

**COURSE
GUIDE**

**EDU 921
ADVANCED EDUCATIONAL RESEARCH METHODS**

Course Team Prof. Ganiyu A. Badmus, Ph.D. (Course
Developer/Editor) - NOUN
Dr. C. A. Okonkwo (Programme Leader/Course
Coordinator) - NOUN
Prof. Nduka Okoh, Ph.D. (Course Writer) – Faculty
of Education University of Benin



NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
14/16 Ahmadu Bello Way
Victoria Island
Lagos

Abuja Office
5, Dar es Salaam Street
Off Aminu Kano Crescent
Wuse II, Abuja
Nigeria

e-mail: centralinfo@noun.edu.ng

URL: www.noun.edu.ng

Published by
National Open University of Nigeria

Printed 2012

ISBN:

All Rights Reserved

Printed by

CONTENTS	PAGE
Introduction.....	iv
What You Will Learn in This Course.....	iv
Course Aims.....	v
Course Objectives.....	v
Working Through This Course.....	vi
The Course Materials.....	vi
Study Units.....	vii
Assignment File.....	viii
Presentation Schedule.....	viii
Assessments.....	viii
Tutor-Marked Assignments (TMAs).....	viii
Final Examination and Grading.....	ix
Course Marking Scheme.....	ix
Course Overview.....	ix
How to Get the Most from This Course.....	x
Tutors and Tutorials.....	xii
References/Further Reading.....	xiii
Summary.....	xiv

INTRODUCTION

EDU 801 Advanced Educational Research Methods is a three- credit unit course developed for the Ph.D. degree programme in Education of the National Open University of Nigeria (NOUN). The course consists of six modules in a total of 21 units, plus a Course Guide designed to introduce you to the course material and how to use it. The Course Guide will also be useful to you as a reference tool to consult if and when you have any questions about the course, e.g. how to plan your time for studying the course, when to submit assessments, and the support system available to you.

The course itself, as part of the doctoral programme, assumes and builds upon previous knowledge of and exposure to Educational Research Methods at the Masters degree level and even possibly also at the undergraduate degree level where applicable. Basic computer literacy is required of candidates taking this course, in order to be able to source relevant data from the internet and other electronic reference accessories such as the CD-ROM. This is of course in addition to the efficient use of the traditional resources of the modern standard library in terms of research textbooks, journals, special collections, etc.

On successful completion of the course, it is hoped that you would have been well grounded in advanced research methods in education, and be fully competent and confident to carry out high quality research work in and outside university tertiary level education.

That it will require hard work and dedication to stay the course is obvious, but the great thing is that if you persevere to the end, the academic and intellectual rewards will be overwhelming. This is what NOUN offers you in this course, and I am sure you are up to it. Welcome.

WHAT YOU WILL LEARN IN THIS COURSE

Meaningful research is central and critical to the science of education, indeed to the science of life generally. In this course therefore we have covered, and you will learn about, the nature of educational research and its ramifications in terms of its various types and procedures, how to formulate research questions and hypotheses, how to source for information, and the major elements of research designs. Proper reviews of literature, as well as sampling, were also treated. Additionally, we have given information about instrumentation, and the main techniques for processing data.

It is hoped that you will have been equipped from this course to read, understand, and apply research results from various research reports, and carry out valid educational research of your own.

COURSE AIMS

The course aims to provide you with a comprehensive, advanced understanding and appreciation of what quality educational research is and how to get about it.

COURSE OBJECTIVES

To actualize the above stated aims, the course has well-defined set of specific objectives at the beginning of each unit. These objectives are meant for you to read and internalise them before you study the unit. You will need to refer to them during your study of the unit to check on your progress. You should also look back at the objectives after completing a unit; this will assist you in ascertaining your level of attainment of the stated objectives of the unit.

Consequently, the overall objectives of the course are given below. By attaining these objectives, you should have attained the aims of the course as a whole. Thus, on successful completion of this course, you should be able to:

- list the basic procedures for conducting educational research
- explain the role of educational research for educational practice
- state the interrelationships between the various classifications of research
- know how to use the computer to source for educational information
- describe historical research with examples
- evaluate critical theory as a paradigm of educational research
- examine the different sampling methods used in data collection for surveys
- list the threats to internal and external validity of a survey instrument
- explain how “correlation does not imply causation”
- comment critically on the differences between causal-comparative and correlation research
- differentiate the several types of quasi-experimental design
- describe a simple experiment
- outline the advantages and limitations of factorial designs
- identify the types of repeated measures design

- distinguish between analysis of variance and analysis of covariance
- express some of the advantages and disadvantages of multivariate analysis
- state the methodology and data analysis of time series design
- give some possible examples of trend studies in Nigeria
- discuss some of the problems of meta-analysis
- explain the types of measurement
- describe the major processes involved in test construction
- discuss the advantages and limitations of questionnaires, interview schedules, rating scales, and attitude/interest inventories as measurement instruments
- distinguish between descriptive and inferential statistical tools in data processing
- exemplify the use of parametric and non-parametric techniques in data analysis.

WORKING THROUGH THIS COURSE

For you to be able to successfully complete this course, you will need to diligently read through this course book from Unit 1 of the first module to the last unit of the sixth module, and carry out/perform all activities therein stated. Each unit contains self-assessment exercises and tutor-marked assignments which you will need to do and submit the latter (TMAs) periodically at designated points in the course. Part of your TMA scores, which are regarded as continuous assessment, will form 50% of your overall assessment for the course, the balance 50% being your score in the final examination.

The course should take you about 42 weeks to complete. Your tutor/facilitator for the course will be able to attend to you in case of any difficulty in comprehending any aspect of the course material. You are therefore advised to read ahead and attend the tutorial sessions where you would be able to ask questions and interact with your colleagues and the facilitator. Below are the components of the course, what you have to do, and how you should allocate your time to each unit to be able to complete the course successfully according to schedule.

THE COURSE MATERIALS

The major components of the course, which will be made available to you, include:

- the course guide
- the course material with the corresponding study units

- assignment file (which will be available from the web CT OLE in due course)
- presentation Schedule.

STUDY UNITS

The study units in this course are as follows:

Module 1 Nature, Classifications and Information Sources of Educational Research

- Unit 1 Definitions, purposes and procedures of educational research
- Unit 2 Classifications of educational research
- Unit 3 Sources of educational information

Module 2 Basic Types of Educational Research

- Unit 1 Humanistic research
- Unit 2 Survey research
- Unit 3 Correlation research

Module 3 Designs of Research I

- Unit 1 Causal-comparative research
- Unit 2 Quasi-experimental research
- Unit 3 Experimental research
- Unit 4 Factorial designs

Module 4 Designs of Research II

- Unit 1 Repeated measures design
- Unit 2 Twin studies
- Unit 3 Analysis of variance and covariance
- Unit 4 Multivariate studies

Module 5 Time Series, Trend, and Meta Analysis Studies

- Unit 1 Time-series design
- Unit 2 Trend studies
- Unit 3 Meta-analysis

Module 6 Instrumentation and General Data Processing Techniques

- Unit 1 Meaning, types and classification of measurement
- Unit 2 Construction and validation of measurement instruments
- Unit 3 Descriptive data analysis

Unit 4 Inferential data analysis

ASSIGNMENT FILE

Your assignment file will be posted on the Web CT OLE in due course. In this course, you will find all the details of the work you must submit to your tutor for marking. The marks you obtain from these assignments will count towards the final mark you obtain for this course. Further information on assignments will be found in the assignment file itself, and later on in the section on assessment in this course guide. There are over twenty one (21) tutor-marked assignments in this course, and you are expected to practice all of them but submit at most six (6).

PRESENTATION SCHEDULE

The presentation schedule included in your course materials gives you the important dates for this year for the completion of tutor-marked assignments (TMAs) and for attending tutorials. Remember, you are required to submit all your assignments by the due dates. You should guard against falling behind in your work.

ASSESSMENTS

There are two aspects to the assessment of this course: first are the tutor-marked assignments (50%), and second is a written examination (50%).

In tackling the assignments, you are expected to apply information, knowledge and techniques gathered during the course. The assignments must be submitted to your tutor for formal assessment in accordance with the deadlines stated in the Presentation Schedule and the Assignment File.

At the end of the course, you will need to sit for a final written examination of three hours duration.

TUTOR-MARKED ASSIGNMENTS (TMAs)

There are over 21 tutor-marked assignments in this course, and you are advised to attempt all but submit six (6). Aside from the course material provided, you are advised to read and research widely using other references which will give you a broader viewpoint and may provide a deeper understanding of the subject. Ensure all your completed assignments are submitted on schedule before set deadlines. If for any reasons you cannot complete your work on time, contact your tutor before the assignment is due to discuss the possibility of an extension of

time. The extension may not however be granted after the due date, except in exceptional circumstances or for very genuine excuse.

FINAL EXAMINATION AND GRADING

The final examination for this course will be of three hours duration and carries a value of 50% of the total course grade. All areas of the course will be assessed, and the examination will consist of questions which reflect the type of self-testing, practice exercises and tutor-marked problems you have previously encountered.

Utilize the time between the conclusion of the last study unit and sitting for the examination to revise the entire course. You may find it useful to review your self-assessment exercises, tutor-marked assignments and comments before the examination.

COURSE MARKING SCHEME

The work you submit will count for 50% of the total course mark. At the end of the course, you will be required to sit for a final examination which will also count for 50% of your total marks. The table below shows how the actual course marking is broken down.

Table I: Course Marking Scheme

ASSESSMENT	MARKS
Assignment 6 (TMAs)	6 assignments, best 5 will be used for CA, i.e. 5 x 10 = 50% of overall course marks
Final Examination	50% of overall course marks
Total	100% of course marks

COURSE OVERVIEW

The table below brings together the units and the number of weeks you should take to complete them and their accompanying assignments.

Unit	Title of Work	Week's activity	Assessment (end of unit)
1.	Definitions, purposes and procedures of educational research		
2.	Classifications of educational research		
3.	Sources of educational information		
4.	Humanistic research		
5.	Survey research		

6.	Correlation research		
7.	Causal-comparative research		
8.	Quasi-experimental research		
9.	Experimental research		
10.	Factorial designs		
11.	Repeated measures design		
12.	Twin studies		
13.	Multivariate studies		
14.	Analysis of variance and covariance		
15.	Time-series design		
16.	Trend studies		
17.	Meta-analysis		
18.	Meaning, types and classification of measurement		
19.	Construction and validation of measurement instruments		
20.	Descriptive data analysis		
21.	Inferential data analysis		
	Revision		
	Total		

HOW TO GET THE MOST FROM THIS COURSE

In distance learning, the study units are specially developed and designed to replace the university lecturer. Hence, you can work through these materials at your own pace, and at a time and place that suit you best. Visualize it as reading the lecture instead listening to a lecturer.

Each of the study units follows a common format as earlier explained under the section on 'Working through the Course'. The first item is an introduction to the subject matter of the unit, and how a particular unit is integrated with the other units and the course as a whole. Next is a set of learning objectives. These objectives let you know what you should be able to do by the time you have completed the unit. You should use these objectives to guide your study. When you have finished the unit, you must go back and check whether you have achieved the objectives. If you make a habit of doing this, you will significantly improve your chances of passing the course.

The main body of the unit guides you through the required reading from other sources. This will usually be either from your set books or from a **Reading Section**.

Self-assessment exercises are interspersed throughout the units, and the answers are given at the end of the units. Working through the self-assessment exercises will help you to achieve the objectives of the units and prepare you for the assignments and examinations. You should do each activity as you come to it in the study unit.

The following is a practical strategy for working through the course. If you run into any trouble, telephone your facilitator or post the questions on the Web CT OLE's discussion board. Remember that your facilitator's job is to help you. When you need help, don't hesitate to call and ask your tutor to provide it. In summary,

- Read this course guide.
- Organize a study schedule. Refer to the course overview for more details. Note the time you are expected to spend on each unit and how the assignments relate to the unit. Important information e.g. details of your tutorials, and date of the first day of the semester is available from the Web CT OLE. You need to gather together all this information in one place, such as your diary or wall calendar. Whatever method you choose to use, you should decide on and write in your own dates for working on each unit.
- Once you have created your own study schedule, do everything you can to stick to it. The major reason that students fail is that they get behind with their coursework. If you get into difficulties with your schedule, please let your facilitator know before it is too late for help.
- Turn to Unit 1 and read the introduction and the objectives for the unit.
- Assemble the study materials. Information about what you need for a unit is given in the 'Overview' at the beginning of each unit.
- Keep an eye on the Web CT OLE. Up-to-date course information will be continuously posted there.
- Well before the relevant due dates (about 4 weeks before the dates) access the Assignment file on the Web CT OLE and download your next required assignment. Keep in mind that you will learn a lot by doing the assignments carefully. They have been designed to help you meet the objectives of the course and, therefore, will help you pass the examination. Submit all assignments not later than the due dates.
- Review the objectives for each study unit to confirm that you have achieved them. If you feel unsure about any of the objectives, review the study material or consult your facilitator.
- When you are confident that you have achieved a unit's objectives, you can then start on the next unit. Proceed unit by unit through the course and try to pace your study so that you keep yourself on schedule.

- When you have submitted an assignment to your facilitator for marking, do not wait for its return before starting on the next unit. Keep to your schedule. When the assignment is returned, pay particular attention to your facilitator's comments. Consult your facilitator as soon as possible if you have any questions or problems.
- After completing the last unit, review the course and prepare yourself for the final examination. Check that you have achieved the unit objectives and the course objectives.

TUTORS AND TUTORIALS

There are 20hours of tutorials (ten 2-hour sessions) provided in support of this course. You will be notified of the dates, times and location of these tutorials, together with the names and phone number of your facilitator, as soon as you are allocated a tutorial group.

Your facilitator will mark and comment on your assignments. Keep a close watch on your progress and on any difficulties you might encounter as they would provide assistance to you during the course. You must mail your tutor-marked assignments to your facilitator well before the due date (at least two working days are required). They will be marked by your facilitator and returned to you as soon as possible. Do not hesitate to contact your facilitator by telephone, e-mail, or discussion board if you need help. The following might be circumstances in which you would find help necessary: when

- you do not understand any part of the study units or the assigned readings.
- you have difficulty with the self-assessment exercises.
- you have a question or problem with an assignment, with your facilitator's comment on an assignment or with the grading of an assignment.

You should try your possible best to attend the tutorials. This is the only chance to have face-to-face contact with your facilitator and to ask questions which are answered instantly. During the tutorial, you can raise any problem encountered in the course of your study. To gain the maximum benefit from course tutorials, prepare a question list before attending them. You will learn a lot from participations in discussions.

REFERENCES/FURTHER READING

- Badmus, G.A. and P. Odor (1996) (eds.): *Challenges of Managing Educational Assessment in Nigeria*. NABTEB, Benin City. Kaduna: Atman Publishers.
- Badmus, G.A. and C.N. Omoifo (1998): *Essentials of Measurement and Evaluation in Education*. Benin City: Osasu Publishers.
- Best, J.W. (1970): *Research in Education*, New Jersey: Prentice Hall.
- Campbell, D.T. and Stanley, J.C. (1963). *Experimental and quasi-experimental designs for Research*. Chicago: Rand McNally.
- Creswell, J.W. (2009). *Educational Research: Planning, conducting and evaluating quantitative and qualitative research*, 3rd edition, Columbus, OH: Prentice Hall.
- Fisher, R.A. (1970). *Statistical Methods for Research Workers 14th Ed.* New York: Hafner Publishing Company Inc.
- Gay, L.R. (1996). *Educational Research: Competencies for Analysis and Application*. Columbus, Ohio: Charles E. Merrill Publishing Company.
- Guilford, J.P. (1965), *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill Book Company.
- Habermas, J. (1972c) *Toward a Rational Society*. Heinemann, London.
- Hassan, T. (1995): *Understanding Research in Education*. Lagos: Merrifield Publishing Co.
- Kaplan A. (1964) *The Conduct of Inquiry: Methodology for Behavioral Sciences*. Chandler, San Francisco, California.
- Keeves, J.P. (ed.) (1990): *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford: Pergamon Press.
- Kerlinger, F.N. (1973). *Foundations of Behavioural Research*, Second Edition. New York: Holt, Rinehart and Winston.
- Koul, R.B. (1992): *Methodology of Educational Research*. India, Vikas Publishing House PVT Ltd.

Nie, N.H. Hull, C.H. Jenkins, J.G., Steinbrenner, K. and Bent, D.H. (1975). *Statistical Package for the Social Sciences*. New York: McGraw-Hill Book Company.

Nwana, O.C (1982): *Introduction to Educational Research*. Ibadan: Heinemann.

Nyanjui, P.J. (2006): *Introduction to Research*. Nairobi: Kenya Institute of Education.

Popkewitz, T.S. (1984). *Paradigm and ideology in educational research*. London: The Falmer Press.

Stevens, S.S. (1951) (Ed.). *Handbook of Experimental Psychology*. New York: Wiley

Tuckman, B.W. (1978). *Conducting Educational Research*, Second Edition. New York: Harcourt Brace Javanovich, Inc.

SUMMARY

This course is aimed at exposing the graduate student in education to advanced research techniques. Specifically, you will be equipped with the knowledge to understand and appreciate, as well as to produce a good research work, and be able to answer questions such as

- what is educational research?
- how do you distinguish between theoretical and action research?
- what do you understand by the terms “causal relationship”, “matching”, “randomization”?
- in what situations can we use factorial design?
- what is the value of twin studies in educational research?
- can we justify meta-analysis in educational research?
- what is the meaning of the F-ratio in analysis of variance?

I wish you success with this course, and hope that you will find your acquaintance with the National Open University of Nigeria (NOUN) interesting, useful and rewarding.

**MAIN
COURSE**

CONTENTS	PAGE
Module 1 Nature, Classifications and Information Sources of Educational Research	1
Unit 1 Definitions, purposes and procedures of educational research	1
Unit 2 Classifications of educational research	8
Unit 3 Sources of educational information	15
Module 2 Basic Types of Educational Research	19
Unit 1 Humanistic research	19
Unit 2 Survey research	27
Unit 3 Correlation research	33
Module 3 Designs of Research I	37
Unit 1 Causal-comparative research	37
Unit 2 Quasi-experimental research	42
Unit 3 Experimental research	46
Unit 4 Factorial designs	51
Module 4 Designs of Research II	55
Unit 1 Repeated measures design	55
Unit 2 Twin studies	59
Unit 3 Analysis of variance and covariance	66
Unit 4 Multivariate studies	72
Module 5 Time Series, Trend, and Meta Analysis Studies	77
Unit 1 Time-series design	77
Unit 2 Trend studies	82
Unit 3 Meta-analysis	85

Module 6	Instrumentation and General Data Processing Techniques	91
Unit 1	Meaning, types and classification of Measurement	91
Unit 2	Construction and validation of measurement instruments	104
Unit 3	Descriptive data analysis	116
Unit 4	Inferential data analysis	128

MODULE 1 NATURE, CLASSIFICATIONS AND INFORMATION SOURCES OF EDUCATIONAL RESEARCH

- Unit 1 Definitions, Purposes, and Procedures of Educational Research
- Unit 2 Classifications of Educational Research
- Unit 3 Sources of Educational Information

UNIT 1 DEFINITIONS, PURPOSES AND PROCEDURES OF EDUCATIONAL RESEARCH

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 What is Educational Research?
 - 3.1.1 Definitions of Educational Research
 - 3.2 Purposes of Studying Educational Research
 - 3.3 Procedures for Conducting Educational Research
 - 3.3.1 Selection and definition of a Research Problem
 - 3.3.2 Conducting a Literature Review
 - 3.3.3 Formulation and Statement of Research Hypothesis
 - 3.3.4 Developing a Research Plan or Design
 - 3.3.5 Subjects and Sampling
 - 3.3.6 Analysis of data
 - 3.3.7 Reporting conclusions
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

No doubt, you are familiar with the term “research” as a systematic way of finding out and solving problems in all facets of life. Meaningful research is central and critical to the science of education, like any of the other sciences. This unit introduces you to the meaning and nature of educational research, and some ways in which such research studies can be characterised. It will also give you an overview of some of the basic procedures in educational research. The next section spells out the objectives of the unit.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define research
- distinguish between scientific and non-scientific research
- recount the purposes for conducting research
- list the basic procedures for conducting education research
- explain the role of educational research for educational practice.

3.0 MAIN CONTENT

3.1 What is Educational Research?

3.1.1 Definitions of Educational Research

As to be expected, definitions abound for the word “research” since it has both everyday and specialized connotations.

Etymologically, the word research itself is derived from the French word “recherche” which means to travel through or to survey. UNESCO (1962) defined research as “the orderly investigation of a subject matter for the purpose of adding to knowledge”.

Kerlinger (1973) defined research as “systematic, controlled, empirical and critical investigation of hypothetical prepositions about the presumed relations among natural phenomena”.

According to Rekha Koul (2008), the nature of educational research is analogous to the nature of research itself, which is a “careful, systematic, reliable and valid method of investigating knowledge and solving problems” (Wiersma 1991).

For purpose of this volume, educational research may be therefore summarily defined as a careful, systematic investigation into any aspect of education; its goal is to explain, predict, and/or control educational phenomena.

Research can also be characterized in several ways. One such characterization is given by Hassan (1995), paraphrasing Best (1985) and Tuckman (1978), who listed the following characteristics of research and the research process: According to Hassan, research

- is directed toward the solution of a problem, i.e. it involves the quest for answers to unsolved problems;

- involves careful collection, organization and articulation of what is already known about the problem and what is yet to be known;
- is a structured process which follows a systematic order or rule of execution;
- is characterized by rigorous logic and objectivity in the carefully designed procedures and analysis;
- demands accurate observation and description of phenomena;
- involves gathering new data from primary or first hand sources, or using existing data for new purposes;
- involves logical and plausible explanation of the findings of the study;
- is replicable and emphasizes the development of generalizations, principles or theories that can be used in predicting future occurrences;
- is expensive in terms of time, money, resources and energy.(pp 10-12)

3.2 Purposes of Studying Educational Research

1. To orient students to the nature of educational research: its purposes, forms and importance.
2. To provide information which helps students become more intelligent consumers of educational research: where to locate it, how to understand it, and critique it.
3. To provide information on the fundamentals of doing educational research such as selecting a problem, using available tools, organizing a project, etc.
4. To generate new theories, confirm existing ones or disapprove them, e.g. the role of punishment in discipline.

SELF – ASSESSMENT EXERCISE 1

- i. What do you understand by educational research? And why do it?
- ii. List some of the characteristics of research and the research process.

3.3 Procedures for Conducting Educational Research

3.3.1 Selection and Definition of a Research Problem

The task of defining the research problem is of the greatest importance in the entire research process. Problems are the basis for research, and the first step in the research process is to identify a well-defined problem. It is said that a problem well formulated is already half solved. For definiteness, problems are usually stated in either question form or

hypothesis form. These two serve the purpose of focusing the problem for the researcher and the reader.

In selecting a research problem, the following points must be borne in mind. The problem must be:

- Interesting, i.e. the topic should be of considerable interest to the researcher and also to others. It should not be dull or mundane but should excite the curiosity of the researcher.
- Significant: The research problem should be significant to education from either a practical or theoretical viewpoint.
- Researchable: Obviously, not all problems in education, or for that matter anywhere else, are researchable. Take for an example a problem statement like: do beautiful school premises and surroundings affect achievement in English Literature? Such a problem will require telescoping the scope and operationalising the definition of the problem, since ‘beauty is in the eyes of the beholder’. As Kerlinger (1976) put it, a problem should imply possibilities for testing. For example, the question, “does appreciation of God affect achievement in religious studies in schools? is hardly testable.
- Feasible: Even if the problem is researchable, doing the research may not be feasible in terms of time, money, and energy resources.

3.3.2 Conducting a Literature Review

Literature review is akin to describing the “state of play” in the area selected for study. It should represent a distillation of the essential issues and inter-relationships associated with the knowledge, arguments and themes that have been explored in the area. Provided it is not merely a catalogue, or even a summary, of research studies and their findings in the area, review of relevant literature will help the researcher gain an understanding of the current state of knowledge pertaining to that research idea. If properly carried out, it will inform you if the research problem or topic has already been explored, and if a revision or replication is needed, how to design your own study, what data collection methods to use, and help make sense of the findings of your study once data analysis is complete.

3.3.3 Formulation and Statement of Research Hypothesis

Statement of the research problem should be concise and provide adequate focus and direction of the study, identify key factors and provide limits. It affects the manner in which tests must be conducted in the analysis of data.

Usually, it takes the form of a question or a statement in the form of a hypothesis. A hypothesis is a clear, unambiguous but tentative conjecture or assumption about the relationship between two or more variables which can be subjected to scientific verification or refutation, and could be stated in the null or alternate/research form.

3.3.4 Developing a Research Plan or Design

A research design is the conceptual framework within which the research will be conducted. It has also been labeled the blueprint of the research, since it is used to structure the research. Development of an overall research design will include specification of the information that is to be collected, from which individuals, and under what research conditions.

The two major types of research designs are the randomized or true experiment, and the non-experiment. These will be elaborated upon in later modules of this course.

3.3.5 Subjects and Sampling

A population refers to all the members or items under consideration in a research. It is very often not possible to study the entire population for obvious reasons of cost, time, energy, volume of data, etc. Therefore, a sample frequently has to be drawn from the population as possible. For this purpose, sampling designs are used, mostly in the form of probability and non-probability samples.

3.3.6 Analysis of Data

Analysis of data involves the transformation of raw data into manageable categories, through coding and tabulation, for further analysis. This analysis is usually based on computation of various statistical measures through (nowadays) analysis software as SPSS, EPI info, Excel and Access. In analysis, relationships or differences that support or conflict the original hypotheses are subjected to tests of significance to determine the validity/confidence with which conclusions can be made. In the absence of hypothesis, the researcher

does his best to explain the findings using intuition, logic and experience.

3.3.7 Reporting Conclusions

Conclusions are statements that interpret and evaluate the results found from the study. Aspects of the Report to be covered include a statement of the problem under investigation, methodology used, scope of the study, and the limitations. Emphasis should be given to the results that relate to the hypotheses or research questions of the study. Relevant factors to note here include tailoring the report content to the audience, integrating your findings with the results from previous research, accession of whatever practical or theoretical implications can be drawn from your study, and suggestions for future research.

SELF – ASSESSMENT EXERCISE

- i. List some of the pertinent points in selecting a research problem.
- ii. What is the importance of literature review in educational research?

