWA-IBOM 03006 SUB TREASURY, ETINAN

0030098	COMPREHENSIVE HIGH SCHOOL, IKOT ESSEN	62
0030099	COMMUNITY SECONDARY SCHOOL,NYA ODIONG	1
0030325	SLAWD PETERS COMPREHENSIVE (TECH) SCHOOL, ETINAN	4
		67

AKWA-IBOM 03008 LOCAL GOVERNMENT AUTHORITY OBOT AKARA

0030121	ST. COLUMBANUS SECONDARY SCHOOL,IKWEN	53
		53

AKWA-IBOM 03010 LOCAL EDUCATION COMMITTEE, IKONO

0030470	PROVICENCE SECONDARY SCHOOL, AKA EKPENE IKONO	1
		1

AKWA-IBOM 03014 DIVISIONAL POLICE STATION, ETIM EKPO

0030389	COMPREHENSIVE SECONDARY SCHOOL, URUK ATA, IKOT OTOK,	93
		93

AKWA-IBOM 03015 UNION BANK PLC, EKET

0030337	SS PETER & PAUL COMPREHENSIVE HIGH SCHOOL,MKPANAK	6
		6

AKWA-IBOM 03016 DIVISIONAL POLICE HEADQUARTERS ESIT EKET

0030467	MODEL INTERNATIONAL SECONDARY SCHOOL, NTAK INYANG	1
		1

AKWA-IBOM 03017 LOCAL EDUCATION AUTHORITY, ONNA

0030042	ONNA PEOPLES HIGH SCHOOL, ABAT	37
		37

AKWA-IBOM 03020 FIRST BANK PLC, ORON

0030301	CHRISTIAN SECONDARY TECHNICAL SCHOOL, OYUBIA	3
		3

ANAMBRA 04001 UNION BANK OF NIGERIA PLC, AJALLI

0040065	COMMUNITY SECONDARY SCHOOL (TECHNICAL), UMUNZE	3
		3

ANAMBRA 04004 FIRST BANK EKWULOBIA

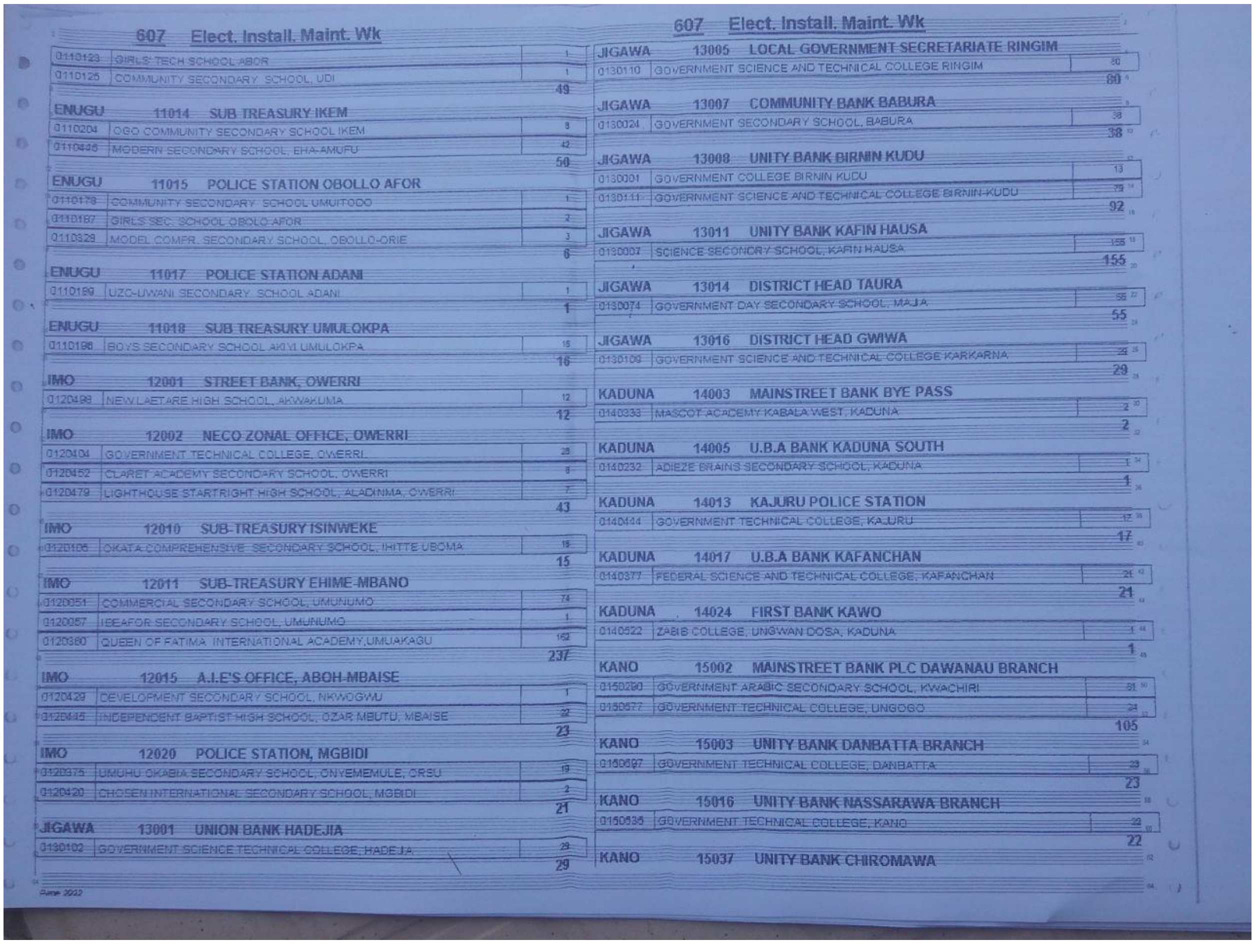