4.0 CONCLUSION

Research is so important in our everyday life today that it is necessary that every ‘educated’ person should be able to at least understand it.

The preceding pages have highlighted various definitions of research and some of its characteristics, purposes and procedures. This should give you a feel of what it is to do research in education and how to get about it.

5.0 SUMMARY

To understand and properly engage in educational research requires an appreciation of what research is, why and how it is done, and what are its characteristics and procedures. That is the summary of this Unit which has introduced you to the early basics of research. The next unit will explore this introductory theme further.

6.0 TUTOR – MARKED ASSIGNMENT

- i. Define “research” in your own words. What are its uses?
- ii. List and explain some of the procedures for conducting educational research.

7.0 REFERENCES/FURTHER READING

Best, J.W. (1970). *Research in Education*, New Jersey: Prentice Hall.

Creswell, J.W. (2009). *Educational Research: Planning, conducting and evaluating quantitative and qualitative research*, 3rd edition, Columbus, OH: Prentice Hall.

Gay, L.R., Geoffrey Mills, & Peter Airasian (2006). *Educational Research: Competencies for Analysis and Application*, 8th edition, Prentice Hall.

Hassan, T. (1995). *Understanding Research in Education*. Lagos: Merrifield Publishing Co.

Kerlinger, F.N. (1973). *Foundations of Behavioral Research*, Second Edition. Holt, Rinehart and Winston, New York.

Koul, R.B. (1992). *Methodology of Educational Research*. India, Vikas Publishing House PVT Ltd.

Nyanjui, P.J. (2006). *Introduction to Research*. Nairobi: Kenya Institute of Education.

Nwana, O.C (1982). *Introduction to Educational Research*. Ibadan: Heinemann.

UNIT 2 CLASSIFICATIONS OF EDUCATIONAL RESEARCH

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Classifications of Educational Research
 - 3.1.1 Basic/Applied Research
 - 3.1.2 Quantitative/Qualitative Research
 - 3.1.3 Theoretical/Action Research
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

In Unit 1 you were introduced to some of the rudiments of educational research in terms of definitions, purposes and procedures of educational research. This second unit takes you further into the ways educational research is classified, and will be followed in the next unit by a look at some of the sources of educational information.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- state what classification is
- distinguish between Basic and Applied Research
- distinguish between Quantitative and Qualitative Research
- distinguish between Theoretical and Action Research
- state the interrelationships between these classifications of research.

3.0 MAIN CONTENT

3.1 Classifications of Educational Research

Classification is the act or system of methodically arranging a phenomenon in distinct divisions or classes. Research studies have been classified in a variety of ways, but for the purposes of this course three such classifications will be considered, viz;

1. Basic/Applied Research
2. Quantitative/Qualitative Research
3. Theoretical/Action Research

It should be emphasized that these are not watertight categories, since most research has elements of all of them.

3.1.1 Basic/Applied Research

Basic research, sometimes also called fundamental or pure research, is primarily concerned with the development and advancement/testing of knowledge or theory through verification of hypothesis (e.g. $e = mc^2$ of Einstein and Newton's $f = ma$). It is not particularly concerned with practical application of knowledge to solve everyday problems or to introduce reforms into present practice. Its rationale is that the purpose of science is to describe and explain the world as it is, and not to change it – in obvious contrast to Karl Marx's famous dictum that the purpose of philosophy is to change the world! This is often referred to as the “scientific or rationalistic paradigm” of enquiry. The word paradigm means/denotes an example or model or pattern of something. Research traditions are best regarded as different paradigms, and a paradigm determines how a problem is formulated and methodologically tackled.

The scientific paradigm of research assumes that the researcher is able to maintain a discrete and inviolable distance from the “object” of enquiry, and asserts that the aim of research is “to develop a nomothetic body of knowledge best stated in generalizations which are truth statements of enduring value that are context free” (Guba, E.G and Lincoln, Y.S. 1982). According to this paradigm, every effect has a cause, and it is the purpose of a “true experiment” to demonstrate this cause-effect relationship.

Quintessentially, basic research is value-free, partly because of the objective/quantitative methods used and therefore their greater mathematical manipulability. Theories under this paradigm are formulated a priori to guide enquiry, and only propositional knowledge, i.e. knowledge that can be stated in language form, is admissible in scientific enquiry. Finally, basic research finds its most comfortable niche in a laboratory setting where control of variables can be maximized.

In contrast to basic research, applied research is generally concerned with the application of theory to the solution of immediate practical problems. This may take three forms:

- Evaluation Research, which is the systematic process of gathering data to make decisions of educational relevance, e.g. how would a program of General Studies (GST) affect the final degree class of undergraduate students in Nigeria?
- Research and Development (R&D), used to develop effective products for use in a school system, e.g. development of student learning materials or teacher training materials. What is the relative worth of 2 or more alternative decisions? Is Otonti eligible to participate in the schools gifted/talented program?
- Action Research is a unique form of applied research which is focused on solving specific practical educational problems through the application of the scientific method. (Gay, 1996). Its value is primarily confined to those conducting it, and its results may not be generalized beyond the sample used for the study. For example, what can be done to significantly reduce high failure rate, in SSCE mathematics examinations in schools in Ika North East LGA of Delta State?

SELF – ASSESSMENT EXERCISE

- i. Distinguish between basic and applied research?

3.1.2 Quantitative/Qualitative Research

This is sometimes also referred to as “empirical/analytic/symbolic/positivist vs interpretive/post positivist” paradigm. Quantitative research, as the names implies, relies mostly on numerical data such as the use of mathematical tools (especially statistics), and on hermeneutical data such as test scores and other measures of performance to explain, predict, and/or control phenomena of interest, using the deductive process for data analysis.

According to Smith and Heshusius (1986), the epistemological rationale for the quantitative researcher is that there exists a mind-independent and tangible reality “out there” that is knowable to some extent. Quantitative research is therefore said to be positivist, hypothesis testing, and statistical.

Many recent writers in education infact distinguish two fundamental paradigms of research: the “scientific” (which is often erroneously identified with positivism) and the “interpretative” or “humanistic”. They additionally claim that the “scientific” and “humanistic” approaches are not mutually exclusive but complementary to each other. A broader, three-way taxonomy of research is given by Popkewitz

(1984) who claims that in educational sciences three paradigms have emerged to give definition and structure to the practice of research, i.e. empirical-analytic (roughly equivalent to the distinction between intent and consequences, i.e. the danger.....shaped. intent cannot be inferred from consequences, and many times, consequences belie intent. (Kaestle, 1988).

Qualitative research, on the other hand, involves the collection of extensive ‘narrative data’ (i.e. non-numerical data) on many variables over a period of time in order to gain insights into phenomena of interest. Example of a qualitative research would be “a case study of parental involvement in Southpoint Royal School, Benin City”.

Its data analysis includes coding of data and production of a verbal synthesis by the inductive process. Two main examples of this approach would include historical research and ethnographic research.

Historical research is concerned with the study of past events. It generates descriptions, and sometimes attempted explanations of conditions, situations and events that have occurred in the past. Some examples of historical research would be:

1. Evolution of teacher-training programs in Nigeria 1900 – 1960
2. History and development of free primary education in Western Nigeria, 1957 -1965

Some of the problems associated with historical research include (i) vagueness in defining key terms and (ii) the problem of “presentism”, (iii) lack of clear distinction between ideas about how people did in fact behave, and how ordinary people did in fact behave (Kaestle, 1988)

Ethnographic research involves the study of current events in a naturalistic setting. It usually consists of a description of events that occur among the life of a group, with particular reference to the interaction of individuals in the context of the sociocultural norms, rituals and beliefs shared by the group. Participant observation, where the researcher lives with the subjects being observed, is frequently used in this kind of research. Child rearing practices among the Igbo ethnic group of the South Eastern Nigeria would be an example of the method.

SELF – ASSESSMENT EXERCISE II

- i. Give and justify an example of action research.

3.1.3 Theoretical/Action Research

A theory arises in answer to the question why. A basic facet of human nature is that man is perennially trying to ascribe meaning to his experience, to find a rational order in the universe. That is the purpose of science or art or philosophy, and indeed of all knowledge. Man's discoveries about himself and his environment, put in an orderly form, have created the vast structure of knowledge that is called 'science' whether in the field of physics, chemistry, biology, or in economics, psychology, sociology or meteorology. Prof Bronowski (1956) defined 'science' as "the organization of our knowledge in such a way that it commands more of the hidden potential in nature". Theoretical research therefore is that research that is interested in discovering knowledge for its own sake and which may not necessarily result in action or any given direction of action.

Action research, on the other hand, can be described as a family of research methodologies which pursue action (or change) and research (or understanding) at the same time, by using a cyclic or spiral process which alternates between action and critical reflection. In most of its forms, it is qualitative and participative. It can be used as a research tool for investigative or pilot research, and generally for diagnosis or evaluation (Hart, 1995).

Action research may attempt to determine the value of a product, procedure or program in a particular setting (e.g. in a school), with the goal of improving same. It is concerned with immediate solutions to real problems, and does not attempt to generalize results for a broader population. According to Gray (1996), "the purpose of action research is to solve practical problems through the application of the scientific method". An example would be "what can be done to prevent the "deliberate foul play" which seems to break out every weekend in the school football field amongst some students?"

Action research is also construed as research into one's own practice, or more accurately into one's praxis (i.e. informed committed action), to enable the practitioner understand his/her practices more deeply and therefore improve on them. Frequently, action research is participatory (i.e. involving a group or more than one participant), and collaborative with, say, other teachers or students or parents. In summary, educational action research is a form of research which places control over processes of educational reform in the hands of those involved in the action. (Kemmis, 1988).

SELF – ASSESSMENT EXERCISE

- i. What are some of the problems associated with ethnographic research?

4.0 CONCLUSION

For convenience, educational research has been classified under three main headings of Basic versus Applied Research, Quantitative versus Qualitative Research, and Theoretical versus Action Research, which have been adumbrated in this unit.

As stated in the introductory remarks, these classifications of research are not mutually exclusive, but nevertheless give useful insights into the various research paradigms. Understanding of their essence will enable you to get the correct initial perspectives of the broad area of educational research, which will help you in further work on this course.

5.0 SUMMARY

In this unit, you have learnt that

- Basic Research primarily serves the need/urge to know or understand a phenomenon
- Applied Research is more concerned with the solution of immediate practical problems.
- Quantitative Research uses mostly numerical data and the deductive method for data analysis.
- Qualitative Research relies on narrative data and the use of the deductive method.
- Theoretical Research is akin to basic research in its focus on discovering knowledge for its own sake.
- Action Research has two connotations: (a) to solve specific practical problems using the scientific method and (b) to enable practitioners understand their practices more deeply and improve on them.

6.0 TUTOR – MARKED ASSIGNMENT

- i. Critically examine the three classifications of Research. How are they interrelated?

7.0 REFERENCES/FURTHER READING

- Best, J.W. (1970). *Research in Education*, New Jersey: Prentice Hall.
- Creswell, J.W. (2009). *Educational Research: Planning, conducting and evaluating quantitative and qualitative research*, 3rd edition, Columbus, OH: Prentice Hall.
- Gay, L.R., Geoffrey Mills, and Peter Airasian (2006). *Educational Research: Competencies for Analysis and Application*, 8th edition, Prentice Hall.
- Hassan, T. (1995). *Understanding Research in Education*. Lagos: Merrifield Publishing Co.
- Kerlinger, F.N. (1973). *Foundations of Behavioral Research*, Second Edition. Holt, Rinehart and Winston, New York.
- Koul, R.B. (1992). *Methodology of Educational Research*. India, Vikas Publishing House PVT Ltd.
- Lincoln Y.S, & Guba E.G. (1985) *Naturalistic Inquiry*. Sage, Beverly Hills, California.
- Nyanjui, P.J. (2006). *Introduction to Research*. Nairobi: Kenya Institute of Education.
- Nwana, O.C (1982). *Introduction to Educational Research*. Ibadan: Heinemann.
- Smith, J.K, Heshusius, L. (1986). Closing down the conversation. The end of the qualitative/quantitative debate among educational inquirers *Educ. Res. AERA* 15(1): 4-12.
- Wiersma, W. (2007). World views, paradigms and the practice of social science research. In Mukta, J. & Rema, N. (Eds.) *Foundations of Qualitative Research: Interpretive and Critical Approaches*, 1-26. Sage, Thousand Oaks, CA.

UNIT 3 SOURCES OF EDUCATIONAL INFORMATION

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Sources of Educational Information
 - 3.1.1 Traditional Sources of Information
 - 3.1.2 The Computer Revolution
 - 3.1.3 Current Common Sources of Information
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Having been introduced to the elementals of research, the next logical step is to try to locate sources of gathering information about your topic. This is the subject of this Unit which discusses traditional and modern sources of educational information and how to access them. These sources include primary and secondary sources, including importantly the use of the computer.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- identify primary and secondary sources of information for educational research
- effectively use the resources of the traditional library to obtain data for educational research
- know how to use the computer to source for educational information.

3.0 MAIN CONTENT

3.1 Sources of Educational Information

3.1.1 Traditional Sources of Information

A basic initial activity/requirement in any research, including educational research, is the acquisition of data or information relevant to the study or research. These data can be obtained from a variety of

sources which may be categorized into primary and secondary sources. In education, some of these sources include books, periodicals, reports (technical, seminar, workshop), conference proceedings, academic theses and manuscripts. Electronic versions of most of these information sources are generally now available on-line and can be downloaded from the computer/internet.

Books and journals are primary sources in educational research, since they contain original work of the writers and form materials for secondary sources. Professional journals especially are usually up to date, since the articles in them give reports of recent research studies. Books, although very useful, are usually not as current as journals.

Primary sources usually provide more information about a study than can be found elsewhere, and are also a good source of the research methodology used, but they can be time consuming to cover the study. Locating sources of educational information can in fact be bewildering to neophyte researchers, but actually today the real problem is delimiting the scope of the vast resources to be consulted. For this, there are indexes, abstracts and other retrieval mechanisms available in the standard library to ameliorate this problem, e.g. the card catalogues and the bibliographic indexes. For example, the card catalogue will help the researcher locate pertinent books, and the indexes will help him/her find articles in journals.

3.1.2 The Computer Revolution

Traditionally, the library has, over the centuries, been the main repository of non-electronic or hard copy sources of information, but the vast explosion of information or knowledge in the last half century or so, due partly to research itself, has rendered impossible the storage of information in the form of books and other hard copies, in one building or location (i.e. the library). The electronic digital computer can store an incredibly vast amount of knowledge or information in micro storage units (e.g. CD-ROM) which can be accessed with phenomenal speed and ease. Therefore, modern libraries, in addition to their shelves of books and periodicals now stock large collections of electronic materials, audio and video, which can be accessed in seconds.

In many universities, computer terminals and keyboards have now replaced the old card catalogs and indexes. With the electronic digital computer, new knowledge/information in all fields is now being generated daily at an exponential rate, which information can now be accessed almost instantaneously from all over the world with the click of a button.

CD-ROM journal indexes and database searches are very useful tools in identifying and locating various references for the researcher. The CD-ROM usually focuses on a single specific database. Online computer searches, on the other hand, can have up to 4000 databases which provide access to literally billions of records. Now a researcher can sit down before a computer monitor in the library or his office or home and with a flick of the button gain almost instantaneous access to previously unimaginable records of educational information from all over the world, information which are frequently updated. It however requires considerable skill and some practical experience and competence on the part of the researcher to sort out and sift through and efficiently manage this bulk of information to serve his/her purposes.

3.1.3 Current Common Sources of Information

In educational research, some of the most common sources of information are

- (1) Education Index
- (2) Educational Resources Information Center (ERIC) – a free bibliographic database of more than 1m citations on education topics since 1966
- (3) Dissertation Abstracts International
- (4) Psychological Abstracts
- (5) Review of Educational Research
- (6) Encyclopedia of Educational Research
- (7) American Education Research Association (AERA)
- (8) British Education Research Association (BERA)
- (9) National Foundation for Educational Research (NFER)
- (10) International Education Research Foundation (IERF)
- (11) Nigerian Educational Research Council (NERC)
- (12) Museums
- (13) National Archives
- (14) Special Collections in Libraries

4.0 CONCLUSION

Progress on the research road requires knowledge and skills of getting information from a number of sources, including the traditional books, periodicals and other hard copy materials from the standard library, as well as on-line computer searches. This Unit has touched on these assistance mechanisms, in the hope that you should be able to use them effectively in carrying out your research objectives. In the next module, you will be introduced to the basic types of educational research.

5.0 SUMMARY

The basic data-gathering tools of educational research and how to access them have been highlighted in this unit. These range from the traditional library resources to the modern electronic computer. Research today has never been so exciting in spite of the heavy work involved.

6.0 TUTOR-MARKED ASSIGNMENT

- i. List and evaluate the various sources of obtaining information for educational research
- ii. Why is it a desideratum that today's educational researcher must be computer literate?

7.0 REFERENCES/FURTHER READING

Cresswell, J.W. (2009). *Educational Research: Planning, conducting and evaluating quantitative and qualitative research*, 3rd edition, Columbus, OH: Prentice Hall.

Gay, L.R., Geoffrey Mills, & Peter Airasian (2006). *Educational Research: Competencies for Analysis and Application*, 8th edition, Prentice Hall.

Hassan, T. (1995). *Understanding Research in Education*. Lagos: Merrifield Publishing Co.

MODULE 2 BASIC TYPES OF EDUCATIONAL RESEARCH

Unit 1	Humanistic Research
Unit 2	Survey Research
Unit 3	Correlation Research

UNIT 1 HUMANISTIC RESEARCH

CONTENTS

1.0	Introduction
2.0	Objectives
3.0	Main Content
3.1	Humanistic Research
3.1.1	Historical Research
3.1.2	Ethnographic Research
3.1.3	Observation Research
3.1.4	Critical Theory
3.1.5	Hermeneutics
4.0	Conclusion
5.0	Summary
6.0	Tutor-Marked Assignment
7.0	References/Further Reading

1.0 INTRODUCTION

Welcome to Module 2 of this course which now looks at three basic types of educational research, i.e. humanistic, survey and correlation research. Humanistic research is so called because it addresses problems or needs that are relevant to people, and is therefore based on a holistic understanding of people and their activities (i.e. their physical, rational, psychological, social and ethical needs).

Humanism is, as a rule, narrative, textual, rhetorical, value – laden, and judgmental (as, for instance, in historical disciplines): It is also idiographic – interested in individual facts or stories – rather than nomothetic. In humanistic research, descriptions very often do matter more than explanation, i.e. relies more on intuitive rather than deductive approach, and, unlike some other forms of human inquiry, is not pursued for its own sake but in order to contribute to society. That is because humanism believes in human rationality, creativity and morality, and recognizes that human values have their source in experience and culture. It emphasizes that all people have the ability to lead meaningful lives. People acquire purpose in life through developing talents and

using them for the service of humanity. Humanistic research should be designed to provide tools and services that empower and enable people themselves to address their social, rational and emotional needs.

Various subdivisions of humanistic research include historical, ethnographical, observation, critical theory and hermeneutics research. These will now be considered in turn in this unit.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- identify the various types of humanistic research
- describe historical research with examples
- compare and contrast ethnographic research with observation research
- evaluate critical theory as a paradigm of educational research
- explain hermeneutics theory in educational research.

3.0 MAIN CONTENT

3.1 Humanistic Research

3.1.1 Historical Research

Historical research attempts to describe and explain conditions, situations and events of the past. An example would be a study that documents the evolution of vocational training programs in Nigeria 1951-1990, with the aim of explaining the historical origins of the content and processes of current programs. Historical research generally relies on qualitative data such as written documents and oral histories, although some quantitative data is also used. Its two main sources of data are primary and secondary sources; primary sources being firsthand knowledge (e.g. eye witness reports, original documents, tape and video recordings, etc.), while secondary sources would be second hand information (e.g. description of an event by a non-eye witness.)

Like in other descriptive research, the process of historical research involves identification of the problem concretized in form of questions and hypotheses, the collection and evaluation of source materials (subject to external and internal criticism), and synthesis of the information in terms of analysis, interpretation and formulation of conclusions. Its validity can be measured in two ways. External validity is establishing that a historical document is genuine, while internal

validity can be obtained through searching out biases in personal accounts of events, diaries, and autobiographies.

According to Kaestle (1988), some of the problems associated with quantitative historical research would include:

- the unfamiliarity of statistics and computers to many historians, apart from their time consumption and expense.
- the average audience for historical research is not particularly inclined to methodological sophistication;
- often, the data are crude, incomplete and mostly cross-sectional, providing only a snapshot of a group at a given moment. But education is a process, and to infer process from cross-sectional data or static information is often defective. For example, many important questions about educational careers, or the influence of education in peoples' lives, can be answered only by data that trace individuals over time, i.e. longitudinal studies. But it is difficult to create time series on educational variables like school attendance, teachers' salaries, educational expenditures, or length of school year because often the items were defined differently in different periods or omitted altogether;
- it is also difficult to know how conscientiously the data were reported in the first place or what biases operated.
- the frequent temptation to confuse correlation and causation, common of course to many other studies, remains a problem. This is compounded by the problems and vagueness of presentism in the definition of key terms – “presentism” meaning the danger of investing key terms from the past with their present connotations, or, conversely, applying to past developments present-day terms that did not exist or meant something else at the time.
- the problem of distinguishing between intent and consequences of historical action. Intent can only be inferred from consequences if there is direct evidence of this at the time the event occurred.

SELF – ASSESSMENT EXERCISE

- i. What do you understand by “presentism” in historical research?

3.1.2 Ethnographic Research

The ethnographic method is qualitative and holistic, making use of the investigator's intuition, empathy and general ability to learn another culture. It relies on rationalistic, qualitative techniques, especially

observation and careful recording of events and social interactions. It attempts to describe group behaviour and interactions in social settings, with particular reference to the interaction of individuals in the context of the socio-cultural norms, rituals and beliefs shared by the group. The researcher generally participates in some part of the normal life of the group and uses what he or she learns from this participation to understand the interactions between the group members.

Participant observation has its advantages and drawbacks. On the one hand, the investigator develops personal knowledge about the rules of the group and begins to perceive the same meanings in events as do the members of the group. Thus he becomes acculturated to the 'silent language' of the group since he shares, probes, asks questions, takes notes, etc. On the other hand, the investigator may become so much absorbed in the life of the group that his or her status as an observer may be compromised. In any case, the observer inevitably influences the behaviour of the people being studied, and this has to be taken into account in the research outcome in terms of reliability and validity of the study. One of the procedures often followed to confirm validity of interpretation in ethnographic studies is to feed them back for comment to selected members of the group or to other persons who know the group, but the application of this procedure is limited.

3.1.3 Observation Research

Observation has been defined (Hassan, 1995) as a systematic procedure of studying and understanding of human behaviour in its natural setting. It is best used for studying aspects of behaviour that cannot be studied through more objective methods, and involves watching, noting and recording of behaviours occurring in a natural setting over a specified period of time which may be long or short, depending on the nature and purpose of the observation.

Observational data are valuable, for they provide hunches and guesses about what is and is not important in any area of scientific interest. They also serve as a common sense check on the conclusions reached in more controlled investigations.

Shetzer and Linden (1979) classified observation methods into systematic, controlled and informal, involving direct observation by time or event sampling, the use of interview and contrived situational tests, as well as anecdotal observation and sociometric methods respectively. Systematic observation is most commonly used in the studies of young children.

SELF – ASSESSMENT EXERCISE

- i. Distinguish between ethnographic and observation research.

3.1.4 Critical Theory

Critical theory as an approach to educational research has a distinctive political orientation. According to Habermas (1976b), critical theory suggests that the current dominance of science and the rise of technology and bureaucracy are developmental tendencies of late capitalism which increasingly encroach on the domain of social life. Science and technology have been shown to be ideological, and the advocates/proponents of critical theory contend that it is a more socially just educational theory and practice. It is claimed that critical theory transcends the distance between the dominant positivist school and its challenger, the interpretivist paradigm, in a higher order synthesis (Lakomski, 1988).

The two central doctrines of critical theory are: (1) the conception of human interests, which provides justification of the theory as knowledge, and (2) the notion of communicative competence which, according to Habermas (1974), solves the problem of ideological domination. The victory of scientism or positivism, which presents itself as the only valid form of knowledge, has made it virtually impossible to reflect critically on current forms of domination. Interests, according to Habermas (following Marx), are grounded in the fundamental human conditions of work and interaction. Habermas argues that just as human beings produce and reproduce themselves through work, so they shape and determine themselves through language and communication in the course of their historical development. “Communicative competence,” Habermas contends, eludes us since there is no complete symmetry of power among the partners of communication, and even the language we use to reach consensus is itself a carrier of ideology.

Critical theory is based upon the so-called “theory vs. practice problem” (or “problematic,” in the language of Tripp (1992)) which holds that traditional (positivist) theory is incapable of properly informing and guiding practice. The concept of ‘socially critical’ research in education, or the ‘critical paradigm’ in educational research is anchored on the idea of a more just society in terms, not just of all people having equal access to the good things of life, but of people being in cultural, economic and political control of their lives. Tripp (1990a) defines critical theory in education as follows:

Socially critical research is informed by principles of social justice, both in terms of its own ways of

working and in terms of its outcomes in and orientation to the community. It involves strategic pedagogic action on the part of classroom teachers aimed at emancipation from overt and covert forms of domination. In practical terms, it is not simply a matter of challenging the existing practices of the system, but of seeking to understand what makes the system be the way it is, and challenging that, whilst remaining conscious that one's own sense of justice and equality are themselves open to question. (p.161)

How things are is never seen as having occurred by chance and for no particular reason; all social systems and their practices are seen to be as they are in order to serve the interests of particular groups.

Critical theory ('criticalism') focuses on the autonomy of the individual and is one of the four main research paradigms in educational research, the others being post-positivism, interpretivism and postmodernism. The critical paradigm promotes the notion of social justice in order to create the world which is "fairer, more equitable, more inclusive and more harmonious (Taylor, 2008). It is concerned with the power and justice of several issues in society such as the economy, race, gender and education. It considers the power of social politics and ideology which influence educational research. According to Morrison (1995), "the task of the researcher is not to be dispassionate, disinterested and objective, but the main objective of critical theory is to improve existing situations through action research". In critical theory, practical issues can construct knowledge, and the critical theory paradigm tends to see the world as something that has to change.

Critical theory has been applied in the areas of curriculum theory, educational admin, and action research.