0040306	HOLY CHILD SECONDARY SCHOOL, ISUOFIA	42
		42

ANAMBRA 04006 EXAMINATIONS DEVELOPMENT CENTER, AWKA

0040423	FEDERAL SCIENCE AND TECHNICAL COLLEGE, AWKA	46
		46

ANAMBRA 04007 FIRST BANK EXPRESS AWKA

0040374	TANSI INTERNATIONAL COLLEGE, AWKA	50
		50



607 Elect. Install. Maint. Wk

0110123	GIRLS' TECH SCHOOL ABOR	1
0110125	COMMUNITY SECONDARY SCHOOL, UDI	1
		49

ENUGU 11014 SUB TREASURY IKEM

0110204	OGO COMMUNITY SECONDARY SCHOOL IKEM	8
0110446	MODERN SECONDARY SCHOOL, EHA-AMUFU	42
		50

ENUGU 11015 POLICE STATION OBOLLO AFOR

0110178	COMMUNITY SECONDARY SCHOOL UMUITODO	1
0110187	GIRLS SEC. SCHOOL OBOLO-AFOR	2
0110328	MODEL COMPR. SECONDARY SCHOOL, OBOLLO-ORIE	3
		6

ENUGU 11017 POLICE STATION ADANI

0110189	UZO-UWANI SECONDARY SCHOOL ADANI	1
		1

ENUGU 11018 SUB TREASURY UMULOKPA

0110198	BOYS SECONDARY SCHOOL AKWI UMULOKPA	16
		16

IMO 12001 STREET BANK, OWERRI

0120498	NEW LAETARE HIGH SCHOOL, AKWAKUMA	12
		12

IMO 12002 NECO ZONAL OFFICE, OWERRI

0120404	GOVERNMENT TECHNICAL COLLEGE, OWERRI	28
0120452	CLARET ACADEMY SECONDARY SCHOOL, OWERRI	8
0120479	LIGHTHOUSE STARTRIGHT HIGH SCHOOL, ALADINMA, OWERRI	7
		43

IMO 12010 SUB-TREASURY ISINWEKE

0120106	OKATA COMPREHENSIVE SECONDARY SCHOOL, IHITTE UBOMA	15
		15

IMO 12011 SUB-TREASURY EHIME-MBANO

0120051	COMMERCIAL SECONDARY SCHOOL, UMUNUMO	74
0120057	IBEAFOR SECONDARY SCHOOL, UMUNUMO	1
0120380	QUEEN OF FATIMA INTERNATIONAL ACADEMY,UMUAKAGU	162
		237

IMO 12015 A.I.E'S OFFICE, ABOH-MBAISE

0120429	DEVELOPMENT SECONDARY SCHOOL, NKWOGWU	1
0120445	INDEPENDENT BAPTIST HIGH SCHOOL, OZAR MBUTU, MBAISE	22
		23

IMO 12020 POLICE STATION, MGBIDI

0120375	UMUHU OKABIA SECONDARY SCHOOL, ONYEMEMULE, ORSU	19
0120420	CHOSEN INTERNATIONAL SECONDARY SCHOOL, MGBIDI	2
		21

JIGAWA 13001 UNION BANK HADEJIA

0130102	GOVERNMENT SCIENCE TECHNICAL COLLEGE, HADEJIA	29
		29

June 2022

607 Elect. Install. Maint. Wk

JIGAWA 13005 LOCAL GOVERNMENT SECRETARIATE RINGIM

0130110	GOVERNMENT SCIENCE AND TECHNICAL COLLEGE RINGIM	80
		80

JIGAWA 13007 COMMUNITY BANK BABURA

0130024	GOVERNMENT SECONDARY SCHOOL, BABURA	38
		38

JIGAWA 13008 UNITY BANK BIRNIN KUDU

0130001	GOVERNMENT COLLEGE BIRNIN KUDU	13
0130111	GOVERNMENT SCIENCE AND TECHNICAL COLLEGE BIRNIN-KUDU	79
		92

JIGAWA 13011 UNITY BANK KAFIN HAUSA

0130007	SCIENCE SECONDRY SCHOOL, KAFIN HAUSA	155
		155

JIGAWA 13014 DISTRICT HEAD TAURA

0130074	GOVERNMENT DAY SECONDARY SCHOOL, MAJA	55