SELF – ASSESSMENT EXERCISE

- i. What are the critical points in the critical theory of educational research?

3.1.5 Hermeneutics

Hermeneutics is the theory of interpretation and understanding in different kinds of human contexts, religions as well as secular, scientific as well as that of everyday life. The purpose of hermeneutics is to increase understanding of other cultures, groups, individuals, conditions, and lifestyles, both in the present as well as the past.

In biblical exegesis, two traditional studies of interpretation were (1) literal interpretation which spells out the messages as more or less explicit in the text itself, and (2) symbolic content interpretation which attempts to reconstruct or create the divine meanings which were not there in a literal sense. In education, the hermeneutical approach involves the study of meaning and comprises of theory and practice of interpretation and understanding in the different social contexts in which human beings live and work. Interpretation consists of two elements, one grammatical (largely in terms of the linguistic principles of a work) and one psychological (i.e. analyzing a text in terms of the thoughts and feelings of the author through identification with or transformation into the other person).

Essentially, hermeneutics involves cultivating the ability to understand things from somebody else's point of view, and appreciating the cultural and social forces that may have influenced their social outlook. Hermeneutics is the process of applying this understanding to interpreting the meaning of written texts and symbolic artifacts (e.g. sculpture and architecture) which may be either historic or contemporary.

SELF-ASSESSMENT EXERCISE

- i. In the hermeneutical approach, how is the ability to understand things from someone else's point of view exemplified?

4.0 CONCLUSION

The various forms of humanitarian research discussed in this Unit pertain to finding out about the human being from a humanistic, naturalistic, holistic and descriptive perspective. Generally, they rely more on qualitative or non-numerical data rather than numerical data, although some quantitative data is also sometimes used, with their accompanying analytical methods. The bottom line of humanistic research is studying the human being from a holistic, not a fragmented dimension.

5.0 SUMMARY

In this Unit you have learnt that the focus of humanistic research is on the human being in his/her totality. Some of the forms of this kind of research were addressed. The meanings, methodologies, uses, and problems of historical, ethnographic, observation, critical theory and hermeneutics research were explained, and the differences between them and the more objective procedures of applied/quantitative research were pointed out. The next unit deals with Survey Research.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Elaborate on the 'distinctive political orientation' of critical theory.
- ii. How can critical theory be concretized in the typical classroom?

7.0 REFERENCES/FURTHER READING

Fiere, P. (1972). *Cultural Action for Freedom*. Penguin, Harmondsworth.

Habermas, J. (1972) *Knowledge and Human Interests*. Heinemann, London.

Kinchelow & Mclauren (2002). Rethinking critical theory and qualitative research. In Zou, Y.E Trueba, E.T. *Ethnography and Schools: Quantitative approaches to Study of Education*, 87 – 125. Maryland: Rowman and Littlefield Publishers.

Koul, R.B. (2008). Educational research and ensuring quality standards. *E-journal of All-India Association for Educational Research*, (2003 and 2004). <http://www.ejournal.iaer.net/vol20208/5.htm>

Okoh, N. & A.B. Oduaran (1984): *Consientization as an awareness process for development in Nigeria*. *Jul. of Education in Developing Areas*, 1,3, University of Port Harcourt, Nigeria.

Tripp, D. (2001). Critical theory and educational research. *Issues in Educational Research*, 2001, 1992, 13-23. <http://education.curtin.edu.au/ier/ier2/trip.html>.

UNIT 2 SURVEY RESEARCH

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Survey Research
 - 3.1.1 Cross-sectional Survey
 - 3.1.2 Longitudinal Survey
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Survey research is one of the most important and most common types of research in applied social research like educational research. The educational research student must therefore be familiar with this broad area of research, its methodologies, instrumentation, its design and procedures, and its pitfalls. This unit introduces you to the what, the how and the why of surveys.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define survey research
- state what factors to consider in designing a valid questionnaire
- describe the various types of survey
- examine the different sampling methods used in data collection for surveys
- choose the correct tools for analysis of the data obtained
- list the threats to internal and external validity of a survey instrument.

3.0 MAIN CONTENT

3.1 Survey Research

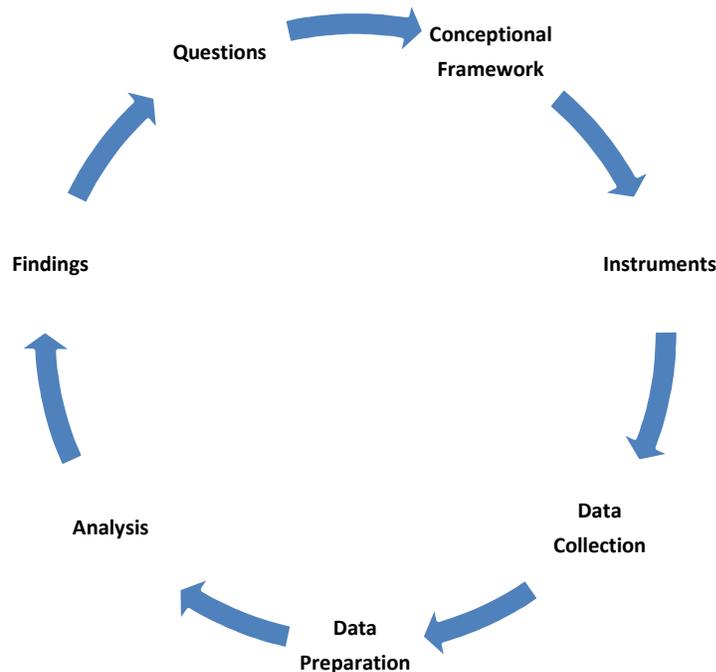
The survey method is a non-experimental method commonly used in descriptive research in education. It encompasses any measurement procedures that involve asking questions of respondents, and it can take any form from a short paper and pencil feedback form to an intensive one-on-one, in-depth interview. It involves the collection of information

from members of a predefined finite population, e.g. a group of students, teachers or other stakeholders in education, and the analysis of this information to illuminate some important educational issues. The information is usually collected by means of questionnaires and tests, although sometimes also by means of observation schedules and interviews. Most surveys use samples of a specified target population with a view to generalising the results to the populations from which the samples were drawn.

Surveys may be used to obtain descriptive information about a target population, (e.g. to measure levels of literacy or numeracy in a school or region), or to examine relationships between various factors (e.g. to explain the differences in mathematics achievement of students in terms of their age, sex, exposure to the mathematics curriculum, and amount of time spent in class learning mathematics) (Rossier 1980).

In educational research, use of the survey method may range from small-scale exploratory survey to stimulate hypothesis generation, to large-scale exploratory survey to obtain evidence in major policy decision making.

The starting point for a survey is the statement of the questions to be addressed by the investigation, followed by the conceptual framework. Next is the development of the instruments for data collection, followed by the data analysis and presentation of findings. Figure 1 below (Runket and Mcgrath, 1972) summarizes the typical survey research cycle.

Figure 1: Typical Survey Research Cycle

The questions must be pertinent to real problems, and the conceptual framework may be derived from a set of propositions which will be examined at the analysis stage.

SELF –ASSESSMENT EXERCISE I

- i. What do the instruments used in survey research attempt to measure?

Instrumentation usually takes the form of existing or constructed questionnaires, tests, attitude scales, etc. Instruments generally attempt to measure four broad categories of: (1) background characteristics of subjects (e.g. sex, educational level of parent, age of teacher, school enrolment, etc), (2) knowledge (e.g. knowledge of science concept), (3) attitudes and opinions (often measured by responses to a range of statements on a topic), and (4) direct observation of behaviour (e.g. of students and teachers during a classroom lesson).

Sampling design and procedure are very important in surveys. Most sampling plans assume random sampling which makes it possible to use data derived from the sample to estimate statistical characteristics of the population. The art of sampling lies in preparing a design to minimize sampling error which depends primarily on the size and structure of the sample.

The two random sampling types most commonly used are (1) stratified random sampling where the target population is divided into strata such as types of school and the areas or regions; and (2) multistage sampling involving for example the initial random selection of areas or regions in the first stage, the selection of schools in the second stage, and the selection of students on the third stage. The response rates to questionnaires should be calculated and where necessary weighting procedures based on the size of the target population and the size of the achieved sample may be used to take account of low response rates.

Analysis of survey data may be in the form of univariate analysis which deals with the variables individually, bivariate analysis which involves pairs of variables, or multivariate analysis which is concerned with the simultaneous effects of more than two variables.

There are two main types of survey. The most common is “cross-sectional survey” which involves collection of data at a specific point in time and mostly for the purpose of describing situations and estimating frequencies rather than establishing causal patterns. In contrast, “longitudinal surveys” involve following a particular group of students or schools over a period of time.

SELF –ASSESSMENT EXERCISE

- i. List and discuss two of the most commonly used random sampling techniques used in survey research.

3.1.1 Cross-sectional Survey

Cross-sectional surveys are used to gather information about present conditions in a population at a single point in time or during a single, relatively brief period, and then comparisons are made across the variables of interest. Generally, the purpose is to describe situations and estimate frequencies rather than to establish causal patterns. An example of cross-sectional survey would be a questionnaire that collects data on how parents feel about school excursions as of September 2009. A different cross-sectional survey questionnaire might try to determine the relationship between two factors like the social class of parents and their views of school excursions, or a comparative study of the educational facilities in private and public primary schools.

3.1.2 Longitudinal Surveys

In longitudinal surveys, the data are collected at more than one point or data collection period, and the researcher is interested in making comparisons across time. The data can be collected on one or more

multiple groups. Repeated observation for at least two points in time is the key characteristic of the longitudinal method since this enables the educational researcher to study the processes and patterns of change and stability in the educational field.

The logic of longitudinal studies is that educational researches are concerned with the process of change, and the study of change requires that observations are made for at least two points in time. Time is significantly related to causal influence, since earlier events usually influence later events.

In longitudinal studies where time plays a key role, the temporal order in which data are gathered on predictor variables with respect to the criteria measures is very important, and constitutes one threat to internal validity. Other threats to internal validity are history (current or past), maturation (physical/psychological), practice effects in testing, reliability of instruments, subject instability, composition of samples, etc. With respect to external validity, the Hawthorne effect, population validity, multiple intervention interference, critical changes in individual's life (e.g. road accident or family break-up), publication of interim reports, etc., provide possible distortions into the data and therefore interpretations.

SELF –ASSESSMENT EXERCISE III

- i. Write a brief description of Cross-Sectional Surveys.

4.0 CONCLUSION

The importance and widespread use of the survey method in research qualifies it for detailed treatment in its many aspects. From this unit, the student would have seen how broad an area survey research is, and how it can be used not only to gather descriptive information but also to test hypothesis arising from studies of smaller, selected groups, and to seek explanations for relationships that have been found between variables. It must also have been noticed that, by their very nature, surveys cannot definitively establish causal explanations on the basis of the empirical data collected. The last will also be taken up in the next unit.

5.0 SUMMARY

In this unit you have learnt that survey research

- involves collection of information from a sample of a specified target population, and the analysis of this information to illuminate some important educational issues;

- uses generally questionnaires and tests, and occasionally interviews and observation schedules, in its instrumentation.
- cross-sectional survey gathers information about present conditions in a population at a single point in time, and comparisons are made across the variables of interest.
- a longitudinal survey collects data at more than one point in time and the comparisons are made across time.
- stratified random sampling and multistage sampling are the most commonly used sampling types.
- processing of survey data may be in the form of univariate or multivariate analysis to confirm or negate the stated hypotheses.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Give two examples of survey research, including their design, sampling method, appropriateness of procedures used, and analysis of data.
- ii.
 - (a) Distinguish critically between cross-sectional and longitudinal survey.
 - (b) Elaborate on how to minimize threats to internal validity in longitudinal surveys.

7.0 REFERENCES/FURTHER READING

Babbie, E.R. (1973). *Survey Research Methods*. Belmont, CA: Wadsworth Publishing Co.

Campbell, D.T and Stanley, J.C. (1963). Experimental and quasi-experimental design for research on teaching. In Gage, N.L. (ed.), *Handbook of Research on Teaching*. Chicago, Illinois: Rand McNally.

Rossier, M.J. (1985). Survey Research Methods, in J.P. Keeves (ed.) *Educational Research, Methodology and Measurement: An International Handbook*. Oxford: Pergamon Press.

UNIT 3 CORRELATION RESEARCH

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Correlation Research
 - 3.1.1 Description
 - 3.1.2 Methods
 - 3.1.3 Data analysis
 - 3.1.4 Limitations
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

In the last unit on survey research, you will remember that it was pointed out that surveys may also be used to examine the relationship between various factors in a multi-factor survey. This is an aspect of the construct of correlation which will be considered more fully in this unit.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- state what correlation is
- give some examples of your own correlational research studies
- exemplify how the results of such studies are analysed and interpreted
- explain why “correlation does not imply causation”.

3.0 MAIN CONTENT

3.1 Correlation Research (See Module 6, Unit 3, for more detailed treatment)

3.1.1 Description

Correlation research attempts to explore a non-cause and effect relationship between two or more variables, through the use of various measures of statistical association. It relies on quantitative or numerical data such as test scores, grade point averages, scores from attitudinal

instruments, etc. which can be correlated and shown that some relationship exists between or among them.

Some examples of correlational questions are:

- Do female students do better than male students in arts subjects in secondary schools?
- Is the educational achievement of primary school children who enjoy parental help with their homework better than that of children who do not have such parental help?
- Can an aptitude test in Mathematics be used to predict success in an algebra course?

3.1.2 Methods

Some common types of correlational methods include

- (i) Observation: This is useful for describing behavior and for suggesting causal hypotheses that could be tested in experiments. Scientific observation should (a) be objective (just the facts and avoid anthropomorphism), (b) systematically record data to avoid memory biases and errors, (c) use good and fair sample in order to be able to generalize results: But observation has its own difficulties, e.g., it can sometimes be impossible (thoughts?), impractical (robbery, fire, or earthquake?), or unethical (spying?).
- (ii) Surveys and tests, which suffer from the following problems: (a) social desirability bias (i.e. people may try to make themselves look better than they are, or say what they think the experimenter wants to hear); (b) memory lapses (people do not accurately remember what they do); (c) people don't always know why they do things; (d) bad questions (e.g. leading questions and confusing questions); (e) poor sampling (a good sample should wherever possible be large in size, but more importantly should be randomly selected).
- (iii) Case Study, which examines some phenomenon in depth e.g. people, programs, policies, decisions, organizations, etc. A case study weaves together data from documents, archives, interviews, participation, observation, artifacts, etc, and attempts to document not only the 'what' but also the 'why'. It may also suffer from (a) small, non-random sample, (b) lack of control group which means one cannot make cause-effect statements.

3.1.3 Data Analysis

In correlation studies, product moment correlations or cross tabulations indices are usually calculated. There are statistical tests which can be applied to determine whether the association is more than would occur by chance.

The most commonly used and understood correlational index is the product moment correlation coefficient where both variables are considered to be single variables, e.g. the association between scores on a reading readiness test and reading achievement. Another such product moment correlation coefficient index is the canonical correlation coefficient between two unobservable (latent) variables. In cases where one or both of the variables is dichotomous, or where the data are measured on the ordinal scale, the point-biserial correlation (rpb) and the Spearman rank order correlation (rho) respectively are used. The formulae for computing these various correlation coefficients can be found in any elementary to middle-level statistics textbook. To control or remove the effect of one or more confounding variables statistically, e.g. the effect of age or gender or intelligence level, this can be done at the sampling stage through partial or semi-partial correlation, although this often has the problem of a negative effect on the generalisability of the findings.

3.1.4 Limitations

But it is always important to point out that association between two or more variables does not necessarily imply that one of the variables causes the other. Correlation implies prediction but not causation. Intervening variables may account for the causation, and this is in fact the distinctive characteristic and the strength of the experimental paradigm. When two variables are correlated, you can use the relationship to predict the value on one variable for a subject if you know that subject's value on the other variable. Variables that are highly related may suggest the possible existence of causation, by experimental studies to determine if the relationships are causal.

SELF – ASSESSMENT EXERCISES

- i. Give some examples of correlation research.
- ii. Why can causal inferences not be drawn from correlation research results?

4.0 CONCLUSION

Correlation research is quite common in educational research. One of the main uses of correlation research is to highlight topics for further controlled experimental study. The results of a correlation study are most times useful and intriguing, but the major drawback is the researcher's inability to draw causal inferences from these results.

5.0 SUMMARY

In this unit, you have learnt, with some examples, what correlation essentially is, its usefulness in educational research, the statistical tools used in analyzing and interpreting results from correlation studies, and some of its limitations in educational research.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Discuss the view that correlation research is an inferior form of educational research.

7.0 REFERENCES/FURTHER READING

- Hassan, T. (1995). *Understanding Research in Education*. Lagos: Merrifield Publishing Co.
- Keeves, J.P. (ed.) (1990). *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford: Pergamon Press.
- Kerlinger, F.N. (1973). *Foundations of Behavioural Research*, Second Edition. New York: Holt, Rinehart and Winston.
- Nwana, O.C (1982). *Introduction to Educational Research*. Ibadan: Heinemann
- Nyanjui, P.J. (2006). *Introduction to Research*. Nairobi: Kenya Institute of Education.
- Shaughnessy, J.J. (2006). *Research Methods in Psychology*. New York: McGraw-Hill.

MODULE 3 DESIGNS OF RESEARCH I

Unit 1	Causal Comparative Research
Unit 2	Quasi-Experimental Research
Unit 3	Experimental Research
Unit 4	Factorial Designs

UNIT 1 CAUSAL COMPARATIVE RESEARCH

CONTENTS

1.0	Introduction
2.0	Objectives
3.0	Main Content
3.1	Causal Comparative Research
3.1.1	Description
3.1.2	Procedure and Analysis
3.1.3	The debate/controversy
4.0	Conclusion
5.0	Summary
6.0	Tutor-Marked Assignment
7.0	References/Further Reading

1.0 INTRODUCTION

Among the designs of research, causal comparative research is very commonly used since it does not involve the more tedious processes of manipulating the independent variables and the random assignment of subjects to treatments. In many real life situations the true experimental design is hardly possible/desirable or very difficult.

Causal comparative research therefore, as a non-experimental research design, comes in handy and useful in educational research, even though it cannot attain the precision of the true experiment.

2.0 OBJECTIVES

After working through the unit, you should be able to:

- describe what causal comparative research is
- give some examples of causal comparative research
- adumbrate the procedures and data analytic methods of the design
- state the strengths and weaknesses of the design

- comment critically on the differences between causal comparative research and correlation research

3.0 MAIN CONTENT

3.1 Causal Comparative Research

3.1.1 Description

Causal-comparative research attempts to establish cause-effect relationship among the variables in a study, or determine the cause or reason for pre-existing differences in groups of individuals. The attempt is to establish that values of the independent variable have a significant effect on the dependent variable. This type of research typically involves comparison of two or more groups and one independent variable. The comparison groups in the study make up the values of the independent variable. For example, the study may be to find out the effect of gender, pre-school attendance, and working-mother status on dependent variables like physics achievement, social maturity at the beginning of primary school, and school absenteeism respectively. The critical factor in causal-comparative research is that the independent variable is not under the experimenter's control, e.g. the experimenter obviously cannot randomly assign the subjects to age or gender classification (male or female) but has to take the values of the independent variables as given.

Causal-comparative research is also often referred to as ex-post facto (Latin for "after the fact"), i.e. research conducted to find causes of events or conditions that have already occurred. Both the effect and the suspected cause have already occurred and must be studied in retrospect.

The basic causal-comparative research approach (retrospective causal comparative research) in education involves starting with an effect and seeking possible causes. Individuals are not randomly assigned to treatment groups because they already were selected into groups before the research began. Independent organismic variables like age and gender, cannot be manipulated, or, for ethical reasons, should not be manipulated. For example, it would be highly unethical, if at all possible, to try to manipulate anxiety level, pain and stress levels in subjects.

Caution therefore must be exercised in interpreting results of causal comparative studies since the alleged cause of an observed effect may in fact be the effect itself, or there may be a third variable. For example, if a researcher hypothesized that reading achievement is a function of self concept, and proceeded to identify high and low self-concept groups and, after comparison, found that the high self-concept group did in fact

show higher reading achievement, it would not be warranted to conclude that high self concept results in high reading achievement. Lack of randomization, manipulation, and control are sources of weakness in causal comparative research. Only in experimental research does the researcher randomly assign participants to treatment groups, and the degree of control is sufficient to establish cause-effect relationship.

3.1.2 Procedure and Analysis

Causal-comparative research is non-experimental research; the independent variable is not manipulated and subjects are not randomly assigned to treatments. In causal-comparative research studies, the researcher selects two groups of participants, experimental and control groups (more accurately referred to as comparison groups). The groups differ in either the possession or degree of possession of a characteristic. The goal is to have groups that are as similar as possible on all relevant variables except the independent variable. The performance of the groups is then compared using some valid dependent variable measure or instrument.

The researcher may control for an identified extraneous variable by random assignment of participants to groups or by pair-wise matching of participants, i.e. the researcher matches each participant in one group with another participant in the other group with the same or very similar score on the control variable. Or, where this is not possible, he/she may eliminate the unmatched participant altogether. However, this procedure may lower the number of participants and so limit the generalisability of the findings.

Data analysis is by use of descriptive statistics (mean, SD), as well as inferential statistics (t-test, chi-square, and analysis of variance). The use of factorial analysis of variance will allow the researcher to determine the effect of the independent/control variable on the dependent variable, both separately and in combination. Analysis of covariance is also used to adjust initial group differences on the variables used so that the results can be fairly compared.

SELF-ASSESSMENT EXERCISE I

- i. Discuss the methods of data analysis appropriate to causal comparative research.

3.1.3 The Debate/Controversy

There is an ongoing controversy among researchers (Fraenkel and Wallen, 1996; Gay, 1996; Charles, 1998; Cook and Campbell, 1979;

Johnson and Christensen, 2000; Johnson, 2000) not only about evidence of causality in both forms of research, but also about the datedness or otherwise of nomenclature of the terms “causal comparative” and “correlational.” Burke Johnson (2000), the leading advocate of this debate, strongly debunks the suggestions that (1) causal comparative research provides better evidence of cause and effect relationships than correlational research, (2) causal comparative research includes categorical independent and/or dependent attribute variable, while correlation research only includes quantitative variables. He argues that both methods are similar in that both are non-experimental methods because they lack manipulation of an independent variable which is under the control of the experimenter.

In any case, both causal comparative research and correlational research can and do use categorical independent variables. For example, causal comparative research uses some categorical independent attribute variables which are not or cannot be manipulated (e.g. gender, parenting style, learning style, ethnic group, parity identification, type of school, marital status of parents, degree of introversion, etc.), and similarly correlation research also uses categorical independent variables which cannot be manipulated (e.g. intelligence, aptitude, age, school, size, income, job satisfaction, GPA, degree of extroversion, etc.). Also, in both types random assignment of participants is not possible. So according to Johnson, *cet par*, causal comparative research is neither better nor worse in establishing evidence of causality than correlation research. In fact, in his view, both ‘outdated terms’ should be replaced by “non-experimental quantitative research”. The debate, it appears, is still not spent.

SELF-ASSESSMENT EXERCISE

- i. What problems arise in interpreting results of causal comparative research?

4.0 CONCLUSION

Causal comparative research is sometimes treated as a type of descriptive research since it describes conditions that already exist. However, it also attempts to determine reasons, or causes, for the current status of the phenomena under study, often not very successfully.

5.0 SUMMARY

This unit has introduced you to the meaning, methodological procedures and data analysis for causal comparative research. It also introduced you to the debate on the terminology and causality attribution of causal

comparative research. The following unit on quasi-experimental research will touch on this further.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Why is causal comparative research also known as ex-post facto research?
- ii. Comment critically on the causal comparative research versus correlation research controversy.

7.0 REFERENCES/FURTHER READING

Babbie, E. (1998). *The practice of social research*. 8th edition. Belmont, CA: Wadsworth.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Fraenkel, J.R., & Wallen, N.E. (1996). *How to design and evaluate research in education*. 3rd edition. New York: McGraw-Hill.

Gay, L.R. (1996). *Educational research: Competencies for analysis and application* 5th edn. Englewood Cliffs, NJ: Prentice-Hall.

Johnson, B. (2000). It's (Beyond) Time to Drop the Terms Causal-Comparative and Correlational Research in Education.
<http://techl.coe.uga.edu/itforum/home.html>

Kerlinger, F.N. (1986). *Foundations of behavioral research*. 3rd ed. New York: Holt, Rinehart & Winston.

UNIT 2 QUASI-EXPERIMENTAL RESEARCH

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Quasi-Experimental Research
 - 3.1.1 Description
 - 3.1.2 Characteristics
 - 3.1.3 Types
 - 3.1.4 Limitations
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Because of the complexity of human circumstances and behaviour, it is not often practical to carry out the traditional controlled experimental research. Therefore, quasi-experimental designs are widely used as a near and useful substitute in areas where it is not feasible or desirable to conduct an experiment or randomized control trial. What quasi-experimental designs are is the subject of this Unit which will discuss the advantages, characteristics, types and limitations of quasi-experimental research.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define quasi-experimental research
- distinguish between quasi-experimental research and experimental research
- differentiate the several types of quasi-experimental design
- explain some common characteristics of quasi-experimental research
- describe the advantages and limitations of this research design.

3.0 MAIN CONTENT

3.1 Quasi-Experimental Research

3.1.1 Description

As the name implies, quasi experimental research is an approximation of experimental research. Here, the researcher controls some but not all of the extraneous factors. An example would be a study in which there is a treatment intervention but no assumption of random assignment of subjects to treatments. A quasi-experiment is similar to a true experiment, in the sense that it has subjects, treatment, etc; but uses non-randomized groups.

Quasi-experimental designs are used when randomization is impossible and/or impractical, and therefore they are typically easier to set up than true experimental designs – since it takes much less effort to study and compare subjects or groups of subjects that are already naturally organized than to have to conduct random assignment of subjects. Also, threats to external validity are minimized since natural environments do not suffer the same problems of artificiality as compared to a well-controlled laboratory setting.

In some quasi-experimental designs, the researcher might have some control over assignment to the treatment condition but use some criteria other than random assignment (e.g. a cut-off score) to determine which participants receive the treatment, or the researcher may have no control over the treatment condition assignment, and the criteria used for assignment may be unknown.