		55

JIGAWA 13016 DISTRICT HEAD GWIWA

0130109	GOVERNMENT SCIENCE AND TECHNICAL COLLEGE KARKARNA	29
		29

KADUNA 14003 MAINSTREET BANK BYE PASS

0140333	MASCOT ACADEMY KABALA WEST, KADUNA	2
		2

KADUNA 14005 U.B.A BANK KADUNA SOUTH

0140232	ADIEZE BRAINS SECONDARY SCHOOL, KADUNA	1
		1

KADUNA 14013 KAJURU POLICE STATION

0140444	GOVERNMENT TECHNICAL COLLEGE, KAJURU	17
		17

KADUNA 14017 U.B.A BANK KAFANCHAN

0140377	FEDERAL SCIENCE AND TECHNICAL COLLEGE, KAFANCHAN	21
		21

KADUNA 14024 FIRST BANK KAWO

0140522	ZABIB COLLEGE, UNGWAN DOSA, KADUNA	1
		1

KANO 15002 MAINSTREET BANK PLC DAWANAU BRANCH

0150290	GOVERNMENT ARABIC SECONDARY SCHOOL, KWACHIRI	81
0150577	GOVERNMENT TECHNICAL COLLEGE, UNGOGO	24
		105

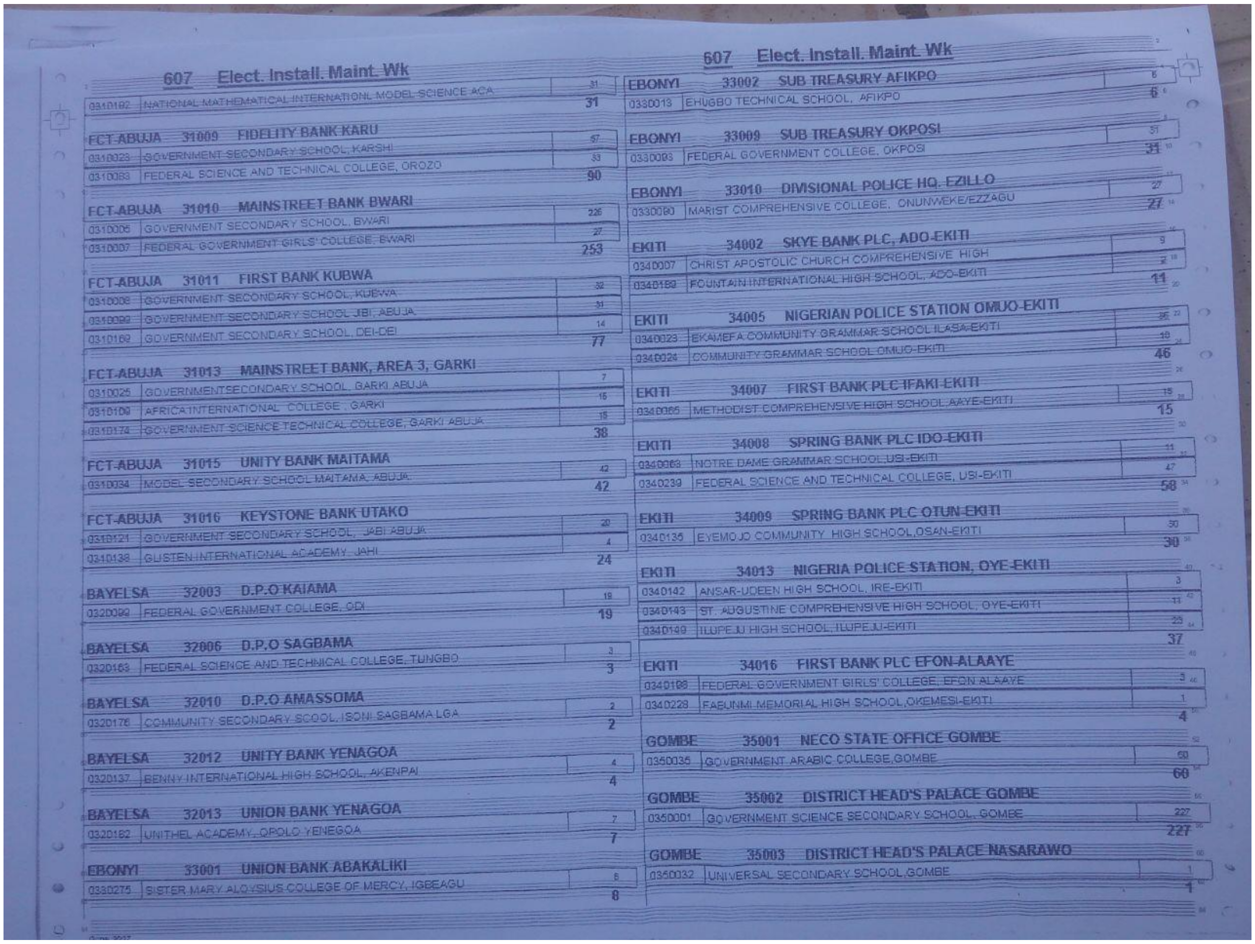
KANO 15003 UNITY BANK DANBATTA BRANCH

0150697	GOVERNMENT TECHNICAL COLLEGE, DANBATTA	23
		23

KANO 15016 UNITY BANK NASSARAWA BRANCH

0150535	GOVERNMENT TECHNICAL COLLEGE, KANO	22
		22

KANO 15037 UNITY BANK CHIROMAWA



607 Elect. Install. Maint. Wk

0310182	NATIONAL MATHEMATICAL INTERNATIONL MODEL SCIENCE ACA	31
		31
FCT-ABUJA 31009	**FIDELITY BANK KARU**	
0310023	GOVERNMENT SECONDARY SCHOOL, KARSHI	57
0310083	FEDERAL SCIENCE AND TECHNICAL COLLEGE, OROZO	33
		90
FCT-ABUJA 31010	**MAINSTREET BANK BWARI**	