Quasi-experimental designs are often chosen for field studies where the random assignment of experimental subjects is impractical, unethical or impossible. Therefore, quasi experiments are sometimes also called natural experiments and are common in many science and social science disciplines including economics, political science, geology, paleontology, ecology, meteorology and astronomy.

SELF-ASSESSMENT EXERCISE

- i. How is quasi-experimental research different from experimental research?

3.1.2 Characteristics

Some common characteristics of quasi-experimental design are: (i) matching instead of randomization is more often used, (ii) time series is

involved (a time series, whether interrupted or non-interrupted, examines changes in the dependent variable over time), and (iii) the unit of analysis is often something different than people, e.g. climate, atmosphere, anomie, morale, crime statistics, quality of life, etc.

In quasi experiments, the word “trend” is often used instead of cause. This is because this kind of research often uncovers several trends, with the major ones designated as syndromes or cycles, and the minor ones referred to as normal or abnormal events. It is advisable that the quasi-experimental design researcher engage in modeling the causal relationships, since the designs tend to involve many different but interlocking relationships between variables. This will allow identification of spurious and intervening variables. Many causal modeling techniques exist, from cross tabulation to partial correlation analysis to path analysis.

3.1.3 Types

There are several types of quasi-experimental designs, ranging from the simple to the complex. These designs include: (1) one-group posttest only, (2) the one group pretest posttest, (3) the removed treatment design, (4) the case-control design, (5) the non-equivalent control groups design, (6) the interrupted time-series design, and (7) the regression discontinuity design. Some of these will be treated in the next section.

3.1.4 Limitations

The deficiency of lack of random assignment makes it harder to rule out confounding bias and introduces threats to internal validity. Also, conclusions of causal relationships are difficult to determine due to a variety of extraneous and confounding variables that exist in a social environment which the researcher does not have total control of, e.g. factors such as cost, feasibility, political concerns, or convenience. However, such bias and other confounding variables, can be controlled for by using various statistical techniques such as multiple regression which can be used to model and partial out the effects of confounding variables.

On the whole, however, quasi-experiments are a valuable tool, especially for the applied researcher. Although on their own, it is difficult to draw definitive causal inferences from quasi-experimental designs, they do provide necessary and valuable information that cannot be obtained from experimental methods alone.

SELF-ASSESSMENT EXERCISE II

- i. Discuss some of the limitations of this mode of research design.

4.0 CONCLUSION

Quasi-experimental research is popular in education because of the many difficulties of conducting full blown experimental studies. Its deficiencies notwithstanding, quasi-experimental design is a very useful tool systematically studying many education phenomena, and its illuminative and suggestive findings should be utilized to the fullest.

5.0 SUMMARY

This unit has introduced you to the use of quasi-experimental research in addressing many educational problems. Its characteristics, advantages and limitations have been highlighted. The next unit on experimental research design is a natural follow-up.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Why do you think quasi-experimental research has been pejoratively labelled “queasy” research?
- ii. Justify with examples the use of quasi-experimental design in educational research.

7.0 REFERENCES/FURTHER READING

- Campbell, D. & J. Stanley. (1963). *Experimental Designs*. Chicago: Rand McNally.
- Cook, T. & D. Campbell (1979). *Quasi Experimental Design*. Chicago: Rand McNally.
- Hassan, T. (1995). *Understanding Research in Education*. Lagos: Merrifield Publishing Co.

UNIT 3 EXPERIMENTAL RESEARCH

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Experimental Research
 - 3.1.1 Description
 - 3.1.2 General Procedures
 - 3.1.3 Types
 - 3.1.4 Data Analysis
 - 3.1.5 Limitations
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Experimental research is unique in that it is the only type of research that directly attempts to influence a particular variable, and it is the only type that, when used properly, can really test hypothesis about cause and effect relationships. Experimental designs are some of the strongest available for educational researchers to use in determining cause and effect.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define experimental research
- explain the terms “causal relationship”, “matching”, “randomization”
- state the general procedures involved in conducting an experimental research
- describe a simple experiment
- distinguish between experimental and quasi-experimental research
- evaluate at least one of the types of experimental research.

3.0 MAIN CONTENT

3.1 Experimental Research

3.1.1 Description

Experimental research is used in settings where variables defining one or more ‘causes’ can be manipulated in a systematic fashion in order to discern ‘effects’ on other variables. In the typical experimental research design, the experimenter randomly assigns subjects to the groups or conditions that constitute the independent variable of the study and then measures the effect this group membership has on another variable, i.e. the dependent variable of the study. In other words, he/she compares the results obtained from an experimental sample against a control sample which is practically identical to the experimental samples, except for the one aspect whose effect is being tested (the independent variable). An example would be an investigation of the effectiveness of two new textbooks in geography teaching, using random assignment of teachers and students to three groups – two groups for each of the new textbooks, and one group as ‘control’ group (i.e. the group with no implemented treatment) to use the existing textbook. The research question would be “would students using Textbook A achieve better than students with Textbook B?”

The “true experiment”, when it is feasible, is a particularly effective way of investigating causal inference. This is due to the random assignment of subjects to treatments. Random assignment creates treatment groups which are initially comparable on all subject attributes, and so it can be concluded in an experimental study that any final outcome differences are due to the treatment alone – subject of course to control of other possible threats to validity, specifically statistical validity, internal validity, construct validity and external validity, (Campbell and Stanley, (1963), Cook and Campbell (1979). These can be taken care of in the design of the experiment, particularly the use of randomized block design and ANCOVA.

3.1.2 General Procedures

In experimental research, the following general procedures are usually applied:

- (1) random selection of control group;
- (2) control group must be statistically equal to the experimental group which is a randomly assigned treatment group;
- (3) both experimental and control groups must come from the same population;

- (4) you can have more than one control and treatment group;
- (5) all groups should have about the same number of subjects, and each group should have not less than 25 subjects as a general rule;
- (6) “Blind” experiment means that subjects do not know which group, i.e., the experimental or control group they have been assigned to. “Double blind” experiment means that the experimenter (and research helpers), as well as the subjects, do not even know who is in what group. These precautions help protect the study from Hawthorne effect and placebo effect.
- (7) findings should be interpreted primarily from differences in a posttest score between experimental and control groups.

3.1.3 Types

True experimental designs vary in complexity from (1) the simple experiment (classical pretest/posttest control group design), to (2) the randomized Solomon four-group design, to (3) the randomized posttest only control group design.

- (1) Classical pretest-posttest (Simple Experiment): The simple experiment is one of the most basic methods of determining if there is a cause effect relationship between two variables. A simple experiment utilizes a control group of participants who receive the treatment and another group who do not. The experimenter then compares the results of the two groups to determine if the treatment had an effect. In this design, the total population of participants is randomly divided into two samples: the control sample and the experimental sample. Only the experimental sample is exposed to the manipulated variable. The researcher then compares the pretest results with the posttest results for both samples. Any divergence between the two samples is assumed to be a result of the experiment.
- (2) Solomon — four group design: Here, the population is randomly divided into four samples, of which two are experimental samples. Two groups experience no experimental manipulation of variables, two groups receive a pretest and a posttest, and the other two groups receive only a posttest. This tends to control for the effect of the pretest.
- (3) Randomised Posttest-Only Control Group Design: Sometimes also referred to as the After-Only Research Design, this design compares the scores of the experimental group (which receives the treatment condition) and that of the control group (which does not) on the dependent variable. Any significant difference between the

two scores is then attributed to the treatment condition. This design is very useful when pretesting is not possible, or where a pretest may interact with the independent variable.

SELF-ASSESSMENT EXERCISE I

- i. Compare and contrast the simple experiment to the randomized posttest only control group design.

3.1.4 Data Analysis

Experimenters use inferential statistics to determine if the results of an experiment are meaningful. Inference is a branch of science that deals with drawing inferences about a population based upon measures taken from a representative sample of that population. The key to determining if a treatment had an effect is to measure the statistical significance of the difference in the group results, which may show that the observed relationship between the variables is probably not due to mere chance but that a real causal relationship likely exists between the two variables. Statistical significance is often represented like this: $P < .05$, which indicates that if the particular result obtained is due mainly to chance the probability of a causal result would be less than 5%. Stronger significant results would be $P < .01$ and $P < .001$.

3.1.5 Limitations

The goal of an experiment is causal inference. If a treatment variable is manipulated or changed while holding constant all other possible determinants of the outcome of interest, and there is a subsequent change in the outcome, then it can be concluded that there is a causal relationship between the manipulated variable and the outcome. But this does not imply that the manipulated variable is the sole cause of the outcome, only that it is one of probably many causes.

An experimental study can be conducted either in a laboratory or in 'real' field settings. However, it is often difficult to conduct true experiments in educational settings for a number of reasons, including the reluctance of the concerned educational authorities ('administrative gatekeepers') to permit the researcher to use 'randomisation' as fully as is required for a valid experiment.

SELF-ASSESSMENT EXERCISE

- i. What are the limitations of the experimental research method?

4.0 CONCLUSION

Having worked through the Unit, you would have seen that experimental research, which is at the apex of research designs, is quite practical to do, subject to the required hard work and care. In fact, experimental research is not only very valuable but can be quite satisfying, and it is hoped that as a doctoral student you will be able to carry out a feasible experimental research at or before the conclusion of the course.

5.0 SUMMARY

Experimental research designs are used for the controlled testing of causal processes. The general procedure is that one or more independent variables are manipulated to determine their effect on a dependent variable. One of its most important requirements is the necessity of eliminating the effects of spurious, intervening, and antecedent variables. Some of the most common ways to control for these extraneous/confounding factors include randomization, holding certain variables constant, building the variable into the design, matching, using subjects as their own controls, and the statistical technique of ANCOVA.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Discuss critically the statement that “experiments are the top guns of research design, the most expensive and powerful techniques available in educational research”?
- ii. Design an experiment to test the hypothesis that “belief in witchcraft causes poor academic performance of students in secondary schools”.

7.0 REFERENCES/FURTHER READING

Campbell, D. & J. Stanley. (1963). *Experimental and Quasi-Experimental Designs*. Chicago: Rand McNally.

Cook, T. & D. Campbell. (1979). *Quasi-Experimental Design*. Chicago: Rand McNally.

<http://en.wikipedia.org/wiki>

Tate, R. (1990). *Experimental Studies*, In J.P. Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford, England, Pergamon Press.

UNIT 4 FACTORIAL DESIGNS

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Factorial Design
 - 3.1.1 Description
 - 3.1.2 Types
 - 3.1.3 Uses
 - 3.1.4 Advantages and Limitations
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

The term “factorial” implies that the design involves several factors. Factorial design therefore studies the joint effects (singly and interactively) of two or more independent variables (or factors) on a given dependent variable. Educational researchers often use factorial designs to assess the effects of educational methods on say academic achievement, whilst taking into account the influence of some economic factors and background.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define a factorial design
- describe the main types of factorial design
- illustrate the use of a factorial design by a concrete example
- outline the advantages and limitations of factorial designs.

3.0 MAIN CONTENT

3.1 Factorial Designs

3.1.1 Description

Traditional experimental research methods generally study the effect of one variable over time, because it is statistically easier to manipulate a single independent variable or factor. However, in many cases two

factors may be interdependent, and it is impractical or false to attempt to analyze them in the traditional way. For this reason, factorial experiments, which allow subtle manipulations of a larger number of interdependent variables, are often used.

The design of a factorial experiment consists of two or more factors, each with discrete possible values or “levels”, and the experimental units take on all possible combinations of these levels across all such factors. Such an experiment makes it possible to study the effect of each factor on the response variable, as well as the effects of interactions between factors on the response variable. In factorial designs, a factor is a major independent variable and a level is a subdivision of a factor. An interaction effect exists when differences on one factor depend on the level you are on another factor, i.e., if the effect of one independent variable is different under one level of another independent variable than it is under another level of the other independent variable. It is important to recognize that interaction occurs between factors, not levels.

3.1.2 Types

There are four general types of factorial design: (i) randomized groups design in which participants are assigned randomly to two or more groups, (ii) matched-subjects design in which participants are first matched into blocks, then randomly assigned to conditions, (iii) repeated measures design in which each participant serves in all experimental conditions, and (iv) mixed factorial design which combines manipulated independent variables and measured participant variables.

The simplest factorial experiment contains two levels for each of two factors. This is the common 2×2 factorial experiment, so named because it considers two levels for each of the two factors, producing $2^2 = 4$ factorial points. A factorial experiment can be analysed using regression analysis. The main effect for a factor is relatively easy to estimate. For example, to compute the main effect of a factor “A”, you subtract the average response of all experimental runs for which A was at its low (or first) level from the average response of all experimental runs for which A was at its high (or second) level.

SELF-ASSESSMENT EXERCISE

- i. Outline and evaluate the main types of factorial design.

3.1.3 Uses

Factorial designs are frequently used by psychologists and field scientists as a preliminary study, allowing them to judge whether there is

a link between variables, whilst reducing the possibility of experimental error and confounding variables. The factorial design, as well as simplifying the research process and making research cheaper, allows many levels of analysis. Apart from highlighting the relationship between variables, it allows the effects of manipulating a single variable to be isolated and analyzed singly.

3.1.4 Advantages and Limitations

Factorial designs, also called multivariate designs, have several important advantages: (i) they have great flexibility for exploring or enhancing the treatment in experimental studies, (ii) they are efficient in the sense that we are able to combine a series of independent studies into one, i.e., several hypotheses may be tested simultaneously, rather than having to conduct a series of single independent variable experiments to study the effects of different treatments on a dependent variable, (iii) they are the only effective way to examine interaction effects between and among different independent variables.

The major drawback of factorial design is the difficulty of experimenting with more than two factors, or many levels, at the same time. For this reason, a factorial design needs to be planned meticulously, for an error in one of the levels, or in the general operationalisation, might jeopardize a great deal of work.

SELF-ASSESSMENT EXERCISE

- i. Discuss some of the advantages, as well as limitations, of factorial design.

4.0 CONCLUSION

Factorial experimental designs are used when you have two or more independent variables and want to know if an interaction is present. They provide information about the effects of each independent variable by itself (the main effects), as well as the combined effects of independent variables (interaction effects).

5.0 SUMMARY

Factorial design is used to investigate the joint effect of two or more independent variables on a dependent variable. More specifically, in factorial design, two or more independent variables are simultaneously studied to determine their independent and interactive effects on the dependent variable. For the vast majority of factorial experiments, each factor has only two levels. The number of experimental runs required for

three-level (or more) factorial designs would be much less manageable, and therefore much less attractive to researchers.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Why is there a need for factorial design in experimental research in education?
- ii. What do you understand by ‘main’ and ‘interaction’ effects in factorial design?

7.0 REFERENCES/FURTHER READING

Box, G.E.P, Hunter, W.G. & Hunter, J.S. (1976). *Statistics for Experimenters*. Wiley: New York.

Hassan, T. (1995): *Understanding Research in Education*. Lagos: Merrifield Publishing Co.

MODULE 4 DESIGNS OF RESEARCH II

Unit 1	Repeated Measures Design
Unit 2	Twin Studies
Unit 3	Analysis of Variance and Covariance
Unit 4	Multivariate Studies

UNIT 1 REPEATED MEASURES DESIGN

CONTENTS

1.0	Introduction
2.0	Objectives
3.0	Main Content
3.1	Repeated Measures Design
3.1.1	Description
3.1.2	Types
3.1.3	Advantages and Limitations
4.0	Conclusion
5.0	Summary
6.0	Tutor-Marked Assignment
7.0	References/Further Reading

1.0 INTRODUCTION

You will remember that under longitudinal surveys (see Module 2, Unit 2), we touched on the topic of the time dimension in research, where data is first collected at the outset of the study and may then be gathered repeatedly throughout the length of the study. The repetition period may be long or short, i.e., spanning from a few days to even over a period of decades. The essence of the repeated measures design is to study the effects on subjects of repeating an experiment over time.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define, with an example, repeated measures design
- identify the types of repeated measures designs
- examine the uses of repeated measures designs
- state the advantages and limitations of this research design.

3.0 MAIN CONTENT

3.1 Repeated Measures Design

3.1.1 Description

Repeated measures designs are a type of longitudinal or panel study in which all participants participate in the treatment conditions. They measure the dependent variable on the same subjects under different conditions (e.g. before vs. after). Hypotheses that compare the same subjects under several different treatments, or those that follow performance over time, can be tested using repeated measures design.

Repeated measures designs differentiate among repeated and non-repeated factors. A “between” variable is a non-repeated or grouping factor, such as gender or experimental group, for which subjects will appear in only one level. A “within” variable is a repeated factor for which subjects will participate in each level, e.g. subjects participate in both experimental conditions, albeit at different times (Stevens, 1996).

3.1.2 Types

Repeated measures designs, often referred to as within-subjects design, offer researchers opportunities to study research effects while “controlling” for subjects. They are characterized by having more than one measurement of at least one given variable for each subject. A well-known repeated measures design is the pretest, posttest experimental design, with intervening treatment; this design measures the same subjects twice on an intervally-scaled variable, and then uses the correlated or dependent-sample t-test in the analysis (Stevens, 1996).

A popular repeated measures design is the crossover study. A crossover study is a longitudinal study in which subjects receive a sequence of different treatments (or exposures). Many important crossover studies are controlled experiments in many scientific disciplines, especially medicine. For example, a crossover clinical trial is a repeated measures design in which each patient is randomly assigned to a sequence of treatments, including at least two treatments (of which one “treatment” may be a standard treatment or a placebo). Thus, each patient crosses over from one treatment to another.

SELF-ASSESSMENT EXERCISE

- i. What is a cross-over study?

3.1.3 Advantages and Limitations

The primary benefit of a repeated measures design is statistical power relative to sample size, which is important in many real world research situations. Repeated measures designs use the same subjects throughout different treatments and thus require fewer subjects overall. Because the subjects are constant, the variance due to subjects can be partitioned out of the error variance term, thereby making any statistical tests more powerful (Stevens, 1996).

The advantages of repeated measures designs are: (1) fewer participants or subjects can be used, (2) allows many experiments to be completed more quickly since only a few groups need to be trained to complete an experiment, (3) allow researchers to monitor how the participants change over time, both in the long term and in the short term.

Repeated measures designs are almost always affected by practice effects, which occur when a subject in an experiment is able to perform a task and then perform it again at some later time, with positive or negative feedback effects. Another disadvantage of repeated measures design is that it may not be possible for each participant to be in all conditions of the experiment, e.g. because of time constraints, location of experiment, etc.

Also, carry-over effects, which are effects from one treatment that may extend into and affect the next treatment, may be detrimental to a repeated measures design study. Examples would be tracking memory over time, investigating practice or fatigue on a targeted behaviour, or if a second drug treatment is administered without the previous drug passing out of the subject's system. This internal validity threat can be controlled through counterbalancing. By varying the presentation order of treatments, either randomly or systematically, interaction between treatment order and main effect can be investigated through data analysis. However, even with counterbalancing, carryover effects can raise issues involving external validity.

SELF-ASSESSMENT EXERCISE II

- i. What are “practice effects” in repeated measures design?

4.0 CONCLUSION

Repeated measures designs allow many experiments to be completed more quickly, as only a few groups need to be trained to complete an entire experiment. For example, there are many experiments where each condition takes only a few minutes, whereas the training to complete the

task takes much more time than that. They also allow researchers to monitor how the participants change over the passage of time, both in the case of long-term situations like longitudinal studies and in the much shorter term case of practice effects.

5.0 SUMMARY

Repeated measures designs are very useful in many educational settings where there is a need to measure the dependent variable (e.g. academic achievement) on the same subjects under different treatment conditions (e.g. before and after a specific learning intervention). As a variant of the experimental research design, the primary strength of the repeated measures design is that it makes an experiment more efficient and helps keep the variability low. This helps to keep the validity of the results higher, while still allowing for smaller than usual subject groups.

6.0 TUTOR-MARKED ASSIGNMENT

- i. What are the advantages and disadvantages of a repeated measures design?
- ii. Elaborate on, with examples, the pretest-posttest experimental design, with intervening treatment.

7.0 REFERENCES/FURTHER READING

Gay, L.R. & Airasian, P. (2000). *Educational Research: Competencies for Analysis and Application*. Prentice-Hall, New Jersey.

Myers, J.L. (1979). *Fundamentals of Experimental Design*, 3rd ed. Allyn and Bacon, Boston, Massachusetts.

Shaughnessy, J.J. (2006). *Research Methods in Psychology*. New York: McGraw-Hill.

UNIT 2 TWIN STUDIES

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Twin Studies
 - 3.1.1 Description
 - 3.1.2 Methods
 - 3.1.3 Assumptions
 - 3.1.4 Limitations
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Twin studies is a fascinating research design which has some long albeit controversial history, and touches on many aspects of human behavior – genetic, embryological, biochemical, immunological, behavioral, anthropological, psychological and sociological. In its essence, it is a research method for investigating heredity – environment interaction and its implications for education. Amongst the questions asked in twin studies are: (1) What genetic, biological or environmental factors cause twinning? (2) How do the inherited and acquired traits of twins differ from those of singletons? (3) How do the traits of identical (monozygotic) twins differ from those of non-identical (dizygotic) twins? (4) Why and how does one twin typically dominate the other? (5) What are the effects when monozygotic twins grow up apart? (6) What are the effects of adoption?

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define twin studies as it pertains to educational research
- outline the methods used in twin studies, both traditional and modern
- state and evaluate the major assumptions of this research method
- discuss the problems and limitations of twin studies.

3.0 MAIN CONTENT

3.1 Twin Studies

3.1.1 Description

Twin studies are one of a family of designs in behavior genetics research which aid the study of individual differences by highlighting the role of environmental and genetic causes on behavior. Twin studies ramify into genetic, embryological, biochemical, immunological, behavioral, anthropological, psychological and sociological aspects.

Twin studies can be considered from two standpoints, (1) as a matter of interest in itself, and (2) as a research method for investigating heredity – environment interaction and its implications for education. In the latter, within-pair differences of identical and fraternal pairs are compared for different variables and ages, and any differences within the identical twin pairs is assumed to be due to environmental causes, whereas differences within fraternal pairs are assumed to be to both environmental and genetic factors.

Modern twin studies have shown that almost all traits are in part influenced by genetic differences, with some characteristics showing a strong influence of genetics (e.g. height), other traits like IQ showing an intermediate level of genetic influence, and some other traits (e.g. autism) showing more complex or mixed heritability. Observed similarities in children of a family may reflect shared environmental influences common to members of the family, e.g. social class, parenting styles, education, etc., but they will also reflect shared genes inherited from parents.

3.1.2 Methods

Developmental psychologists who want to rule out the effects of heredity in their investigations often use the co-twin study as a method of research. Co-twin studies typically compare identical twins who have been reared apart, or who have been given different kinds of training. Hilgard's work (1933) in which he trained one twin to remember digits in the first year, then trained the other twin in the second year, and compared their performance on frequent memory tests; is an example of a controlled experiment in twin studies. Time of training was the independent variable manipulated, whilst performance on digit memory tasks was the dependent variable. He found that although both twins benefited from the training, the twin trained later did better than the twin trained in the first year. However, both twins lost their achievement gains after training was ended.

The classical twin study design relies on studying twins raised in the same family environments. Monozygotic (identical) twins share all their genes, while dizygotic (fraternal) twins share only about 50% of them. So, the hunch is that if a researcher compares the similarity between sets of fraternal twins for a particular trait, e.g. intelligence, then any difference between the identical twins should be due to genes rather than the environment. Researchers use this method, and variations on it, to estimate the heritability of traits, i.e. the percentage of variability due to genes. Modern twin studies also try to quantify the effect of a person's shared environment (family) on his or her social environment (the individual events that shape a life) on a trait.

Like all behavior genetic research, the classic twin study begins from assessing the variance of a behavior phenotype in a large group, and attempts to estimate how much of this is due to genetic effects (heritability), how much appears to be due to shared environmental effects (e.g. family) and how much is due to unique environmental effects (e.g. events that occur to one twin but not another, or events that affect each twin in different ways). Typically, those three components are called A (additive genetics), C (common environment), and E (unique environment) – the so called ACE model. Given the ACE model, researchers can determine what proportion of variance of a trait is heritable, versus the proportions which are due to shared or unshared environment.

Emerging genome research methods are increasingly being used to shed light on human behavioral genetics, by modeling genetic – environmental effects, taking into account genetic – environment covariance and interactions, as well as non-additive effects on behavior, using maximum likelihood methods (Martin and Eaves, 1977).

A principal benefit of modeling is the ability to explicitly compare models, including multivariate modeling. This is invaluable in answering questions about the genetic relationship between apparently different variables, e.g. do IQ and long term memory share genes? Do they share environmental causes? Additional benefits of modeling include the ability to deal with interval, threshold, and continuous data, retaining full information from data with missing values, integrating the latent modeling with measured variables, whether measured environment or measured molecular genetic markers. In addition, models avoid the constraint problems in the normal correlation method, so that all parameters will lie between 0 – 1.

Model building in twin studies relies upon an analysis of variance and covariance for twins brought up together and brought up apart. The model should embody a testable null hypothesis. Wilson (1981) has

used a repeated measures analysis of variance model specifically designed for twin data to estimate the developmental consistency of each pair. He found that the intra-pair correlations will tend to increase with age for Dz twins and remain constant for Mz twins, especially for physical traits like height and weight.

SELF-ASSESSMENT EXERCISE I

- i. Describe the classical methodology used in twin studies research.

3.1.3 Assumptions

Assumptions on which twin studies rest include:

- (1) Equal environment: Twin researchers assume that fraternal and identical twins raised in the same home share equally similar environments. But some research suggests that parents, teachers, peers and others may treat identical twins more similarly than fraternal twins.
- (2) Random mating: Twin researchers also assume that people are as likely to choose partners who are different from them as they are to choose partners who are similar for a particular trait. If the latter, i.e. that people tend to choose mates similar to themselves, then fraternal twins could share more than 50% of their genes – and hence more similarities on genetic traits – because they would receive similar genes from their mothers and fathers.

Overall, twin studies assumptions remain controversial.

SELF-ASSESSMENT EXERCISE

- i. Explain the following terms found in twin studies research:
 - ii. “equal environments”
 - iii. “random mating”
 - iv. “ACE model”.

3.1.4 Limitations

Some psychologists have long questioned the assumptions that underlie twin studies, i.e. the assumption that fraternal and identical twins share equal environments, or that people choose mates similar to equal environments assumption. An initial limitation of the twin design is that it does not consider both shared environment and non-additive genetic factors simultaneously. This limitation can be addressed by including

additional siblings to the design. A second limitation is that the gene-environment correlation is not always detectable as a distinct effect. This can be remedied by incorporating adoption models, or children of twins designs, to access family influences uncorrelated with shared genetic effects.

The twin method has been criticized from the standpoints of statistical genetics, statistics, and psychology. The statistical genetics critiques argue that heritability estimates used for most twin studies rest on restrictive and mostly untested assumptions, i.e. that the heritability estimate from a twin study may reflect factors other than shared genes. From the psychological viewpoint, it has been pointed out that the results of twin studies cannot be automatically generalized beyond the population in which they have been derived. It is therefore important to understand the particular sample studied, and the nature of the population since they differ in their developmental environment and in that sense they are not representative. Fraternal twins are a function of many factors, e.g. frequency of producing more than one egg by the mother, genetic factors in the family, age of mother, number of earlier children in the family, in vitro fertilization, etc.

Twin studies are in virtually all cases observational, in contrast to studies in plants or animal breeding where the effects of experimentally randomised genotypes and environment combinations are measured. But it is also argued, on the other hand, that the observational method and its inherent confounding of causes is not limited to twin studies and is in fact common in psychology.

The study of twins in the behavioral sciences has been specifically used as a method to estimate heritability in such different areas as personality, school achievement, and studies of intelligence. The results have mostly shown a higher within-pair similarity for identical twins as compared to fraternal twins, hence the results have been interpreted as due to genetic factors. But there has been criticism and debate about this interpretation because of the difficulty of generalizing results from one twin population to another, the inadequacy of traditional additive model in twin research, and the necessity to take interactional and correlational effects into account. (Fischbein, 1980)

The additive model for interpreting twin data implies that an environmental change will contribute to an equal or comparable change in all genotypes. Interactional effects, on the other hand, will lead to different reactions in genotypes exposed to the same environmental impact. It has therefore been argued that the additive model confuses genetic and interactional effects in interpreting data from different types of twins who differentially react to and construct their own environment.

Genotype and environment may co-vary, so that it is seen that bright children tend to live in a home conducive to their intellectual development, or Mz twins tend to be treated more alike than Dz twins, or people tend to create for themselves different environments related to their genetic potentials (e.g. in the same home, a brighter child may choose to read more books than a less bright child, thereby creating for himself or herself a more stimulating environment).

SELF-ASSESSMENT EXERCISE

- i. Outline some of the limitations of the twin studies design.

4.0 CONCLUSION

Twin studies are one of a family of designs in behavior genetics research which aid the study of individual differences by highlighting the role of environmental and genetic determinants of human behavior. Twins are invaluable for studying these important questions because they help to disentangle the phenomenon of the sharing of genes and environments in human behavior.

5.0 SUMMARY

Systematic twin studies has come a long way from its early 19th century beginnings to its modern use of modeling techniques to fully capture the interaction effects of the many variables, genetic as well as environmental, today considered. In this unit, we have outlined the description, assumptions, methods and limitations of this research design and its current direction in educational research. The future of twin research will likely involve combining traditional twin studies with molecular genetics i.e. DNA technology. It is hoped that you the student have been better informed about this specialized research tool and can see its place alongside other research methods earlier discussed and to be discussed.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Discuss the assumptions which underlie twin studies.
- ii. How does the emerging modeling design assist in strengthening twin studies?

7.0 REFERENCES/FURTHER READING

- Creswell, J.W. (2004). *Educational Research: Quantitative, Qualitative and mixed approaches*. 2nd ed., Boston: Allyn and Bacon.
- Fischbein, S.E.C. (1990). Twin Studies. In J.P. Keeves (ed.) (1990), *Educational Research, Methodology and Measurement: An International Handbook*. Oxford, England: Pergamon Press.
- Husen, T. (1959). *Psychological Twin Research: A Methodological Study*. Stockholin: Almqvist and Wiksell.
- Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1- 123.
- Neale, M.C. and Cardon, L.R. (1992). *Methodology for genetic studies of twins and families*. Dorderecht, The Netherlands: Kluwer Academic Press.

UNIT 3 ANALYSIS OF VARIANCE/COVARIANCE

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Analysis of Variance and Covariance
 - 3.1.1 Description
 - 3.1.2 Uses
 - 3.1.3 Types
 - 3.1.4 Data analysis
 - 3.1.5 Assumptions
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

In its simplest form, analysis of variance is a statistical test of whether the means of two or more groups on a single variable are all equal. Studies with three or more variables classified into two or more ways use multivariate analysis of variance discussed in the next unit.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- explain what analysis of variance is
- state the main assumptions underlying the use of analysis of variance
- list and illustrate three types of analysis of variance
- describe the procedures for analyzing data from analysis of variance
- distinguish between analysis of variance and analysis of covariance.

3.0 MAIN CONTENT

3.1 Analyses of Variance and Covariance

3.1.1 Description

The main function of analysis of variance (ANOVA) is to compare systematically the mean response levels of two or more individual groups of observations or set of observations measured at two or more points in time. ANOVA procedures are based on linear models that depict the partition of scores on a measure into pre-specified components. The analysis of variance model describes a typical score as the sum of three components: (1) a common response level in the total population, (2) deviation from this level according to treatments, and (3) deviation from the average response under a particular treatment. Fitting the analysis of variance model to empirical data consists of performing two general statistical functions, i.e. tests of significance and estimation of effects.

Analysis of variance techniques are validly applied in each of three types of investigations (Bock 1975): (a) experiments in which subjects are assigned at random to treatments determined by the investigator; (b) comparative studies, whose purpose is to describe differences among naturally occurring populations; and (c) surveys, in which the responses of subjects in a single population and, if necessary, differential responses of sub-classes of the population, are to be described.

Analysis of Covariance (ANCOVA), is an extension of ANOVA procedures to incorporate one or more measured scales as additional antecedent variables (covariate). ANCOVA was originally developed for use in experiments where ipso facto the measurement of the covariate must precede the experimental manipulation. In non-experimental studies the covariates may be measured simultaneously with the dependent variables. The same assumptions apply as in ANOVA, i.e. independence of observations, normal distribution of population variance, equal means, and equal population variances.

3.1.2 Uses

ANOVA is typically used in a situation when there are one or more independent variables and two or more dependent variables. For example, to find out the effects of study habits, age, sex, and socio-economic background on college grades, one can do a simple ANOVA which analyses all four sets of data (study habits and grades, age and grades, sex and grades, socio-economic status and grades) at one time — which is an obvious advantage over the t-test statistic. With this

technique, what are called Main effects (e.g. study habits and college grade), as well as Interaction effects (e.g. study habits with age and grades) are clearly partialled out.

SELF-ASSESSMENT EXERCISE

- i. What advantages does ANOVA have over a two-sample t-test?

3.1.3 Types

There are several types of analysis of variance depending on the number of treatments and the way they are applied to the subjects in the experiment.

- One-way ANOVA is typically used to test for differences among at least three groups, since the two-group case can be covered by a t-test. For example, in an experiment in which three Groups A, B & C are given respectively gin, wine and a placebo, and all groups are then tested with a memory test, a one-way ANOVA can be used to assess the effects of the various treatments.
- Two-way ANOVA is used when the subjects are subjected to repeated measures, in which the same subjects are used for each treatment. For example, in an experiment in which Group A is given gin and tested on a memory task, allowed to rest for a period of five days and the experiment is repeated with wine. The procedure is repeated using a placebo. A two-way ANOVA with repeated measures can be used to assess the effects of the gin versus the impact of the placebo. One flaw of this method is that it can be subjected to carryover effects.
- Factorial ANOVA (FANOVA) is used when the experimenter wants to study the effects of two or more treatment variables. The most common of this type is the 2x2 (two by two) design, where there are two independent variables and each variable has two levels or distinct values.
- Multivariate analysis of variance (MANOVA) is used in studies that yield two or more interrelated response measures (i.e. multiple dependent variables).

3.1.4 Data analysis

The fundamental technique in analysis of variance is a partitioning of the total sum of squares (SS) into components related to the effects used in the model. The variance estimate is the SS divided by its degree of

freedom. Effect size measures are used in analysis of variance to describe the degree of relationship between a predictor or set of predictors and the dependent variable. Effect size estimates are reported in ANOVA to allow researchers to compare findings in studies and across disciplines.

When scores vary widely from the mean, the sums of the squares will be large. When they cluster around the mean, the sums of squares are small. The sum of squares is affected by the number of scores in a sample, more scores yield larger sums of squares. This is adjusted for by calculating and applying the degrees of freedom. By calculating the variability in data and producing an F-ratio, this ratio of the variance provides an estimate of the statistical difference between the mean scores. If the ratio of F is equal to or greater than the critical value of the F distribution at the chosen level of significance, then the null hypothesis is rejected and it can be concluded that there are statistically significant differences between at least some of the means. On the other hand, if the calculated F ratio is smaller than the table value of F the null hypothesis is retained and no significant difference in the means is affirmed.

The measure of differences among means is the “mean square between groups” (MS_B). Mean square between groups is in fact a summary of these differences, and is generally larger when group means are far from one another and smaller if the subgroup means are close together. Mean square within groups (MS_W) on the other hand, is the measure of differences among individual subjects within the groups. The greater the variation among individuals within the groups, the greater the mean square within groups will be.

The test statistic used here is the F-test or F-ratio (MS_B/MS_W) which can be found in tables of the F distribution. Where F exceeds the tabled critical value, it means that MS_B is sufficiently larger than MS_W to conclude that group differences predominate, and H_0 is rejected. If on the other hand the critical F value is not exceeded, the null hypothesis of number difference is maintained.

SELF-ASSESSMENT EXERCISE

- i. Elaborate on the meaning of F-ratio and how it is applied in ANOVA.

3.1.5 Assumptions

The major assumption underlying analysis of variance is that the observations must be sampled and respond independently of one

another. If dependencies arise because the same subjects or groups are measured repeatedly, then multivariate analysis (which is primarily concerned with the study of relationships between and within one, two or more sets of variables that are sequentially ordered with respect to time) should be employed. Other assumptions underlying analysis of variance are: (1) normal distribution of the population variance, (2) equal population variances (homoscedacity), and (3) equal means.

A typical presentation format for analysis of variance is as under:

Sample ANOVA Summary Table

Source of Variance	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F
Between-groups	114.96	2	57.48	7.82*
Within-groups	176.45	24	7.35	
Total	291.41	26		

* $P < .05$

(Source: Hassan, p.229)

4.0 CONCLUSION

At the heart of the variance calculation is the sum of squares (SS):, which measures the variability of the scores from the mean of the sample. Analysis of variance calculates an F-statistic (F-ratio) for group mean differences, obtained by the formula MS_B/MS_W . The calculated F is then compared to the table F at the desired level of significance with the appropriate number of degrees of freedom to determine whether the null hypothesis of no significant difference in the means is to be retained or the alternative hypothesis of a significant difference in the means of the various samples is to be accepted.

5.0 SUMMARY

Analysis of variance is a collection of statistical models, and their associated procedures, in which the observed variance in the scores is partitioned into components due to the different explanatory variables. It is superior to the t-test and similar parametric measures for complex analyses of data, for two reasons: (1) its ability to combine complex data into one statistical procedure, and (2) its ability to determine what are called interaction effects.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Explain what you understand by the 'logic of ANOVA'.

- ii. Differentiate between ANOVA and ANCOVA.

7.0 REFERENCES/FURTHER READING

Bock, R.D. (1975). *Multivariate Statistical Methods in Behavioral Research*. New York: McGraw-Hill.

Finn, J.D. (1990). *Analysis of Variance and Covariance*. In J.P. Keeves, (ed.) (1990): *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford: Pergamon Press.

Glass, G.V. & Stanley, J.C. (1970). *Statistical Methods in Education and Psychology*. Englewood Cliffs, New Jersey: Prentice-Hall.

Hassan, T. (1995). *Understanding Research in Education*. Lagos: Merrifield Publishing Co.

[http://en.wikipedia.org/wiki/Analysis of variance](http://en.wikipedia.org/wiki/Analysis_of_variance)

UNIT 4 MULTIVARIATE STUDIES

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Multivariate Studies
 - 3.1.1 Description
 - 3.1.2 Methods
 - 3.1.3 Advantages
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Because of the complex nature of the human organism, single variables rarely can validly account for or explain human behaviour. There is the need therefore, in the social sciences, including education, for taking into account several other intervening and interactive factors that impinge on human actions. Multivariate analysis attempts to do just this by studying the relationships between one, two or more sets of variables that are sequentially ordered with respect to time. The first measured set of variables forms the predictor set, whilst the second measured set forms the criterion set.

2.0 OBJECTIVES

At the end of his unit, you should enable you to:

- define what is multivariate analysis
- describe the nature of the variables examined in multivariate analysis
- explain some of the techniques for significance testing in multivariate analysis
- express some of the advantages and disadvantages of multivariate analysis.

3.0 MAIN CONTENT

3.1 Multivariate Studies

3.1.1 Description

In most studies, there are one or more outcome (or response) variables and several explanatory variables, together with a variety of variables which add a characteristic particularity to the study. For example, in a study of infant feeding and growth in children, growth is the outcome variable and the method of feeding the explanatory variable. All additional information like mother's age, her social class, her education, income, number of siblings, etc., provide the particularity of each mother-infant pair.

Method	Type of Dependent Variable	Type of Independent Variable	Purpose
Multiple Regression Analysis	Continuous (numerical)	Mostly continuous but in practice categorical can be used	To describe the extent, direction, and strength of the relationship between several independent variables and a continuous dependent variable
Logistic Regression Analysis	Categorical dichotomous	Continuous and categorical	To describe how many times more likely is the event in one group compared to the other
Analysis of variance	Continuous	All nominal	To describe the relationship between a continuous dependent variable and one or more nominal independent variable
Discriminant Analysis	Nominal (polychotomous)	Commonly all continuous, but in practice a mixture of various types can be used as long as some are continuous	To determine how one or more independent variables can be used to discriminate among different categories of a nominal dependent variable
Factor analysis	Commonly continuous, but in practice may be of any type. The variables are not initially identified as dependent or independent, but the resulting factors may be used as dependent or independent variables in later analysis.	Commonly continuous, but in practice may be of any type. The variables are not initially identified as dependent or independent, but the resulting factors may be used as dependent or independent variables in a later analysis.	To define one or more new composite variables called factors.
Analysis of covariance	Continuous	Mixture of nominal and continuous variables. The continuous variables are used as control variables	To describe the relationship between a continuous dependent and one or more nominal independent variables, controlling for the effect of one or more variables

Multivariate analysis is the research method which takes into account several variables acting simultaneously instead of singly. In most such studies, the researcher is attempting to describe the relationship between an outcome variable and its several determinants. For such purpose, simple univariate methods would not do, since they would provide a one-sided or biased view.

SELF-ASSESSMENT EXERCISE

- i. Define, in your own words, multivariate analysis.

3.1.2 Methods

The table below sets out the main methods of multivariate analysis.

The variables examined in multivariate analyses may involve nominal data with two categories (dichotomous data), or with more than two categories (polychotomous data), or the variables may involve ordinal, interval, or ratio-scaled data.

It is also commonly necessary to transform variables that are members of the criterion set to ensure that they are normally distributed and that the use of the multivariate normal distribution model is appropriate. Although several different approaches have been advanced with regard to testing for significance in multivariate analysis, the general principles of testing employed in multivariate analysis are parallel to those used in univariate analysis.

The internal analysis techniques (which seek interrelations between the variables within a single set) used in multivariate analysis include contingency table analysis, cluster analysis, factor analysis, principal components analysis, etc; whilst the external analysis techniques (which seek not only interrelations between the variables but also the interrelations between the two sets of variables) include multiple regression analysis, canonical correlation analysis, multiple analysis of variance and covariance, discriminant analysis, multiple classification analysis, etc.

SELF-ASSESSMENT EXERCISE

- i. Describe briefly the nature and treatment of the variables examined in multivariate analysis.

3.1.3 Advantages

In general, multivariate methods have the advantage of bringing in more information to bear on a specific outcome, and allow one to take into account the continuing relationship among several variables, especially in observational studies where total control is never possible.

The specific advantages of multivariate studies are as follows:

- (1) They resemble closely how the researcher thinks about the data.
- (2) They allow easier visualization and interpretation of data.
- (3) More data can be analysed simultaneously, thereby providing greater statistical power.
- (4) Regression models can give more insight into relationships between variables.
- (5) The focus is on relationships among variables rather than on isolated individual factors.

SELF-ASSESSMENT EXERCISE

- i. What are the advantages of multivariate analysis?

4.0 CONCLUSION

The main theory of multivariate analysis is based upon the multivariate normal distribution which enables the testing of the significance of relationships between variables obtained from multivariate analyses. A major assumption associated with the use of this model is that observations on a predictor set of variables must be independent of one another, with respondents being both independently sampled and providing information without reference to one another.

5.0 SUMMARY

The many-faceted nature of educational processes demands that measurements should be made on many variables, and that the procedures of analysis employed should be capable of the simultaneous examination and analysis of the many variables on which data have been collected. With the availability of the high speed computer, it is now possible to undertake the tedious analysis of data involving a large number of cases and many variables, and thus achieve greater exactness in the making of observations and therefore increasing their potential usefulness.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Comment briefly on any two of the analytical techniques used in multivariate analysis.
- ii. What are the benefits and drawbacks of multivariate analysis?

7.0 REFERENCES/FURTHER READING

Bock, R.D. & Haggard, E.A. (1968). *The use of multivariate analysis in behavioral research*. In D.K. White (ed.), 1968, *Handbook of Measurement and Assessment in Behavioral Sciences*. Reading, Mass: Addison-Wesley.

Finn, J.D. (1974). *A General Model for Multivariate Analysis*. New York: Holt, Rinehart and Winston.

Keeves, J.P (1986). *Aspiration, motivation and achievement: Different methods of analysis and different results*. *Int. J. Educ. Res.* 10(2): 117-243.

Keeves, J.P. (1990). *Multivariate Analysis*. In J.P. Keeves, (ed.) (1990): *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford: Pergamon Press.

MODULE 5 TIME SERIES, TREND, AND META ANALYSIS STUDIES

Unit 1	Time Series Design
Unit 2	Trend Studies
Unit 3	Meta Analysis Studies

UNIT 1 TIME SERIES DESIGN

CONTENTS

1.0	Introduction
2.0	Objectives
3.0	Main Content
3.1	Time Series Design
3.1.1	Definition
3.1.2	Methodology
3.1.3	Strengths and weaknesses
4.0	Conclusion
5.0	Summary
6.0	Tutor-Marked Assignment
7.0	References/Further Reading

1.0 INTRODUCTION

Time-series design is a type of longitudinal design where data are collected over long periods of time, and measurements are taken on each variable over two or more distinct time periods. This allows the researcher to measure change in variables over time.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define time series design
- explain the uses of time series studies
- state how time-series experiments are carried out
- explain why time-series studies cannot establish cause-effect relationship.

3.0 MAIN CONTENT

3.1 Time Series Design

3.1.1 Definition

Time-series design is one of the types of longitudinal studies which generally assume that human development is a continuous process which can be meaningfully examined by a series of “snapshots” recorded at appropriate points in time.

True experiments satisfy three conditions: (i) the experimenter sets up two or more conditions whose effects are to be evaluated subsequently; (ii) persons or groups of persons are then assigned strictly at random, i.e. by chance (e.g. by flip of a coin or pick from a hat), to the conditions; (iii) the eventual differences between the conditions on the measure of effect (e.g. the pupils’ achievement) are compared with the differences of chance or random magnitude. The time-series design is one of the various quasi-experimental designs that is nearest to the true experiment whose hallmark is the causal proof.

Two factors can adversely affect a causal relationship: (1) an ambiguous direction of influence. For example, does enhanced motivation cause pupils to learn successfully in school, or is it the other way round, i.e. success in learning causes an increase in motivation to learn? The truth is probably somewhere in between, (2) the confounding effect of third variables, when two things are related because each is causally related to a third variable, not because of any causal link between each other. An example would be the relationship between teaching method and pupil achievement, where the third factor may well be pupil motivation. In fact, “spurious” factors like age, height, weight, intelligence, experience, motivation, socio-economical status, nationality, etc. always have to be considered in any causal statement or assumption.

SELF-ASSESSMENT EXERCISE

- i. Define time-series design in educational research.

3.1.2 Methodology

A time-series design collects data on the same variable at regular intervals (weeks, months, years, etc.) in the form of aggregate measures of a population. Time series designs are useful for:

- establishing a baseline measure;
- describing changes over time;

- keeping track of trends;
- forecasting future (short-term) trends.

In time-series experiments, there are treatment interventions at various points. In fact, its foremost requirement is that data must be recorded for many consecutive points in time before and after a treatment is introduced. Since it incorporates many pre-intervention and post-intervention observations, it is argued that this permits separating real intervention effects from other long-term trends in a time-series.

Time series data are nearly always presented in the form of a chart or a graph. The horizontal (or x) axis is divided into time intervals, and the vertical (y) axis shows the values of the dependent variable as they fluctuate over time. Researchers inspect a time series graph to look for four types of patterns:

- (i) long term trends (increases or decreases over the whole time span);
- (ii) cyclical variations (short-term, valley to valley, or peak to peak cycles);
- (iii) seasonal variations (due holidays or weather);
- (iv) irregular fluctuations (none of the above).

SELF-ASSESSMENT EXERCISE II

- i. Comment critically on the methodology used in time series design.

3.1.3 Strengths and Weakness

The simple logic espoused of the time series experiment is that if the graph of the dependent variable shows an abrupt shift in level or direction precisely at the point of intervention, then the intervention is a cause of the effect on the dependent variable.

Unfortunately, time series experimenters have remained resistant to most considerations of statistical analysis, preferring to rely on visual inspection of the graph to judge the presence of an effect despite its heavy taint of unreliability. Some of the researchers have argued that impressionistic visual examination of time experiments is superior over statistical analysis, but Glass (1996) contends that the statistical problems of interpreting time series experiments are sufficiently complex that merely eyeballing graphs and drawing conclusions is not enough. Proper distinction should be made between exploratory (hypothesis generating) research investigations typical of the time series research design, on the one hand, and confirmatory (hypothesis testing)

studies, typical of experimental studies, on the other. Besides, the “eyeball” test is differently applied by different persons, some seeing effects, some seeing none at all.

He has also drawn a distinction between stationary and non-stationary time series, and has argued that it is far easier to detect an intervention effect in a stationary process than in a non-stationary one. He further claims that time-series of observations of a single individual are often stationary, whereas time series based on large groups (e.g. classrooms, cities, states or national populations) are often non-stationary. Time-series experiments however, require long baseline periods for establishing whether the process is stationary or non-stationary. (Glass, 1996).

Also, the conduct of time-series studies is expensive since it is usually very costly to maintain contact with a significant number of sample members over an extended period of time, and sample losses can give rise to substantial distortions of observed relationships. This problem has led to suggestions of retrospective time series design in which a sample is selected and the members of the sample are invited to recall events in their lives at particular times or at specific ages. But this is fraught with two problems: (a) sample bias, and (b) distorted recall of past events either deliberately or unintentionally.

SELF-ASSESSMENT EXERCISE

- i. How is time series study different from true experimental study?

4.0 CONCLUSION

In conclusion, Glass (1996) states that “at its easiest, time series experiments require little more than good graphing skills, a skeptical attitude towards one’s pet hypotheses and the capacity to subdivide the data to locate hidden effects. At its most complex, it involves complicated statistical analyses to separate the unaccountable variation of indices across time from the determinant effects of planned interventions”.

5.0 SUMMARY

Separating potential reasons for effects into those essentially related to the intervention and those only accidentally related remains the principal task in analyzing time-series experiments. In this sense, it is analogous to the pretest-posttest research design where many other influences other than the treatment can account for a change in scores from pre to post. For example, the person or persons in the experiment could grow tired,

smarter, or less cooperative from the single pre-measure to the post-measure.

6.0 TUTOR-MARKED ASSIGNMENT

- i. What are the uses of the time-series design?
- ii. Why can't time-series studies establish cause-effect relationship?

7.0 REFERENCES/FURTHER READING

Campbell, D.T. & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research on teaching*. In N.L. Gage (ed.), *Handbook of research on teaching*. Chicago: Rand-McNally.

Glass, G.V., Wilson, V.L & Gottman, I.M.(1975). *Design and analysis of time-series experiments*. Boulder, co: Colorado Associated University Press.

Glass, G.V. (1996). *Interrupted Time-Series Quasi-Experiments*. Arizona State University.

UNIT 2 TREND STUDIES

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Trend Studies
 - 3.1.1 Description
 - 3.1.2 Methodology
 - 3.1.3 Advantages and Drawbacks
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Certain observable trends are present both in nature and in human activities, including education (e.g. the “project method” in teaching). In this unit, you will learn how and why these trends occur, how they can be measured, and what net benefits accrue from them (if any.)

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define what trend studies is
- give examples of possible trend studies in Nigeria
- explain the methodology used in trend studies
- enumerate the benefits and drawbacks of trend studies.

3.0 MAIN CONTENT

3.1 Trend Studies

3.1.1 Description

Trend study is one of the most common types of longitudinal study. A trend study samples different groups of people at different points in time from the same population, and provides information about net changes at an aggregate level. A public opinion poll about whether or not the downstream oil sector should be deregulated in Nigeria, taken from the same population of adults at two time points, say in September 2009 and a year later in October 2010, would be an example of a trend study. In this example, public opinion may for instance have shifted positively in the one year interval, from say 35% to 68% approval, but it is not

possible to know how many people actually changed their position or how many stayed with their original opinion. To determine both the gross change and the net change, a panel study would be necessary.

Another example of trend studies would be the almost unending annual debate, after the release of results of say WASC, NECO, NABTEB, etc. examinations, about the rates and trends of passes and failures in single subjects as well as in the entire examination – leading to a flurry of newspaper and magazine articles, academic papers, radio/television discussions, etc. on the familiar refrain of falling/rising? standards of education in Nigeria!

3.1.2 Methodology

In this kind of study, although data are collected from the population at more than one point in time this does not always mean that the same subjects are used to collect data at more than one point in time, but that the subjects are selected from the population for data at more than one point in time. There is no experimental manipulation of variables, or more specifically, the investigator has no control over the independent variable. Also, no intervention is made by the investigator other than his or her method or tool used in collecting data.

In analyzing the data, the investigator draws conclusions and may attempt to find correlations between variables. Therefore, trend studies are uniquely appropriate and valuable for describing long-term changes in a population over time, and for prediction questions because variables are measured at more than one time. However, the method is not appropriate for drawing causal conclusions since there is no manipulation of the independent variable.