0310005	GOVERNMENT SECONDARY SCHOOL, BWARI	226
0310007	FEDERAL GOVERNMENT GIRLS' COLLEGE, BWARI	27
		253
FCT-ABUJA 31011	**FIRST BANK KUBWA**	
0310008	GOVERNMENT SECONDARY SCHOOL, KUBWA	32
0310099	GOVERNMENT SECONDARY SCHOOL JBI, ABUJA	31
0310169	GOVERNMENT SECONDARY SCHOOL, DEI-DEI	14
		77
FCT-ABUJA 31013	**MAINSTREET BANK, AREA 3, GARKI**	
0310025	GOVERNMENTSECONDARY SCHOOL, GARKI ABUJA	7
0310109	AFRICAINTERNATIONAL COLLEGE , GARKI	15
0310174	GOVERNMENT SCIENCE TECHNICAL COLLEGE, GARKI ABUJA	16
		38
FCT-ABUJA 31015	**UNITY BANK MAITAMA**	
0310034	MODEL SECONDARY SCHOOL MAITAMA, ABUJA	42
		42
FCT-ABUJA 31016	**KEYSTONE BANK UTAKO**	
0310121	GOVERNMENT SECONDARY SCHOOL, JABI ABUJA	20
0310138	GUSTEN INTERNATIONAL ACADEMY, JAHI	4
		24
BAYELSA 32003	**D.P.O KAIAMA**	
0320099	FEDERAL GOVERNMENT COLLEGE, ODI	19
		19
BAYELSA 32006	**D.P.O SAGBAMA**	
0320163	FEDERAL SCIENCE AND TECHNICAL COLLEGE, TUNGBO	3
		3
BAYELSA 32010	**D.P.O AMASSOMA**	
0320176	COMMUNITY SECONDARY SCOOL, ISONI SAGBAMA LGA	2
		2
BAYELSA 32012	**UNITY BANK YENAGOA**	
0320137	BENNY INTERNATIONAL HIGH SCHOOL, AKENPAI	4
		4
BAYELSA 32013	**UNION BANK YENAGOA**	
0320182	UNITHEL ACADEMY, OPOLO YENEGOA	7
		7
EBONYI 33001	**UNION BANK ABAKALIKI**	
0330275	SISTER MARY ALOYSIUS COLLEGE OF MERCY, IGBEAGU	8
		8