SELF-ASSESSMENT EXERCISE I

- i. How are trend studies carried out?

3.1.3 Advantages and Drawbacks

Two important advantages of trend studies are flexibility and cost-effectiveness. A trend study can be based on a comparison of survey data originally constructed for other purposes, subject of course to necessary adjustments such as question wording, contexts, sampling, or analysis techniques. Also for this reason, i.e. use of secondary data, it saves time, money and personnel.

At the same time, there is less concern for internal validity in trend studies since they do not aim to produce causal inferences as in experimental studies or in some panel studies. Also, if study data are

unreliable or inconsistent (e.g. changes in the way indexes are constructed or the way questions are asked), false trends may show up in the results or the results will be biased, just like instrumentation threat can bias experimental studies. In the worst case scenario, the changes in measures alone can produce a pseudo trend which might becloud the results for both the researchers and the readers.

SELF-ASSESSMENT EXERCISE II

- i. Why cannot a trend study draw causal conclusions?

4.0 CONCLUSION

Trend analysis provides descriptive trends of some topic in a certain period of time. Being a longitudinal study, it does not aim to provide causal inferences, although the derivable information can be enlightening and trigger off some experimental investigations.

5.0 SUMMARY

Trend studies are valuable in describing long-term changes in a population. They can establish a pattern over time to detect shifts and changes in some events, e.g. changes in the number of people using commercial motorcycles (okada) in Lagos metropolis over a time period of say two years, or changes in the pattern of consumption of packaged dietary supplements in Abuja capital city between January 2010 to December 2010, or the acquisition rate of digital satellite television (dstv) in Ikot Ekpene for a specified time period. Such patterns are obviously transitory, but they do serve a purpose of providing exploratory (and perhaps predictive) information for further use in several human situations.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Give and elaborate on an example of a possible trend study in Nigeria.
- ii. Outline the advantages and disadvantages of trend studies.

7.0 REFERENCES/FURTHER READING

Hassan, T. (1995). *Understanding Research in Education*. Lagos: Merrifield Publishing Co.

Keeves, J.P. (1990). *Longitudinal Research Methods*. In J.P. Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford, England, Pergamon Press.

UNIT 3 META ANALYSIS STUDIES

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Meta Analysis Studies
 - 3.1.1 Description
 - 3.1.2 Uses
 - 3.1.3 Procedures
 - 3.1.4 Problems
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Meta-analysis is essentially a research integration strategy. It is a quantitative analysis of a sample of existing research studies on a particular topic. It is used to draw conclusions about the topic from a range of existing studies, e.g. identify aspects of a program associated with program success. It may also generate new hypotheses for future research.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- meaningfully define meta-analysis in educational research
- enumerate some of its uses
- evaluate its common procedures
- discuss some of its problems
- conduct an actual meta-analytic study of your choice.

3.0 MAIN CONTENT

3.1 Meta Analysis Studies

3.1.1 Description

Educational research, by its very nature, often produces inconsistent or contradictory results. Differences among studies in treatment, settings, measurement instruments, and research methods make research findings

difficult to compare. Even frequent replications can prove inconclusive. Literature on a topic may be so extensive as to obscure trends with an overwhelming amount of information. Therefore, confronted with an overwhelming number of studies on a given topic, with no two studies exactly alike, it is difficult to determine if the differences between the study outcomes are due to chance, to inadequate study methods, or to systematic differences in the characteristics of the studies.

Meta-analysis was created out of the need to extract useful information from the cryptic records of inferential data analyses in the abbreviated reports of research in journals and other printed sources. Meta-analysis is a collection of systematic techniques for resolving apparent contradictions in research findings, by amalgamating, summarizing and reviewing previous quantitative research on a topic. Meta-analysts translate results from different studies to a common metric, and statistically explore relationships between characteristics and findings.

SELF-ASSESSMENT EXERCISE I

- i. What are the rationales for meta-analysis?

3.1.2 Uses

Meta-analysis responds to several problems in educational research. First, important issues are or have been studied by numerous investigators. The amount of information on a given topic therefore is often overwhelming and not amenable to normal summary. Even when there are relatively few studies on a given topic, it is difficult to determine if outcome differences are attributable to chance, to methodological inadequacies, or to systematic differences in study characteristics.

The appeal of meta-analysis is that it in effect combines all the previous research on one topic into one large study with many participants. Its objectivity is another appeal. Meta-analysis has been used to gain helpful insight into

- (a) the overall effectiveness of interventions (e.g. in psychotherapy, outdoor education, etc).
- (b) the relative impact of independent variables (e.g. the effect of different types of therapy).
- (c) The strength of relationships between variables.

3.1.3 Procedures

Meta-analysis typically follows the same steps as in primary research. (1) The meta-analyst first defines the review's purpose. (2) Sample selection consists of applying specified procedures for locating studies that meet specified criteria for inclusion. (3) Data are collected from previous studies in two ways: (a) study features are coded according to the objectives of the review, and as checks on threats to validity, (b) study outcomes are transformed to a common metric so that they can be compared; (4) Statistical procedures are then used to investigate relationships among study characteristics and findings.

In the classic or Glassian meta-analysis (Glass 1976), the approach is to (i) define questions to be examined, (ii) collect studies, (iii) code study features and outcomes, and (iv) analyze relationships between study features and outcomes. In this approach, (1) liberal inclusion criteria are used; (2) the unit of analysis is the study finding; (3) the effects from different dependent variables are averaged, even when these measure different constructs which later can confuse the reliability of findings.

Meta-analysis report findings in terms of "effect sizes". Glass (1976) defined effect size as a standardized mean difference between the treatment and control groups. The effect size provides information about how much change is evident across all studies and for subsets of studies. There are two main types of effect size: (a) standardized mean difference, and (b) correlation (e.g. Pearson's r). The standardized mean effect size is basically computed as the difference score divided by the SD of the scores. It is possible to convert one effect-size into another, so each really just offers a differently scaled measure of the strength of an effect or a relationship.

In meta-analysis, the effect sizes are usually reported with (i) the number of studies and the number of effects used to create the estimate, and (ii) confidence intervals, to help readers determine the consistency and reliability of the mean estimated effect size. Tests of statistical significance can also be conducted and on the effect sizes, and different effect sizes are calculated for different constructs of interest. Rules of thumb and field-specific benchmarks can be used to interpret effect sizes. For example, according to Cohen (1988) a standardized mean effect size of zero means no change, negative effect sizes mean a negative change, 0.2 a small change, 0.5 a moderate change, and 0.8 a large change. Wilson (1986), on the other hand, suggests that 0.25 is educationally significant, and 0.50 is clinically significant.

There is often overestimate of effect sizes because of underestimates of the within-groups standard deviation. The ultimate purpose of meta-analysis, in seeking the integration of a body of literature, is both to

determine what average effect size is reported in the studies and to establish whether variability in the effect sizes across studies can be accounted for in terms of covariability with characteristics of the studies.

In terms of estimating variation in effect sizes, the more typical procedure in meta-analysis has been to use parametric procedures to calculate the variance of the effect size estimates and then, in many cases, to establish a confidence interval around their mean. But classical statistics seem not to be able presently to fully reproduce the complex cognitive processes that are commonly applied with success by data analysts.

SELF-ASSESSMENT EXERCISE

- i. What do you understand by effect-sizes in meta-analysis studies?

3.1.4 Problems

Problems with doing meta-analysis are:

- (1) locating suitable studies;
- (2) studies may not all have the same dependent variable;
- (3) only findings of statistical significance are usually published;
- (4) different studies may have many dissimilar aspects, making them difficult to compare.

Criticisms of meta-analysis tend to fall into two categories: (1) meta-analysis obscures important qualitative information by “averaging” simple numerical representations across studies; (2) research is best reviewed by a reflexive expert who can sift kernels of insight from the confusing argumentation of a field. Also to be considered is whether the accessible literature is itself biased, aside from any bias by the reviewer himself/herself.

Another criticism of meta-analysis is that the capacity to cope with large bodies of literature causes the net to be cast too widely and encourages the attempted integration of studies which are insufficiently similar. But this does not justify arbitrary exclusion of studies to make the task manageable, since the use of the modern computer can help the reviewer surmount the complexity problem. The reviewer can code hundreds of studies into a data set. The data set can then be manipulated, measured and displayed by the computer in a variety of ways.

SELF-ASSESSMENT EXERCISE III

- i. Discuss one of the criticisms of meta-analysis.

4.0 CONCLUSION

In the words of Glass (1976), meta-analysis is a “rigorous alternative to the causal, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature”. In meta-analysis, research studies are collected, coded and interpreted using statistical methods similar to those used in primary data analysis. The result is an integrated review of findings that is more objective and exact than a narrative review.

5.0 SUMMARY

The human mind is not particularly equipped to consider simultaneously a large number of alternatives, which is exactly the scope of the problem faced by a researcher attempting to integrate the results from a large number of studies. As research results accumulate on any given topic, it becomes increasingly difficult to understand what they tell us, and to find the knowledge in this flood of information. Meta-analysis is the statistical analysis of a collection of individual studies to try to make sense of the multifarious results and conclusions from these studies.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Critically examine the procedures typically followed in meta-analysis studies.
- ii. Discuss critically the problems posed in doing meta-analysis.

7.0 REFERENCES/FURTHER READING

- Glass, G.V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5*, 351-379.
- Glass, G.V, McGraw, B., & M.L. Smith (1981). *Meta-analysis in Social Research*. Beverly Hills, CA: SAGE.
- Glass, G.V. (2000). *Meta-analysis at 25*. College of Education, Arizona State University.
- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York: Russell SAGE Foundation.
- Kulik, J.A., & C.-L.C. Kulik (1989). Meta-analysis in education. *International Journal of Educational Research, 13*, 221-340.

McGraw, B. (1990). Meta-analysis. In J.P. Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford, England, Pergamon Press.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Revised edition. Newbury Park, CA: SAGE Publications.

Rudner, L., Glass, G., Evaritt, D. & Emery, P. (2002). *A user's guide to the meta-analysis of research studies*. University of Maryland.

Wolf, F.M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: SAGE.

MODULE 6 INSTRUMENTATION AND GENERAL DATA PROCESSING TECHNIQUES

Unit 1	Meaning, Types and Classification of Measurement
Unit 2	Construction and Validation of Measurement Instruments
Unit 3	Descriptive Data Analysis
Unit 4	Inferential Data Analysis

UNIT 1 MEANING, TYPES AND CLASSIFICATION OF MEASUREMENT

CONTENTS

1.0	Introduction
2.0	Objectives
3.0	Main Content
3.1	Definition of Measurement
3.2	Attributes of Measurement
3.3	Types of Measurement
3.4	Scales of Measurement
4.0	Conclusion
5.0	Summary
6.0	Tutor-Marked Assignment
7.0	References/Further Reading

1.0 INTRODUCTION

Measurement in fact can have two meanings: there is a broad sense in which it can mean mere identification or typification of phenomena without quantification, and there is the other more usual sense in which it can mean the ascertainment of dimensions or quantity of phenomena. In both senses, especially of the latter, it is the key to psychology as a scientific study of human behavior, and is what significantly differentiates modern psychology from earlier 'philosophical' or 'armchair' psychology. Its purpose is to elicit reliable information about the behavior of persons, both as individuals and as groups, and to detect differences between them. It enables the psychologist to draw precise and valid conclusions from controlled and not so controlled experiments, and to correctly evaluate results.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define measurement as applied in educational research
- identify the main attributes of measurement
- explain the major types of measurement
- discuss the various scales of measurement.

3.0 MAIN CONTENT

3.1 Definition of Measurement

E.L. Thorndike (1927) once pontificated that “if a thing exists, it exists in some amount; if it exists in some amount, it can be measured”. This is true enough in a theoretical, abstract way, but it is obviously not always true in an absolute sense. Umbrellas exist, and no doubt it is quite possible to measure their number (given boundless effort, resources, patience, and luck!). At the same time, the universe does exist, but our conception of its size or quantity is not and cannot be absolute, and will always be limited by our contemporary capacity to comprehend its extent.

A well-known and popular definition of measurement is “the assignment of numerals to objects or events according to rules” (Kerlinger, 1963), (Stevens, 1946, 1951), (Campbell, 1940). Although this definition has attracted some degree of controversy lately (e.g. Luce 1986, Michell, 1999), it is broad and cogent enough to be acceptable to most researchers in the social sciences.

As applied in education, measurement may be defined as the process of assigning numbers or other symbols to persons, objects or events in such a way that relationships of the numbers or symbols reflect relationships of the attributes being measured. Put another way, it is the use of a number or other symbol to express in quantitative terms the degree to which an individual possesses a given attribute or characteristic.

Although measurement is essentially a descriptive process, the attribute of quantification tends to increase the precision and objectivity of the description, so that it will have the same meaning from time to time, place to place, person to person. For example, the statement that Omonye’s score on a standard achievement test in mathematics falls at the 80th percentile, based on the norms for that test, is much less ambiguous than the statement that Omonye has a high score in mathematics. For describing the behavior and performance of pupils or

students, it is best, whenever possible, to avoid the use of descriptive adjectives which tend to have different meanings for different people.

SELF-ASSESSMENT EXERCISE I

- i. How would you define measurement?

3.2 Attributes of Measurement

The process of measurement involves three steps: (i) identifying and defining the quality or attribute to be measured; (ii) determining a set of operations by which the attribute may be made manifest and perceivable and (iii) establishing a set of procedures or definitions for translating observations into quantitative statements of degree or amount.

Three main attributes of any good measurement are validity, reliability, and usability. Validity of a test refers to its ability to actually measure what it claims to measure, or what it was designed, or being used, to measure. Reliability means that a test gives dependable or consistent scores. Usability includes all such practical factors as cost, ease of scoring, time required, etc. The first two elements, validity and reliability, are in fact, the sine qua non of any measurement.

Validity: We may distinguish three types of “non-technical validity”, i.e. face validity, content analytic, and faith validity; and three types of “technical validity”, i.e. content validity, empirical validity, construct validity.

Non-technical Validity

- (1) *Face Validity:* is concerned with whether a measurement instrument appears to measure what it was designed to measure, i.e. the test looks as if it would be valid or reasonable. Whilst face validity has no technical or statistical basis, it must not be overlooked if a test is to be accepted by its user clientele. Good face validity also helps to keep motivation high since people are likely to try harder when the test seems reasonable.
- (2) *Content-Analytic validity:* sometimes also referred to as course or curricular validity. Here, the content of the test is examined in detail to ensure that the test items are representative of the domain to be measured. For example, a spelling test containing only the names of professional footballers would be of poor content validity as a general purpose measure of spelling; as would be an achievement test whose items do not cover important bits of knowledge or skills related to a particular training programme.

- (3) *Faith Validity*: somewhat similar to face validity, is simply a conviction, a belief or blind faith that says a selection test is valid. There is no empirical evidence nor is any such evidence wanted.

Technical Validity

- (1) *Content validity*: the concept of content validity (also known as logical validity) is more commonly met in connection with attainment tests, although an appraisal of test content must inevitably enter into the construction of ability and interest measures. This is the extent to which a test covers the entire range of relevant behaviours, thoughts and feelings about a given social construct. For example, a course examination has good content validity if it covers all the relevant material that students are supposed to learn, and poor content validity if it does not.
- (2) *Empirical validity*: (sometimes also called “criterion-related validity”). This connotes how well the test measures what we want it to in practical situations, i.e. how closely the test relates to some criterion or standard of performance. When empirical validity is high, we can use the test for predicting performance on the criterion variable. Empirical validity may be either (a) concurrent or (b) predictive.
- (a) *Concurrent Validity*: is the relationship between test scores and some criterion of performance obtained at the same time, e.g. relating scores in computer programming to supervisors’ ratings of work performance in a computer programming office. However, although a test may be of high concurrent validity does not necessarily mean that it will be useful in predicting later performance. For example, a high score in a multiple choice test in French vocabulary may not successfully predict those pupils best able to learn French before formal instruction had begun.
- (b) *Predictive validity*: i.e the extent to which a test predicts some future outcome or criterion, e.g. WASC with JME, or JME with degree class. This form of validity is very useful in placement and personnel selection, but its major problem, as with concurrent validation, is that of deciding the job or placement criteria.
- (3) *Construct Validity*: is more abstract than the other forms of validity, and is the extent to which a test measures a theoretical construct or trait. A construct is any hypothetical attribute such as

love, duty, happiness, verbal/spatial mechanical ability, emotional stability, intelligence, imagination. One cannot see a construct (e.g. intelligence), but one can see the results emanating from that construct (e.g. intelligent action).

In general, construct validity is concerned with the psychological meaningfulness of a test. Building up a picture of the construct validity of a test can be a long process, and involves any information which throws some light on the nature of the construct under investigation. For example, correlations with other tests may provide useful information on test construct validation (e.g. we would expect a mechanical ability test to be more highly related to other measures of mechanical ability than to tests of clerical speed and accuracy).

The statistical technique for identifying the basic dimensions causing the inter-correlations between different tests, which is often met with in construct validation, is known as factor analysis. Hence, factorial validity is sometimes used synonymously with construct validity. A test is said to have high factorial validity if it seems to be a good measure of some dimension which has been isolated or identified through factor analysis.

SELF-ASSESSMENT EXERCISE II

- i. What do you understand by construct validity of a test?

Reliability

Reliability refers to the precision and consistency of measurement, and is important because of its relationship to validity. A test cannot measure anything well unless it measures something consistently, but the converse is not true, i.e. a test may measure consistently without measuring well the characteristic we are interested in. For example, a broken ruler may give consistent repeated readings of length which totaled together may not give us a valid measurement of the length of the object we are interested in.

Reliability can be assessed in two forms:

- (1) Absolute consistency, refers to the variability in score if a person were tested repeatedly with the same test (or parallel forms of the test). This way of viewing reliability in test score units is expressed through the Standard Error of Measurement. (SEM), whose formula is given as

$$SEM = SD \sqrt{1 - r}$$

Where SD = standard deviation of scores in the group
 r = reliability coefficient of test

- (2) Relative consistency refers to the ability of the test to yield scores which place examinees in the same position relative to each other. In other words, relative consistency provides us with an index of the overall dependability of scores in the form of a correlation coefficient known as the coefficient of reliability or internal consistency. It is important to remember that a reasonably high reliability e.g. ($r \geq 0.75$) is necessary for good validity of a test, but that reliability does not ensure validity; reliability is necessary for, but is not a guarantee of, validity.

Three types of reliability have been distinguished:

- (1) Scorer Reliability: Unlike in objective tests, there is considerable evidence that most supply-response tests (such as the typical essay examinations given in school and university courses) have very low scorer reliability. The grade an essay receives seems to depend more on who grades it than on its inherent quality (“beauty is in the eye of the beholder”). However, there is also evidence that this reliability can be increased materially by careful preparation of items and by thorough training of graders, as the public examination bodies do.
- (2) Content reliability, This is evidence that test items are measuring the same thing. In principle, all items on a test should be centered on the same general content area, even though any two items may be quite independent of each other. If the test is not highly speeded, evidence of content reliability may be obtained from one administration of a single form of a test. A common way of doing this is by the use of an internal consistency measure such as the Kuder-Richardson formula given below.

KR²¹ Formula

$$r = \frac{k}{k-1} \left[1 - \frac{m(k-m)}{ks^2} \right] \quad (\text{for internal consistency})$$

Where r = correlational coefficient
 k = number of items in the test
 m = mean score in the test
 s = S.D

Another common way of measuring reliability is by the use of split-half (sometimes called odd/even) reliability coefficient

measured in the Spearman-Brown coefficient of equivalence formula below

$$r_{X_1X_2} = \frac{2r^{1/2}}{1 + r^{1/2}} \quad \text{or} \quad \frac{N\sum X_1X_2 - (\sum X_1)(\sum X_2)}{\sqrt{[N\sum X_1^2 - (\sum X_1)^2][N\sum X_2^2 - (\sum X_2)^2]}}$$

Where $r_{X_1X_2}$ = split-half reliability

X_1X_2 = scores from the two halves of the test respectively.

However, this criterion correlation coefficient is often an underestimate of the test's reliability, for longer tests tend to be more reliable than shorter ones and we have correlated two half-length tests. It is important to emphasize that neither the internal consistency nor split-half approaches in calculating reliability of a test may be used where speed is an important factor in determining a person's test score.

- (3) Temporal reliability, i.e. the stability of a test over time: This is assessed by giving the same test to the same group of subjects at two different times and correlating the scores of the first and second testings. The flaw in this procedure is the possibility of some subjects remembering specific items if the time interval between the first and second administration of the test is very close, which will of course influence the results.

Other factors affecting reliability are:

- (a) Length of test: As a general rule, the longer the test the more reliable it is likely to be cet par. (i.e. that the group tested is the same, that the new items are as good as those on the shorter test, and that the test is not too long to produce fatigue). In other words, chance plays a much greater role in influencing test scores on short tests than on longer ones
- (b) Heterogeneity of group: Other things being equal, higher reliability coefficients will be found for groups which vary more in ability.
- (c) Irregularities will reduce reliability coefficient, e.g. failure to follow directions (hot, poorly-ventilated, ill-lit, noisy room), personal illness, uneven motivation of tester and subject, cheating, etc.

SELF-ASSESSMENT EXERCISE

- i. What factors affect the reliability of a test?

3.3 Types of Measurement

Psycho-educational tests may be classified under three major types

- (1) **Norm-referenced test (NRT)** is a type of test, assessment or evaluation which yields an estimate of the position of the tested individual in a predefined population, with respect to the trait being measured. Tests that set goals for students based on the average student's performance are norm-referenced tests. Many college entrance examinations and nationally used school tests e.g. SAT, GRE, WASC use norm-referenced tests which compare individual student performance to the performance of a normative sample, usually given as age norms, grade level norms, gender norms, etc., and expressed in percentile scores. With a norm-referenced test, grade level pass was traditionally set at the middle 50% of scores.

An obvious limitation of norm-referenced test is that it cannot measure progress of the population as a whole, only where individuals fall within the whole. A norm-referenced test does not seek to enforce any expectation of what all students should know or be able to do, other than what actual students do demonstrate. Present levels of performance and any inequity are taken as fact, not as defects to be removed by a redesigned system.

- (2) **Criterion-referenced test (CRT)** is one that provides for translating test scores into a statement about the behavior to be expected of a person with that score, or their relationship to a specified subject matter. Most tests and quizzes written by school teachers are criterion-referenced tests. The objective is simply to see whether or not the student has learned the material.

The word "criterion" is infact often misunderstood. Many criterions involve a cut-off score, where the examinee passes if their score exceeds the cut-off score and fails if it does not. But this does not mean that the criterion is the cut-off score; the criterion is the domain of subject matter the test is designed to assess. For example, the criterion may be that "students should be able to correctly add two single-digit numbers", and the cut-off score may be that students should correctly answer a minimum of 80% of the questions to pass.

Mastery level measurement, a variant of criterion-referenced measurement, operates on the fundamental belief that all students will perform at one uniformly high level in a standard based system if enough incentives and punishments are put in place.

Standard based educational reform is based on the belief that public education should establish what every student should know and be able to do. Students should be tested against a fixed yardstick rather than against each other or sorted out into a mathematical bell curve. A rank-based system produces data which tell which average students perform at an average level, which students do better, and which students do worse.

A mastery test does mean identifying whether the examinee has “mastered” a specified level of the subject matter, by comparing their score to the cut-off score. But not all criterion-referenced tests have a cut-off score, and the score can simply refer to a person’s standing on the subject domain. A test which is designed to accurately assess mastery may use different questions than one which is intended to show relative ranking, even when testing similar topics.

A criterion-referenced test will use questions which were correctly answered by students who know the specific material, while a norm-referenced test will use questions which were correctly answered by the “best” students and not correctly answered by the “worst” students.

- (3) **Ipsative measurement** connotes a specific type of measure in which respondents compare two or more desirable options and pick one which is most preferred (otherwise known as a forced choice scale). In this sense, ipsative measures may be more useful, than say the Likert-type scales, in evaluating traits within an individual, and for identifying faking.

In education, ipsative assessment is the practice of assessing present performance against the past performance of the person being assessed. Ipsative assessment features heavily in physical education and also in computer games. Encouraging pupils to beat their previous scores can take peer pressure out of students, and eliminates the competitive element associated with norm-referenced test.

Other methods of gathering data about individuals and groups apart from tests, include questionnaires, interviews, observation procedures, inventories (personality, anxiety, adjustment, interests, study habit), check list/rating scales (values, beliefs, etc). (See Unit 2 for more information on these other methods).

3.4 Scales of Measurement

A particular way of assigning numbers or symbols to measure something is called a scale of measurement. There are different scales or levels of measurement that involve different properties (relations and operations) of the numbers or symbols that constitute the measurements. It should be noted that measurement level depends on the correspondence between the measurements and the attributes.

Stevens (1946, 1951, 1975) identified four basic levels of measurement: (i) nominal, (ii) ordinal, (iii) interval, and (iv) ratio. This basic typology is still widely adopted in psycho-educational measurement, although some controversies have been raised, particularly about nominal and ordinal classifications, e.g. Coombs (1953), Tukey (1959), Labovitz (1970), Duncan (1986), Michell (1986, 1999), Rasch (1960), Moyer (1981), Luce (1997).

In **Nominal Level Measurement**, the numerical value assigned to the data category serves merely as a symbol, a label or name, and there is no assumption of inherent ordering or distance. For example, the village of one's birth is a nominal variable, since there is no inherent ordering implied in such a variable. Variables/data assessed on a nominal scale are also called categorical variables/data. The properties of the real number system (i.e. being able to add and multiply numbers, etc.) cannot be applied to such numerically coded categories. Other examples would include the jersey numbers of football players, or religious identification.

Ordinal Level Measurement implies some order relation or rank-ordering of the categories according to some criterion. Each category has a unique position relative to the other categories, i.e. it can be said that a particular category is higher in value than some categories and lower in value than others (except of course, the highest and lowest categories). Examples of this are social class classification, or grades (A,B,C,D, etc.) for academic performance. But the rank ordering does not tell us the distance between the categories. The mode or median score can be used as a measure of central tendency or location of an ordinal attribute, but not the mean.

Interval level measurement: In addition to ordering, the interval level measurement has the additional property that the distances between the categories are defined in terms of fixed and equal units. The interval scale allows us to study differences between things but not their proportionate magnitudes, since an interval scale does not have an inherently determined zero point. Any zero point on an interval scale is arbitrary, and negative values can be used.

In interval level data, there is a meaningful continuous scale of measurement such that equal differences between values in the scale genuinely correspond to real differences between the physical quantities that the scale measures. A good example is the thermometer which records temperature in terms of degrees, and a single degree implies the same amount of heat whether at the lower or upper end of the temperature scale. But, while it is true that the difference between say 30°F and 31°F is the same as the difference between 80°F and 81°F, it would not be correct to say that twice as much as heat is present at 80°F than at 40°F, since Fahrenheit is not a ratio scale. Other examples of this are calendar dates and the IQ metric.

Ratio level measurement: According to Michell (1997, 1999), the ratio scale is the estimation of the ratio between a magnitude of a continuous quantity and a unit magnitude of the same kind. In ratio level measurement, all the properties of an interval scale are present, with the additional property of an inherently defined, non-arbitrary zero point. This property of a fixed and given zero point means that ratio comparisons can be made, as well as distance comparisons. For example, it is quite meaningful to say that a 6ft tall man is not only 3ft taller than but twice as tall as a 3ft boy. Similar examples would be weight in kilograms, duration in seconds, and temperatures in degrees, Kelvin electric charge in volts, the pitch of a tone, the volume of a sound (in decibels), the strength of an attitude.

Since ratio-level measurements satisfy all the properties of the real number system, any mathematical operations appropriate for real numbers can be applied to ratio-level measures. Also, the central tendency of a distribution using ratio-level variables can be represented by the mode, median and arithmetic and geometric means, while the statistical dispersion of the scores of such a distribution can be measured by the range, SD, and a coefficient of variation.

In many real-life situations, a scale of measurement may not correspond precisely to any of these measurements. For example, there can be a mixture of nominal and ordinal information in a single scale, such as in questionnaires that have non-response categories. It is common to have scales that lie somewhere between the ordinal and interval levels in that the measurements can be assured to be a smooth monotone function of the attribute.

4.0 CONCLUSION

Having stated an operational definition of measurement, its attributes, types and scales, were next elaborated upon against the background of the given definition. Thus, the attributes of validity, reliability, and

usability of a measurement instrument and process were discussed; the concepts of norm-referenced and criterion-referenced tests, as well as ipsative measurement, and the conventionally accepted scales of measurement were adumbrated.

5.0 SUMMARY

In this unit, we have affirmed that ability to measure any phenomenon in the scientific sense is obviously crucial to research in general and to psycho-educational research in particular. It is therefore important for the researcher to understand clearly what is meant by measurement, and its main attributes, types and scales. This will assist him/her in the whole process of measurement instrumentation and general data processing techniques discussed in later units.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Differentiate between norm-referenced and criterion-referenced measurement.
- ii. Write a short essay not exceeding two pages on the validity of measurement procedures.

7.0 REFERENCES/FURTHER READING

- Badmus, G.A. (1977). *Problems and Issues of Criterion-referenced Measures: A cautionary note to curriculum evaluators*. West African Jul. of Educational and Vocational Measurement, 3,3, 5 & 13, Ahmadu Bello University, Zaria, Nigeria.
- Badmus, G.A. (1997). *Assurance of quality in educational assessment through application of recent advances in Testing, Measurement and Statistical Techniques*. Benin City: National Business and Technical Examination Board.
- Badmus, G.A. & C.N. Omoifo (1998). *Essentials of Measurement and Evaluation in Education*. Benin City: Osasu Publishers.
- Cronbach L.J. (1970). *Essentials of Psychological Testing*. Harper and Row, New York.
- Guilford, J.P. (1965). *Fundamental Statistic in Psychology and Education*. New York: McGraw-Hill Book Company.
- Netty, H.J. (1996). *Advances in Test Validation*. In G.A. Badmus & P. Odor (eds), *Challenges of managing Educational Assessment in Nigeria*, pp. 137-147, Kaduna: Atman Publishers.

- Nwana, O.C. (1981). *Educational Measurement for Teachers*. London: Nelson.
- Ohuche, R.O & Akeju, S.A (1977). *Testing and Evaluation in Education*. Lagos: African Educational Resources.
- Okoh, N. (1983). The Problems of Measurement in Psychology and Education. In Nduka Okoh (ed.), *Professional Education: A Book of Readings*. Benin City: Ethiope Publishing Corporation.
- Okoh, N. (1984). *Construct Validation of the Torrance Tests of Creative Thinking (TTCT): A Cross-cultural Study*. Nigerian Educational Forum, 7,1, Ahmadu Bello University, Zaria, Nigeria.
- Okoh, N. (1986). *Cross-cultural Study of the concurrent validation of M-Test of non-verbal intelligence*. Benin Jul. of Social Sciences, 1,2, University of Benin, Nigeria.
- Stevens, S.S. (1946). *On the theory of scales of measurement*. Science, 103, 677-680
- Thorndike, E.L. (1927). *The Measurement of Intelligence*. New York: Wiley.
- Thorndike R.L & Hagen E. (1986). *Measurement and Evaluation in Psychology and Education*, 4th ed. Wiley, New York.

UNIT 2 CONSTRUCTION AND VALIDATION OF MEASUREMENT INSTRUMENTS

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Construction and Validation of Measurement Instruments
 - 3.1.1 Test Construction
 - 3.1.2 Questionnaires
 - 3.1.3 Interview Schedules
 - 3.1.4 Rating Scales/Checklists
 - 3.1.5 Attitude Scales/Interest Inventories
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Now that you have been introduced to the meaning and the rationale of measurement in educational research, the basic steps in test construction as well as discussion of some of the data-gathering instruments, will next be considered.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- describe the major processes involved in test construction
- list some of the general guidelines for writing multiple-choice items
- assess the merits and shortcomings of either (a) questionnaire or (b) interview schedules
- differentiate between rating scales and checklists
- contrast attitude scales and interest inventories.

3.0 MAIN CONTENT

3.1 Construction and Validation of Measurement Instruments

3.1.1 Test Construction

Three major processes are involved in test construction:

- (1) Formulation of objectives of the test.
- (2) Construction of a table of specifications or blue print.
- (3) Writing, organization and validation of test items.

A common starting point in test construction is the description and clarification of the objectives of the instructional program that the test purports to measure. Objectives are the skeletal foundation of any learning task and in this case are generally derived from subject syllabuses or curricula. They state in general and specific behavioral terms what the learner is supposed to have learnt or be able to do at the end of the learning exercise. They should therefore be clearly defined in terms of levels and specificity of understanding of the subject. For this purpose, Bloom's taxonomy of education objectives (1956) is the most often used classification, especially the cognitive domain which Bloom categorized into knowledge, comprehension, analysis, synthesis and evaluation. Tests and examinations are used for checking achievement of the objectives stated.

The content area of the test, which is derived from the subject syllabus or curriculum, is usually broken down into about four or five chunks which must of course adequately sample the universe or domain of knowledge to be examined. For this, the test constructor will need to produce a number of draft test items, about 50% more than the number in the final test, show these to subject experts and test construction experts, revise the questions if necessary, arrange the questions in apparent ascending order of difficulty, and try them out on a relevant group of testers. Here, it should be borne in mind that the greater the number of items in a test, the greater the reliability of the test.

The next step is to develop a blueprint or table of specifications, which is simply a two-dimensional cross-tabulated grid or matrix specifying the processes, objectives and content area to be assessed. The idea is to ensure congruence between your instructional objectives and the test items. In other words, your test items should match the objectives, by providing opportunities for testees to exhibit that behavior. (See Table 1 below). Each cell in the table corresponds to a particular task and subject content. By specifying the number of test items you want for each task

and each cell, you can determine how much emphasis to give each task and each content area.

Table 1: Specimen Test Blueprint (Or Table Of Specifications) In Test Construction

	O B J E C T I V E (Cognitive domain)						
CONTENT	Know 34%	Comp. 20%	App. 18%	Analysis 12%	Synthesis 10%	Evaluation 6%	Total Items
1.	4	2	2	2	1	1	12
2.	5	3	2	2	1	1	14
3.	5	3	3	2	2	1	16
4.	4	2	2	2	1	1	12
5.	2	2	1	1	0	0	6
Total Items	20	12	10	9	5	4	60

Test format may take the form of essay or objective tests; the former allows students to select, organize and supply the answer in essay form, whilst the latter allows the students to select the one correct answer from a number of given alternative answers or options (as in multiple-choice and matching questions), in which the incorrect options are called distractors. In terms of comparative advantage, the objective test, in addition to measuring recall, can also be designed to measure understanding, application and other more complex outcomes, whilst the essay test is more efficient in measuring deeper and higher levels of learning.

The large number of questions in an objective test permits a broader sampling of course content, in contrast to the essay test which has fewer tests. However, the highly structured nature of objective tests leaves it prone to guessing, while on the other hand, the less structured essay test may be influenced by writing ability and by bluffing. Objective tests encourage development of broad background of knowledge and abilities, whilst essay tests encourage organization, integration and effective expression of ideas.

In terms of scoring, objective tests are usually marked only right or wrong, and therefore scoring is easily accomplished with consistent results, while essay tests are time consuming to score and require special measures for consistent results.

Guessing, however, in case of multiple choice questions can be ameliorated by use of the formula:

$$K = \frac{1}{x - 1},$$

(Where k = correction factor for wrong answer, and x = number of alternative answers given). A more general formula to control for guessing in the entire test is

$$S = R - \frac{W}{x - 1}$$

(Where S = True score, R = number of correct answers, W = number of wrong answers and x = number of alternative answers).

SELF-ASSESSMENT EXERCISE

- i. Explain the three major processes involved in test construction.

Some general guidelines for writing test items are shown in the figure 1 below (Gronlund 1971, Conoley and O'Neil 1979)

FIGURE I: GENERAL GUIDELINES FOR ITEMS WRITING***Typical rules for multiple-choice items:***

1. The stem of the item should be meaningful itself and should present a clear problem.
2. The stem should be free from irrelevant material.
3. The stem should include as much of the item as possible except where an inclusion would clue. Repetitive phrases should be included in the rather than being restated in each alternative.
4. All alternatives should be grammatically consistent item stem and of similar length, so as not to provide a clue to the answer.
5. An item should include only one correct or clearly best answer.
6. Items used to measure understanding should contain some novelty and merely repeat verbatim materials or problems presented in instruction.
7. All distractors should be plausible and related to the body of knowledge and learning experiences measured.
8. Verbal associations between the stem and correct answer or stereotyped phrases should be avoided.
9. The correct answer should appear in each of the alternative positions with approximately equal frequency.
10. Special alternatives such as “none”, “all of the above” should be used sparingly.
11. Avoid items that contain inclusive terms (e.g., “never”, “always”, “all”) in the wrong answer.
12. Negatively stated items should be used sparingly.
13. Avoid alternatives that are opposite in meaning or that are paraphrases of each other.
14. Avoid items which ask for opinions.
15. Avoid items that contain irrelevant sources of difficulty, such as vocabulary, or sentence structure.
16. Avoid interlocking items, items whose answers clue responses to subsequent items.
17. Don't use multiple choice items where other item Formats are more appropriate.

Typical rules for short answer and completion items:

1. A direct question is generally better than an by incomplete statement.
2. Word the item so that the required answer is both brief and unambiguous.
3. Where an answer is to be expressed in numerical units, indicate the type of units stem wanted.
4. Blanks for answers should be equal in with the length. Scoring is facilitated if the blanks are provided in a column to the right of the question.
5. Where completion items are used, do not leave many blanks.
6. For completion items, leave blank only those things that are important to remember.
7. In completion items, don't take statements overbatim from students' textbook or instruction.

Typical rules for true-false or alternative response items:

1. Avoid broad general statements for true-false items.
2. Avoid trivial statements.
3. Avoid negative statements and especially double negatives.
4. Avoid long complex sentences.
5. Avoid including two ideas in a single statement unless cause-effect relationships are being measured.
6. Include opinion statements only if they are attributed to particular sources.
7. True statements and false statements should be approximately the same length.
8. The number of true statements and false statements should be approximately equal.
9. Avoid taking statements verbatim from students' text or instruction.

(Gronlund 1971, Conoley and O'Neil 1979)

SELF-ASSESSMENT EXERCISE

- i. State in your own words some of the general guidelines for constructing multiple-choice questions.

Other methods of gathering data about individuals and groups apart from tests, include questionnaires, interviews, observation procedures, inventories (personality, anxiety, adjustment, interests, study habit), check list/rating scales (values, beliefs, etc).

3.1.2 Questionnaires

A questionnaire is a self-report instrument for gathering information from a respondent about variables of interest to an investigator, consisting of a number of questions or items on paper that the respondent reads and answers (Wolf, 1988). The items may be structured (e.g. gender, male/female) or unstructured (e.g. how did you spend the last Xmas holiday?), and the investigator will need to maintain a delicate balance between both types of items. A questionnaire of course assumes that (1) the respondent can read and understand the questions or items, and (2) crucially that he/she is able and willing to answer the questions or items honestly.

The content of a questionnaire can be almost limitless, but clearly the interests of the investigator, the sensitivity or delicacy of the questions (i.e. privacy concerns), and time factor are obvious constraints to what can be included in a questionnaire. The last element of completion time will definitely affect issues of respondent fatigue, boredom, and cooperation.

The interests of the investigator should be limited to variables of primary interest to him/her, and should as much as possible be explicitly related to a particular research question or hypothesis — in an effort to avoid an unduly lengthy questionnaire which may increase the likelihood that respondents may not answer. The second constraint is obvious: asking pejorative or highly personal questions (such as sexual behavior and religious attitudes) generally has the effect of making respondents refuse to answer such questions, or give false (e.g. socially desirable) responses, or to simply throw away the entire questionnaire. Ideally, it is better to set a reasonable time limit within which nearly everyone can complete the questionnaire (roughly 15 – 30 minutes).

Obvious steps in questionnaire development are: (i) identification of the population and other variables to be studied; (ii) collection of data which will depend on the nature of the target population and the amount of resources available for the study; (iii) writing, organization and try-out of the questions with a pilot population from the target population. Research results and experience have shown that it is better to place supplementary classificatory items such as age, sex, parent's occupation, etc. at the end of the questionnaire rather than at the beginning; (iv) administration of the questionnaire.

3.1.3 Interview Schedules

The interview process is a face to face encounter which emphasizes more of the interpersonal relationship as well as a “shared meaning”

between respondent and interviewer. An interview can in fact be regarded as an oral questionnaire.

The steps in conducting an interview are similar to those of producing a good questionnaire given in the earlier section: i.e. (a) defining the general goals and objectives of the interview; (b) developing the interview schedule in terms of question format and appropriate questions; (c) developing appropriate interviewing techniques; and (d) developing satisfactory recording/coding of responses.

Like the questionnaire method, the interview may be formal/structured or informal/unstructured. The former uses an interview schedule which contains set questions to be asked in a specific order and where answers are either written down or coded in some way. Unstructured interview, on the other hand, usually uses fewer questions whose order and form are not specified, and the interviewee may be probed or encouraged to expand on his/her answers.

The response process involves (1) understanding and proper interpretation of the questions (which are a function of vocabulary level, clarity of concept, complexity of sentence structure, and other familiar issues of question wording), (2) cognitive and affective processing of information by the respondent and deciding how best to respond. Here, inaccurate and incomplete responses may arise from social desirability bias (need for approval), acquiescence bias, conformity bias, and other such distortions. The issue of rapport (gaining the trust and perhaps affection of the respondent) is of course crucial in interview technique of gathering information, but a good balance should be maintained between a friendly, empathic style and a neutral, standardised and non-directive approach.

The four main components of the interviewer's job are, briefly: (a) to introduce the interview to the person and convince him/her to accept it and to comply with the expected respondent role; (b) to administer the questionnaire using standard procedures; (c) to follow-up inadequate responses, using acceptable non-biasing and non-directive probes or feedback; and (d) to record the response accurately and completely and appropriately code the protocols (Miller and Carnell, 1988).

Successful interviewing often requires some training and selection of interviewer in acquiring interviewing skills involving role play and evaluation of speaking, reading and writing modes.

3.1.4 Rating Scales/Checklists

A rating scale is a paper and pencil device used to describe or appraise human performances or products (Wolf, 1988). Typically, it consists of a number of trait names and a number of categories that are used to represent varying degrees of the traits, and the rater is asked to rate one or more persons or objects on the trait by assigning a number, letter, adjective or description that is judged to best fit the trait. In education and psychology, rating scales are commonly used in the measurement of attitudes and in personality assessment. They are developed to ameliorate some of the subjectivity in unstructured interviews, referees reports and other such reports by the incorporation of some quantitative scale treatment of the data, and systematic descriptions and appraisals of the performances and products.

However, rating scales do not often meet with the above expectations if the rater is unable or unwilling to rate accurately or conscientiously, for the following reasons: (i) feeling of botheration, (ii) friendship ties, (iii) insufficient knowledge of the person or object/phenomenon to be rated, (iv) ambiguity in what is expected of him/her, (v) the lack of a uniform standard, (vi) extraneous factors like generosity error, halo effect, central tendency error. All these factors contribute to lowering the reliability and validity of rating scales. However, these problems can be ameliorated by refining the form of the response variables, e.g. use of Flanders interactive categories and also through training programs for raters.

A Checklist is a list of things or statements which a respondent has to go through and check (or mark) those that apply to him/her under the stated situation. It is often used in observation situations.

SELF-ASSESSMENT EXERCISE

- i. How are rating scales and checklists used in psycho-educational measurement?

3.1.5 Attitude Scales/Interest Inventories

Both attitudes and interests have been described as measuring the outer layers of personality in settled ways of behavior, and are measured by self-report inventories or scales in which the individual expresses his/her preferences of given opinions, activities or job labels by way of ranking True/False or Likert-type responses, or by combinations of these forms of responses.

Attitude has been defined briefly as “a non-observable but organized and enduring personal disposition which forms the basis of human behavior” (Okoh, 1968). Attitudes determine for each individual what he will see and hear, what he will think and what he will do, and what he will get out of any situation or activity. As William. James succinctly put it, attitudes “engender meaning upon the world”.

An attitude scale is a measuring device which consists of a series of statements which the individual is invited to express his/her degree of agreement or disagreement with, or sometimes simply whether he or she agrees with the statement or not.

Three main types of attitude scale are frequently used:

- (1) Likert scale usually consisting of any number of statements about an attitude object and scored on a five point rating system of Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), and Strongly Disagree (SD) to which are assigned numerical response values of 5, 4, 3, 2 & 1 respectively. The scoring is reversed for negatively worded items. Some examples are
 - (i) University students should take an active part in politics
SA A U D SD
 - (ii) Seminar (or tutorial) meetings are much more useful than lectures
 - (iii) “A little learning is a dangerous thing”
 - (iv) Examinations are an unreliable index of academic activity
 - (v) Education creates a lack of respect for traditional customs and values

The Likert scale is very popular in behavioral research since it is easy to construct, and yields greater variance results than the two or three response categories options.

- (2) Thurstone scale: is an equal appearing interval scale and uses the method of paired comparison which hypothesizes that it is possible to find scale values for statements reflecting positive or negative attitudes and that subjects could be measured in terms of the scale values for statements they agreed with. So, essentially the method depends on the possibility of reliably scaling statements in terms of the attitude level of persons who would endorse the statement.
- (3) Guttman scale: is a cumulative unidimensional scale in which items are arranged in a scalable order such that an individual who responds positively to any particular item also responds positively to all other items having a lower rank, and vice versa. The items in the scale are assumed to be differentially ordered, rather than equal, and arranged in random order. Each item is

assigned a scale value indicating the strength of attitude for an agreement response to that item. An example would be: “With regard to foreigners (non-Nigerians), how close a relationship would you like to have with them? (i) marriage, (ii) personal friend or club member, (iii) street or quarters neighbour, (iv) colleague in the same office or occupation, (v) citizen of my country, (vi) visitor to my country, (vii) personae non grata (to be excluded from my country if possible).

Interests, on the other hand and more specifically, are expressions of what phenomena an individual personally likes or dislikes, and the strength and direction of that preference. They include (1) measures of patterns of general interest, (2) measures of patterns of occupational interest, and (3) measures of value.

Interests measurement may take two forms: (a) expressed interest which basically involves asking the person what activities or jobs he/she likes and dislikes, and (b) inventoried interests. Campbell (1973) has distinguished a number of uses for interest inventories, among which he listed

- (a) to aid the individual in making educational and vocational choices
- (b) to confirm choices already made
- (c) as an aid in understanding the world of work
- (d) as a guide in planning self-development
- (e) to help people understand or, possibly, diagnose job dissatisfaction.

The first widely used and still the most popular interest inventory was the Strong Vocational Interest Blank, developed in 1927 by E.K Strong, and since 1974 (further revised in 1981) merged into the Strong-Campbell Interest Inventory. The test contains 325 activities, subjects, etc. Takers of this test are asked whether they like, dislike, or are indifferent to 325 items representing a wide variety of school subjects, occupations, activities, and types of people. They are also asked to choose their favorite among pairs of activities and indicate which of 14 selected characteristics apply to them. The Strong-Campbell test is scored according to 162 separate occupational scales as well as 23 scales that group together various types of occupations (“basic interest scales”). Examinees are also scored on six “general occupational themes” derived from J.L. Holland’s interest classification scheme (realistic, investigative, artistic, social, enterprising, and conventional).

The other most commonly administered interest inventory is the Kuder Preference Record, originally developed in 1939. the Kuder Preference

Record contains 168 items, each of which lists three broad choices concerning occupational interests, from which the individual selects the one that is most preferred. The test is scored on 10 interest scales consisting of items having a high degree of correlation with each other. A typical score profile will have high and low scores on one or more of the scales and average scores on the rest.

Interest inventories are widely used in vocational counseling, both with adolescents and adults. Since these tests measure only interest and not ability, their value as predictors of occupational success, while significant, is limited. They are especially useful in helping high school and college students become familiar with career options and aware of their vocational interests. Interest inventories are also used in employee selection and classification.

The rationale behind vocational interest inventories are that:

- (1) persons in different occupational groups will in fact have different interests or preferred activities, and
- (2) since the majority of people in a particular occupation will have similar interests, a person having a pattern of interests typical of an occupational group will find satisfaction in that field or activity.

4.0 CONCLUSION

This unit has covered the major processes involved in test construction, and briefly discussed some specific measurement instruments like questionnaires, interview schedules, rating scales and checklists, attitude scales and interest inventories in terms of their construction, use and typical problems.

5.0 SUMMARY

Measurement is the process of assigning numbers or other symbols to some attribute of a set of things in such a way that relationships of the numbers or symbols reflect relationships of the attribute being measured. How this concept is applied in test construction generally, and specifically in some measurement instruments like questionnaires, interview schedules, rating scales/checklist, and attitude scales/interest inventories was the focus of this Unit. This will be followed in Units 3 and 4 of this module by the techniques of data analysis.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Compare and contrast questionnaires and interview schedules in human behavior measurement.
- ii. Discuss attitude scales and interest inventories as measures of the outer layers of personality.

7.0 REFERENCES/FURTHER READING

- Allport, G.W, & Vernon P.E. (1931). *A Study of Values: A Scale for Measuring the Dominant Interests in Personality: Manual of Directions*. Houghton Mifflin, Boston, Massachusetts.
- Ebel, R.L. (1979). *Essentials of Educational Measurement*, 3rd ed. Prentice-Hall, Englewood Cliffs, New Jersey.
- Likert, R. (1932). *A technique for the measurement of attitudes*. *Arch. Psychol.*140.
- Miles, J. (1973). *Eliminating the guessing factor in the multiple choice test*. *Educ. Psychol. Meas.* 33: 937-51.
- Okoh, N. (1968). “*Values in Education*”. Unpublished M.Ed. Thesis, University of Glasgow, Scotland.
- Shaw, M.E & Wright, J.M. (1967). *Scales for the Measurement of Attitudes*. McGraw-Hill, New York.
- Thorndike, R.L & Hagen, E. (1977). *Measurement and Evaluation in Psychology and Education*, 4th ed. Wiley, New York.
- Wolf, R.M. (1990). *Rating Scales*. In J.P. Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford, England, Pergamon Press.
- Wolf, R.M. (1990). *Questionnaires*. In J.P. Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook*. Oxford, England, Pergamon Press.

UNIT 3 DESCRIPTIVE DATA ANALYSIS

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Descriptive Data Analysis
 - 3.1.1 Basics of Data Processing
 - 3.1.2 Descriptive Data Analysis
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Descriptive data analysis uses the statistical measures of central tendency, measures of dispersion or variability, measures of relationship, and measures of relative standing to try to give meaning to findings of research which used mostly nominal or ordinal data.

2.0 OBJECTIVES

After successfully studying this unit, you should be able to:

- define descriptive data analysis
- describe and apply the main measures of central tendency
- explain the principles underlying the normal distribution
- discuss and evaluate the various types of test norms
- apply the formulae for the computation of correlation coefficients.

3.0 MAIN CONTENT

3.1 Descriptive Data Analysis

3.1.1 Basics of Data Processing

The last fundamental stage of the research exercise is the processing of data derived from the study, in order to be able to answer/test meaningfully the questions/hypotheses earlier raised in the research problems/hypothesis. For this purpose statistical procedures are generally employed, especially in quantitative research, since statistics is concerned with describing, synthesizing, analyzing and interpreting

quantitative data. In psycho-educational research, two main types of statistics are employed, i.e. (1) descriptive statistics and (2) inferential statistics.

3.1.2 Descriptive Data Analysis

This is used to describe large number of scores with a small number of indices raised from either a sample (statistic) or the total population (parameter). The main types of descriptive statistics are (i) measures of central tendency, (ii) measures of spread or variability, (iii) measures of relationship, and (iv) measures of relative standing.

Measures of central tendency

These reflect the typical or average score in a distribution of raw scores, and are denoted by indices of mode, median and mean. The Mode is the score which is obtained by more members of the group than any other score. Most distributions of scores are unimodal in the sense that there tends to be only one peak of scores within the distribution, but occasionally some distributions do have two or more peaks of scores and are said to be bimodal or multimodal. The Median on the other hand is that figure below which and above which 50% of the group comes, i.e. the middle score. By inspection, a score higher than the median comes in the top half of the group whilst a score lower than the median comes in the bottom half of the group. In a distribution with an even number of cases, the median score becomes the midpoint between the two scores on either side of the middle. For example, the median score in a score distribution of 20, 21, 23, 24, 26, 28 is 23.5 (i.e. $23 + \frac{1}{2}$). Finally, the Mean is simply the average mark obtained by the members of the group, i.e adding up all the raw scores obtained by the members of the group and dividing this total by the number of members of the group. The formula for calculating the mean score is

$$\bar{X} = \frac{\sum X}{N} \quad \text{Where,}$$

\bar{X} = mean or average score

\sum = the sum of

X = each individual raw score obtained by members of the group

N = number of members of the group

We can therefore inspect a distribution to see if a particular score is above or below the mean or average score for that group. The mean can also be used to compare the performance of two different groups on the same test. The mean can also be calculated from grouped scores by the use of frequency table and class intervals (detailed calculation of these central tendency measures can be seen in any elementary statistics text).