607 Elect. Install. Maint. Wk

EBONYI 33002	**SUB TREASURY AFIKPO**	
0330013	EHUGBO TECHNICAL SCHOOL, AFIKPO	6
		6
EBONYI 33009	**SUB TREASURY OKPOSI**	
0330093	FEDERAL GOVERNMENT COLLEGE, OKPOSI	31
		31
EBONYI 33010	**DIVISIONAL POLICE HQ. EZILLO**	
0330080	MARIST COMPREHENSIVE COLLEGE, ONUNWEKE/EZZAGU	27
		27
EKITI 34002	**SKYE BANK PLC, ADO-EKITI**	
0340007	CHRIST APOSTOLIC CHURCH COMPREHENSIVE HIGH	9
0340189	FOUNTAIN INTERNATIONAL HIGH SCHOOL, ADO-EKITI	2
		11
EKITI 34005	**NIGERIAN POLICE STATION OMUO-EKITI**	
0340023	EKAMEFA COMMUNITY GRAMMAR SCHOOL ILASA-EKITI	36
0340024	COMMUNITY GRAMMAR SCHOOL OMUO-EKITI	10
		46
EKITI 34007	**FIRST BANK PLC IFAKI-EKITI**	
0340065	METHODIST COMPREHENSIVE HIGH SCHOOL,AAYE-EKITI	15
		15
EKITI 34008	**SPRING BANK PLC IDO-EKITI**	
0340063	NOTRE DAME GRAMMAR SCHOOL,USI-EKITI	11
0340239	FEDERAL SCIENCE AND TECHNICAL COLLEGE, USI-EKITI	47
		58
EKITI 34009	**SPRING BANK PLC OTUN-EKITI**	
0340135	EYEMOJO COMMUNITY HIGH SCHOOL,OSAN-EKITI	30
		30
EKITI 34013	**NIGERIA POLICE STATION, OYE-EKITI**	
0340142	ANSAR-UDEEN HIGH SCHOOL, IRE-EKITI	3
0340143	ST. AUGUSTINE COMPREHENSIVE HIGH SCHOOL, OYE-EKITI	11
0340149	ILUPEJU HIGH SCHOOL, ILUPEJU-EKITI	23
		37
EKITI 34016	**FIRST BANK PLC EFON-ALAAYE**	
0340198	FEDERAL GOVERNMENT GIRLS' COLLEGE, EFON ALAAYE	3
0340228	FABUNMI MEMORIAL HIGH SCHOOL,OKEMESI-EKITI	1
		4
GOMBE 35001	**NECO STATE OFFICE GOMBE**	
0350035	GOVERNMENT ARABIC COLLEGE,GOMBE	60
		60
GOMBE 35002	**DISTRICT HEAD'S PALACE GOMBE**	
0350001	GOVERNMENT SCIENCE SECONDARY SCHOOL, GOMBE	227
		227
GOMBE 35003	**DISTRICT HEAD'S PALACE NASARAWO**	
0350032	UNIVERSAL SECONDARY SCHOOL,GOMBE	1
		1

APPENDIX VIII

NATIONAL EXAMINATION COUNCIL ADOPTED ELECTRICAL INSTALLATION AND MAINTENANCE ACHIEVEMENT TEST (NECOAEIMAT)

1. A township distribution network (TDN) an …………………. Tension distribution A. extra low B. extra high C. high D. low E. medium

2. The following are causes of accident in an electrical industry, except

A. careless in work B. climbing pole with no belt C. ignorance of the work

D. over speeding a machine E. using draw –vice for stringing

3. Photometric bench is an instrument used in measuring

A. heat formation B. humidity C. luminous intensity D. magnetic flux E. speed of light

4. In electrical installation, cable, H.S.O.S means

A. head of service and overhead source B. head of state overhead service

C. heading for service in overhead system D. houses service overhead system E. housing system overhead service

5. The following are overhead line support except ……………. Insulator,

A. disc B. pin C. shackle D. stay E. wooden

6. Which of these provide protection against an electrical fault in a house?

A. A/C plug B. circuit breaker C. energy meter D. extension box E. socket

7. Which of these regulations is not tenable in the court of law?

A. 1999 constitution of Nigerian B. factory act 1908 C. factory act 1944 D. I.E.E regulation E. penal code Nigeria

8. Which of the following is not a type of cell

A. alkaline B. incandescent C. lead acid D. leclanache E. mercury

9. The suitable ampere rating of stab-lock in a distribution board for bathroom water heater is

A. 1A, B. 5A. C. 10A. D. 15A. E. 30A.

10. Which of these is an electrical appliance

A. blender B. cable C. hammer D. gimlet E. socket outlet

11. A prepaid KWH meter is an example of a/an …………………………… meter

A. analogue B. computer C. digital D. resistive E. semi-conductor

12. Any apparatus with an exposed metal work should be ………….. to reduce risk of shock

A. earthed B. galvanized C. insulated D. painted E. sheradized

13. Which of these is not a property of an insulation

A. high resistivity to the flow of electric current B. low resistance to the flow of current C. resistance to corrosive effect D. withstanding high temperature E. withstanding mechanical stress

14. The maintenance area of a buried earth rod can be applying

A. coal, pepper and coke B. engine oil, salt and coke C. fuel, coal, and salt D. pepper, salt and coal E. salt, coke and coal

15. I.E.E. regulation disallows the installation --------in a bathroom.

A. cable B. clips C. conduit D. lighting fining E. socket cutlet

16. Which of these is not an electrical protective device?

A. scatridge fust breaker C. dimmer D. relay E.rewire able fuse

17. A typical consumer control sequence comprises the following except

A. circuit breaker B. cut-one-fuse C. distribution board D. energy meter E. feeder pillar

18. Which of the following factors is not to be proposing a wiring system?

A. cost of materials B. durability Celectrical part seller's shop D. suitability E.time available

19. which of the following cables is most suitable for filling station wiring?

A. flexible cord B. mineral insulated copper sheated (C)PVC (poly-vingl-chloride)

D. T.R.S (tough rubber sheated) E. vulcanized rubber insulated

20. A bell transformer delivers 12V from a supply of 240V, if the input is at 20A. What is the output current at bell terminal?

A. 0.6A B.1A C. 2A D. 4A E.12A

21. One effective way of providing artificial respiration to an electrocuted person is--------- method

A. mouth to heart B. mouth to mouth C. mouth to nose D. nose to eye E. staring at eye

22. When the flux is used in a soldering work, its function is to

A. align the joint B. sellotape the joint C. elongate melting point of solders

D. protect the joint E. remove dirt from the joint

23. which of these commissions ministries register`s business enterprises in Nigeria? A. corporate affair commission B. due process commission C.ministry of commission D. ministry of labour E. national population commission.

24. which of the following is not a tool for electrical work?

A. hammer B. plier C. scissors D. side cutter E. screw driver

25.The following tool/materials are used in electrical soldering work, except

A. driller B. electrode C. lead E. plier (D)soldering iron.

26. circuit breaker is rated according to its

A. consumer control panel. B. current carrying capacity C. final sub-circuit capacity D. resistance capacity

27. the following instrument are for measuring electrical quantity, except

A. ammeter B. hydro-meter C. ohm-meter D. voltmeter E. watthour-meter

28. stroboscopic effect is a defect associated with

A. discharge lighting B. four plate cooker C. incandescent lamp D.ring boiler

E. surface heater

29.an electrical merchant who wish to ship his consignment from Europe, should receive the goods in Nigeria at

A. Apapa port Lagos B. Ido rail terminus Lagos C. Kure modern market Minna

D. Maiduguri motor park (E) Tejuosho market Yaba

30. The two main types of fused used in electrical wiring are (A)capacitive and resistive B. catridge and capacitive C. catridge rewireable E. rewireable and resistive

31. The following are some of the safety rules to be observed in a workshop except

A. Allowing the workshop floor to be slippery B. keeping tools in the locker after use C. not running around the work shop D. safe work in habits (E)selecting the right tools for the job

32. Which of this is not an electrical appliance?

A. air conditioner B. electric blender   C. grinding machine D. pumping machine E. tumbler switch

33.which of these times is suitable for an off-penk triff billing?

A. 12 midnight B. 8a.m C. 12noon D. 1p.m E. 8p.m

34.which of these materials is used on a high tension over head line?

A. D-iron B. disc insulator C .i.t pole (D)sheckle expansion (E)sheckle insulator

35. An immersion heater of 2KV is connected to a 230V supply. Calculate the current flowing in the heater.

A. 2.4A B. 3.0A C. 8.7A D. 4.5A E. 9.2A

36. The I.EE regulation permits the following in a bathroom except

A. all insulated ceiling shaver unit B. all insulated emf fixing

c.ceiling switch D. earthed portable ceiling fan E. shaver unit

37. Which of the following is not found in a distribution sub-station

A. bus- bar B. feeder pillar C insulator D transformer E. wall bracket

38. To prepare an estimate for electrical work, the following factors are considered, Except.

A. description of item B. Quantity of item required C. the colour of each item

D. total cost of the work E. unit price of item

39. The primary turns of a step-up transformer are 100 and the primary voltage is 240V, if the secondary voltage is 240V, the secondary winding will be ……… Turns

A. 4.16 B. 10 C. 24 D. 100 C. 1000

40. The diagram shown below is that of a ……………. generator

A. Compound
ammerture
B. excited
C. parallel
D. series
E. shunt