In practice, in the majority of distributions the mean score the median score and the modal score tend to be fairly close to one another.

Measures of Dispersion

Measures of variability describe how variable, dispersed or spread the scores are with respect to the typical score or central tendency. The first of such measures is the Range which is a measure of the dispersion of scores in a distribution and is calculated as the lowest score obtained in a group subtracted from the highest. Another and more important measure of dispersion is the Standard Deviation (SD) which is in effect an average of the departure of people's scores from the group mean. It is given by formula

$$SD = \frac{1}{N} \sqrt{N \sum X^2 - (\sum X)^2} \quad \text{or} \quad \sqrt{\frac{\sum f(X - \bar{X})^2}{N}} \quad (\text{for grouped data})$$

Where

$\sum X^2$ = the sum of all the scores squared

$(\sum X)$ = square of the sum of the scores

f = frequency of the individual class intervals

The S.D gives an indication of the degree of dispersion of the scores from the mean and an estimate of the variability in the total sample. It is most usefully employed in comparing the variability of different groups.

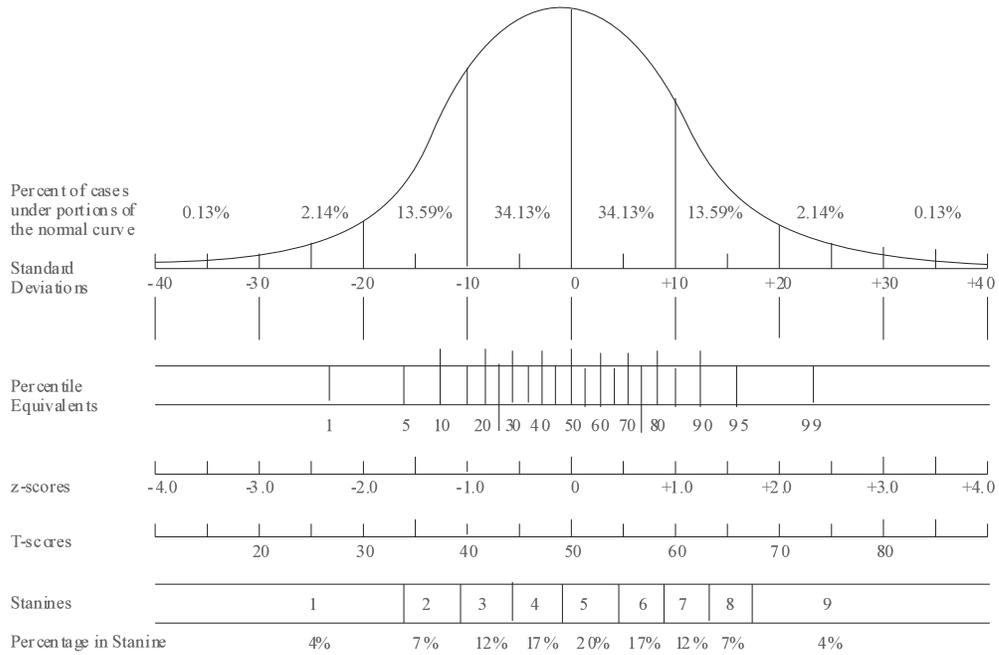
The Variance is another measure of dispersion. It is closely related to the S.D, and is simply the SD squared

$$V = SD^2$$

The Normal Distribution

If you took a large random sample of adult Nigerian males and measured their height, you would notice that there were a lot more men of about average height than there were very tall men or very short men. If you construct a frequency polygon of their heights, you would see that they were distributed in the form of what is called a normal distribution (sometimes also called the bell-shaped curve or the Gaussian curve). It is assumed, for example, that intelligence and aptitude are distributed normally throughout the population. The point of interest is the relationship between the normal distribution and the SD. In a normal distribution, 68.26% of the scores lie within one SD of the mean score, 95.44% of the scores lie within two SDs of the mean, and 99.73% of the scores lie within three SDs of the mean (see Figure I below)

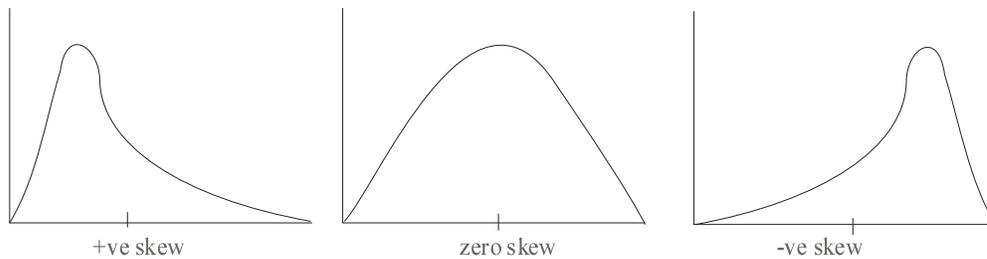
Figure 1 Relationship Between Some Common Score Scales in the Normal Distribution



Skew

It is sometimes found that a distribution is not normal. A distribution is said to be skewed when scores are massed at one or the other end of the score scale. Scores are negatively skewed when they are massed at the upper end as a result of the test being relatively easy for the group concerned, whilst positive skew is the result of the test being relatively difficult

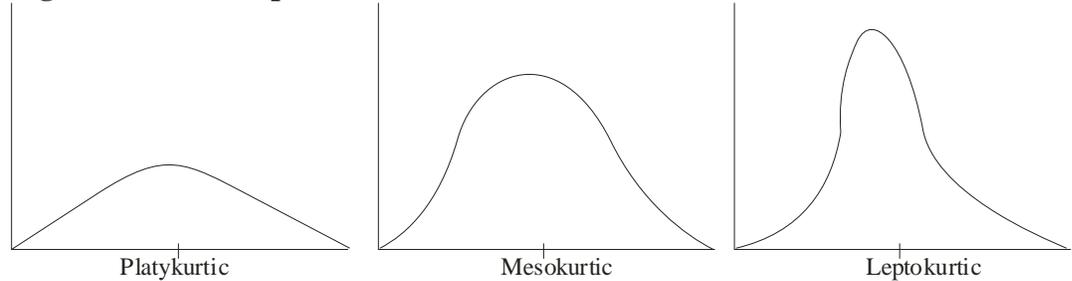
Figure 2 Examples of Skewed Distributions



Kurtosis

The term kurtosis refers to the peakness or flatness of a frequency distribution. A peaked distribution is said to be leptokurtic, a flat distribution platykurtic, whilst the normal distribution is a mesokurtic distribution of zero skew.

Figure 3 Examples of Kurtosis



Test Norms

Norms are data which inform us how other people have performed on a test. In a norm referenced test, to know that an individual has a raw score of 47 out of a total possible score of 50 is not meaningful without appropriate and adequate norms to refer to. It is from the relevant norm tables that percentiles, IQ and the various types of standard scores are obtained. A testee's raw score (i.e. the sum of the "correct" or keyed item responses) is referred to the distribution of scores obtained by the standardization sample in order to discover where he/she comes in the norm group distribution. Among the common norm systems used in psycho-educational testing are (1) percentile, (2) letter-grade system, and (3) stanines scores.

- (1) *Percentiles:* A percentile is a score below which a certain percentage of the members in the group come. For example, the 90th percentile is the point below which 90% of the members of a group come, and the 20th percentile is the point below which 20% of the members of the group come. The 50th percentile is the same as the median. From a percentile norms table, raw scores are read against the appropriate percentiles for that group. A good norm table must contain the size of the standardization sample, the mean raw score, and SD raw score units.

Percentiles have the advantage that they are easily obtained and understood. At the same time they have the drawback that they are not equal units of measurement, and tend to exaggerate differences near the mean and collapse differences at the extremes. Accordingly, percentiles must not be averaged nor treated in any other way mathematically. Percentiles are an

example of ordinal measurement, i.e. they only give an indication of rank order.

- (2) *Letter-grade norm system:* Letter grades are frequently used to denote performance, instead of raw marks, in many educational and testing institutions. This may vary from the 5 points scale (A to E), to the 9 points scale which is a refinement of the 5 points scale by introducing intermediate scale points.

A letter grade represents a range of performance, since letter grading groups a number of consecutive marks and assigns them the same grade (e.g. 70 – 100 = A). The letter grade system has the advantage of being easy to interpret, and is useful in that it encourages us to think in terms of bands of scores. But it has the disadvantage that it can be misleading for those cases on the borders of the grade, e.g. a raw score difference of 1 which may have occurred purely by chance can mean the difference between A or B grade. Also, because of the loss of precision in letter grades the adding and averaging of such grades is not feasible.

- (3) *Stanines:*

The stanine method of reporting achievement is essentially the same as the 9-point letter grade scale, except that instead of using letters as identification marks for the grades, it uses the numbers 9 down to 1 such that stanine 9 will represent 4% of the distribution, as illustrated below:

9	4%
8	7%
7	12%
6	17%
5	20%
4	17%
3	12%
2	7%
1	4%

This implies that the range of marks in the distribution must be wide enough and the sample population also relatively large, as with many public examinations like GCE, SSCE and UME. Stanines, unlike the 9

point letter grades, can be added and averaged since they are numerical scores.

SELF-ASSESSMENT EXERCISE I

- i. Describe the common norm systems used in psycho-educational testing.

Standard Score

A standard score is a derived score based on the mean and SD of distribution i.e. it indicates how many SDs above or below the mean a score is.

Z-score, which is the basic standard score, is merely a raw score which has been changed to SD units. It is calculated by the formula

$$z = \frac{x - \bar{X}}{SD}$$

Where z = standard score

X = individual raw score

\bar{X} = mean raw score of sample

SD = standard deviation of standardisation sample

Usually when standard scores are used they are interpreted in relation to the normal distribution curve. As can be seen in Figure 1 on page 120, the z-scores are marked out on either side of the mean with those above the means positive and those below the mean negative in sign. Therefore by the calculation of the z – score it can be seen where the individual's score lies in relation to the rest of the distribution.

The standard score is very important when comparing scores from different tests since the scores have been converted to a common scale such as the z-score which can then be used to express an individual's score on different tests in terms of norms. One important advantage in using the normal distribution as a basis for test norms is that the standard deviation has a precise relationship with the area under the curve. One SD above and below the mean includes approximately 68% of the sample. From figure 1 on page 120 it can be seen that standard z-scores have a mean of 0 and a standard deviation of 1. Because of this, z- scores can be rather difficult and cumbersome to handle since most of them are decimals and approximately half of them can be expected to be negative. To remedy these drawbacks, various transformed standard score systems have been derived. These simply entail multiplying the

obtained z-score by a new SD and adding it to a new mean. One example of this is the T-score.

The T-score

One commonly used transformation of the z-score is based on a mean of 50 and a standard deviation of 10. This is known as the T-score and is denoted by the formula $T = z(10)+50$, so that given a z-score of 1.53 translates to $1.53 \times 10+50 = 65$.

The relationship between some common score scales in the normal distribution is also given in Figure 1 of page 120. It will be seen that the mean raw score corresponds to a z-score of 0, to a percentile of 50 and an 1Q of 100. Notice also that percentiles are not equal units of measurement. By comparing their scaling properties with z-scores, at the middle and tail ends of the distribution of raw scores, it is possible to construct a table of z-scores equivalents for the raw scores. This could serve as a norm table, and some tests do in fact present normative data in terms of z-scores, rather than percentiles, letter grades or T-scores.

In summary, standard scores have the advantage that they are equal units of measurement and thus may be manipulated mathematically. They show immediately where a person falls in a distribution and give us scores that can be directly compared both within and between individuals.

SELF-ASSESSMENT EXERCISE II

- i. Comment briefly on standard scores, their rationale and use.

Measure of Relationship

Correlation is concerned with the degree of relationship between two or more variables. In everyday life we constantly infer all sorts of relationships between things, people and their properties, — e.g. clever people have high foreheads, fat people are jolly, clouds bring rain, etc.

Relationships vary in closeness and direction. For example, as the length of one side of a rectangle increases, the area increases in a fixed, perfectly close (or predictable) relationship. The relationship between height and body weight on the other hand is not so close, in that although weight tends to increase with height there are many exceptions to the general rule.

The direction of relationship refers to whether one variable increases as the other increases (positive), or whether one variable decreases as the other increases (negative). With some variables there is a combination of

both positive and negative relationships. For example, strength increases with adulthood and slowly declines with the approach of old age. This is known as a curvilinear relationship for which more complex statistics are required.

The correlation coefficient is the statistic we most often use to measure the relationship between variables. It is in fact a summary index of the relationship between two variables, and is so worked out that it ranges from +1.00 through zero to -1.00. The sign of the coefficient has no meaning other than to indicate the direction of the relationship; +1 indicates perfect positive relationship, and -1 indicates perfect negative relationship. If there is no relationship between the two variables, the correlation is zero. It is important to note that the correlation coefficient is not a ratio scale and therefore it cannot be measured as percentage or probability scales are. For example, a correlation coefficient of 0.2 is not half as good as one of 0.4.

There are several different methods of calculating correlation coefficients. Perhaps one of the most simple is the Spearman Rank Order coefficient of correlation(rho) which is used where the two variables to be correlated are expressed in rank order, and based on the principle that the closer the similarity between the two sets of rankings the higher the correlation. The formula for calculating rho is given as

$$\text{rho } (r) = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

Where D is the difference between ranks, and N is the number of pairs of ranks. The Spearman Rank Order Correlation Coefficient is especially useful where one or both of the variables to be correlated are already in rank order. Unlike the Pearson Product Moment Correlation Coefficient described below, the Rank Order Correlation Coefficient does not rest on the assumption that the variables to be correlated are normally distributed, and it is therefore a better statistic to use with highly skewed distributions.

The most powerful and commonly used of the correlation coefficients is the Pearson Product Moment Correlation Coefficient (r). It is employed with interval or ratio scale variables and is used when a linear (straight line) relationship is suspected between two variables. Ordinarily, it is not affected by the number of people in the group or by differences in the units of measurement. It is achieved first by expressing all raw scores as z-scores, and secondly dividing by the number of people in the group. The computational formula used here is

$$r = \frac{\sum xy}{N SD_x SD_y}$$

Where

r = the product moment correlation coefficient

X = the deviation of a person's raw score on the first variable from the mean of all persons on that variable

y = the deviation of a person's raw score on the second variable from the mean of all persons on that variable

SD_x = the standard deviation of the first variable

SD_y = the standard deviation of the second variable

ξ = the sum of

N = the number in the group

This is the formula often used in hand calculations because it has the advantage that rather than having to transform every raw score to a z-score, we deal in raw score deviations, dividing through by the SDs of the two distributions at the final stage.

Another equivalent formula more commonly used in computer programs is

$$r = \frac{\sum xy - \frac{1}{n}(\sum X)(\sum Y)}{\sqrt{[\sum X^2 - \frac{1}{n}(\sum X)^2][\sum Y^2 - \frac{1}{n}(\sum Y)^2]}}$$

The formula is often easier to apply where the means on the two variables are not whole numbers.

Multiple Correlation

Multiple correlation is used in studies where scores in more than two variables are to be correlated. For example, in occupational selection where more than one set of information or test scores (e.g. interview results, IQ or paper form board) are to be related to a criterion of job success. To calculate the multiple correlation we must know not only the correlation of each test with the criterion but also the correlation between the two tests. The formula is given as:

$$R^2_{123} = r^2_{12} + r^2_{13} \text{ (assuming zero correlation between the two tests)}$$

Where R^2_{123} = the multiple correlation
 r_{12} = the correlation between the first test and the criterion

r_{13} = the correlation between the second test and the criterion

It is obvious that performance on a single aptitude test at selection will not correlate too highly with job performance for any occupational group. Other characteristics, such as IQ, nAch, etc. should be considered and weighted for improved prediction validity. But it should be noted that as we add more and more information or variables, we get less and less of a return. Indeed, adding more than three or four variables usually does little to improve the multiple predictions.

On correlations generally, one should always remember that:

- (1) Correlation does not imply cause and effect, i.e. if A correlates with B does not necessarily mean that A causes B.
- (2) The larger our sample the smaller the correlation coefficient has to be in order to be significant at a particular level. This means that
 - (a) a high correlation coefficient may have occurred by chance alone if our sample is small in size;
 - (b) a statistically significant correlation coefficient may not be very high, and therefore by itself may not be of much particular significance and usefulness, if our sample is large.
- (3) the correlation coefficient reflects the characteristics of the group on which it was calculated. For example, the height and verbal ability correlation would be much higher in 2 – 8 year old children than in adults.
- (4) the correlation coefficient is dependent upon the variability of scores in a sample. Validity coefficients, for example, are largest in a group with a wide range of ability and tend to be small in a restricted group.

What does a correlation coefficient of say +.45 mean? In a test manual we may read that a correlation of +.45 between a verbal ability test and typing examination performance was significant at the 5% or .05 level. This means that there were less than five chances in a hundred of obtaining this correlation of 0.05 by chance. If we read that a correlation was significant at the 1% or .01 level, it would mean that there was less than one chance in a hundred of obtaining this correlation by chance: Similarly, a 0.1% or .001 level means that there was less than one chance in a thousand of getting this relationship by chance.

SELF-ASSESSMENT EXERCISE

- i. How do you interpret the statement that “a correlation coefficient of 0.45 between verbal ability and typing examination performance was significant at the 0.05 level”?

4.0 CONCLUSION

You have learnt from this unit how to calculate and interpret the various measures of central tendency, spread, relationship and relative standing used in educational research, in furtherance of your goal of answering problems raised in your own research.

5.0 SUMMARY

The conclusion of research is to present findings that are meaningful and acceptable to the professionals in the academic community and beyond. This unit is part of the process of doing just that, i.e. using descriptive statistics to analyse the findings of studies which used data measured at the nominal and ordinal levels. The next unit, inferential analysis, will conclude the discussion.

6.0 TUTOR-MARKED ASSIGNMENT

- i. Discuss the principles underlying the concept of the normal distribution curve in psycho-educational measurement.
- ii. Describe the main types of correlation indices and their limitations.

7.0 REFERENCES/FURTHER READING

- Fisher, R.A. (1970). *Statistical Methods for Research Workers* 14th Ed. New York: Hafner Publishing Company Inc.
- Guilford, J.P. (1965). *Fundamental Statistic in Psychology and Education*. New York: McGraw-Hill Book Company.
- Thorndike R.L, Hagen E. (1977). *Measurement and Evaluation in Psychology and Education*, 4th ed. Wiley, New York.
- Stevens, S.S. (1975). *Psychophysics*. New York: Wiley.

UNIT 4 INFERENTIAL DATA ANALYSIS

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Inferential Data Analysis
 - 3.1.1 Definition
 - 3.1.2 Parametric Tests
 - 3.1.3 Non-parametric Tests
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Reading

1.0 INTRODUCTION

Following from the last unit which introduced descriptive statistical data analysis, this unit focuses on inferential data analysis which deals with drawing inferences about a population from sample data obtained from the same population. The two common methods used in inferential data analysis are parametric and non-parametric techniques which are adumbrated below.

2.0 OBJECTIVES

At the end of this unit, you should be able to

- distinguish between parametric and non-parametric tests
- calculate the t-test for the means of dependent and independent samples
- explain how the F-ratio is calculated and applied in ANOVA
- outline what the chi-square statistics is used for
- show how the Wilcoxon Matched Pairs Signed Tank test is ‘the non-parametric equivalent of the correlated t-test foe independent samples
- demonstrate how the statistic U in the Mann-Whitney U-test is computed and interpreted.

3.0 MAIN CONTENT

3.1 Inferential Data Analysis

3.1.1 Definition

Inferential statistics is the branch of statistics that deals with drawing an inference about a population on the basis of sample data from the population. Put another way, we use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one or one that might have happened by chance in the study.

There are two broad techniques used in inferential statistics: (1) Parametric techniques which assume that the data are normally distributed and are measured with interval or ratio scales. (2) Non-parametric techniques which are used when the data depart from the normal distribution and the measure can be nominal or ordinal.

The parametric tests commonly applied in psycho-educational research are the t-test and the analysis of variance. The non-parametric tests include the chi-square, the sign test, the Wilcoxon on matched pairs signed rank test, the Median test, the Maun Whitney U-test. Some of these will now be discussed.

SELF-ASSESSMENT EXERCISE

- i. Distinguish between parametric and non-parametric tests.

3.1.2 Parametric Tests

t-test

The purpose of t-test is to determine if a difference exists between the means of two sample groups. The samples may be independent or dependent. Data are regarded as dependent (or related) if they are obtained when each score or value in one set of data are paired with a score or value in another set. Use of the t-test assumes importantly a normal distribution of the scores in both samples.

For independent (or unrelated) samples, the t-test is a process for determining if there is a statistically significant difference between the means of two different groups of research subjects on a given dependent variable. A typical example of this use of t-test is where the first (experimental) group receives some treatment (e.g. using a new mathematics syllabus) while the second (control) group receives no

treatment (i.e. using the old existing mathematics syllabus). The formula for the calculation of the t-score in this case is given as:

$$t_{\text{obs}} = \frac{M_1 - M_2}{\sqrt{\left[\frac{(N_1 - 1)(S_1^2) + (N_2 - 1)(S_2^2)}{(N_1 + N_2) - 2} \right] \left[\frac{1}{N_1} + \frac{1}{N_2} \right]}} \quad \text{df} = (N_1 + N_2) - 2$$

For dependent (or related) samples, the t-test is also used in cases where two treatments are administered (at different times) to a single random sample, and the means of the two treatment samples are compared. In this case, all subjects contribute a score to both samples, and the calculation formula is given below:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n(n-1)}}$$

Where

- M_1 = mean for sample 1
- M_2 = mean for sample 2
- d = difference in any pair of scores
- $\sum d^2$ = difference in pairs of scores squared and summed
- $(\sum d)^2$ = difference in pairs of scores summed and total value squared
- n = number of pairs of scores

The Analysis of Variance

The main function of analysis of variance (ANOVA) is to compare systematically the mean response levels of two or more individual groups of observations, or set of observations, measured at two or more points in time.

ANOVA is typically used in a situation when there are one or more independent variables and two or more dependent variables. For example, to find out the effects of study habits, age, sex, and socio-economic background on college grades, one can do a simple ANOVA which analyses all four sets of data (study habits and grades, age and grades, sex and grades, socio-economic status and grades) at one time — which is an obvious advantage over the t-test statistic. With this technique, what are called Main effects (e.g. study habits and college grade), as well as Interaction effects (e.g. study habits with age and grades) are clearly partialled out.

The fundamental technique in analysis of variance is a partitioning of the total sum of squares (SS) into components related to the effects used in the model. Effect size measures are used in analysis of variance to describe the degree of relationship between a predictor or set of predictors and the dependent variable. Effect size estimates are reported in ANOVA to allow researchers to compare findings in studies and across disciplines.

When scores vary widely from the mean, the sums of the squares will be large. When they cluster around the mean, the sums of squares are small. The sum of squares is affected by the number of scores in a sample, more scores yield larger sums of squares. This is adjusted for by calculating and applying the degrees of freedom. By calculating the variability in data and producing an F-ratio, this ratio of the variance provides an estimate of the statistical difference between the mean scores. If the ratio of F is greater than the critical value of the F distribution at the chosen level of significance, then the null hypothesis is rejected and it can be concluded that there are statistically significant differences between at least some of the means.

The measure of differences among means is the “mean square between groups” (MS_B). Mean square between groups is in fact a summary of these differences, and is generally larger when group means are far from one another and smaller if the subgroup means are close together. Mean square within groups (MS_W) on the other hand, is the measure of differences among individual subjects within the groups. The greater the variation among individuals within the groups, the greater the mean square within groups will be.

The test statistic used here is the F-test or F-ratio (MS_B/MS_W) which can be found in tables of the F distribution. Where F exceeds the tabled critical value, MS_B is sufficiently larger than MS_W to conclude that group differences predominate, and H_0 is rejected. If on the other hand the critical F value is not exceeded, the null hypothesis of number difference is maintained.

There are several types of analysis of variance depending on the number of treatments and the way they are applied to the subjects in the experiment.

- One-way ANOVA is typically used to test for differences among at least three groups, since the two group case can be covered by a t-test. For example, in an experiment in which three Groups A, B & C are given respectively gin, wine and a placebo, and all groups are then tested with a memory test, a one-way ANOVA can be used to assess the effects of the various treatments.

- Two-way ANOVA is used when the subjects are subjected to repeated measures, in which the same subjects are used for each treatment. For example, in an experiment in which Group A is given gin and tested on a memory task, allowed to rest for a period of five days and the experiment is repeated with wine. The procedure is repeated using a placebo. A two-way ANOVA with repeated measures can be used to assess the effects of the gin versus the impact of the placebo. One flaw of this method is that it can be subjected to carryover effects.
- Factorial ANOVA is used when the experimenter wants to study the effects of two or more treatment variables. The most common of this type is the 2x2 (two by two) design, where there are two independent variables and each variable has two levels or distinct values.
- Multivariate analysis of variance (MANOVA) is used in studies that yield two or more interrelated response measures (i.e. multiple dependent variables).

The major assumption underlying analysis of variance is that the observations must be sampled and respond independently of one another. If dependencies arise because the same subjects or groups are measured repeatedly, then multivariate analysis (which is primarily concerned with the study of relationships between and within one, two or more sets of variables that are sequentially ordered with respect to time) should be employed. Other assumptions underlying analysis of variance are: (1) normal distribution of the population variance, (2) equal population variances, and (3) equal means.

A typical format, or summary table of how analysis of variance is usually presented in a research report, is as under:

Table 1: Typical Summary Table of ANOVA

Source of Variance	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F
Between-groups	114.96	2	57.48	7.82*
Within-groups	176.45	24	7.35	
Total	291.41	26		

* $P < .05$

(Source: Hassan, p.229)

3.1.3 Non-parametric (or distribution-free) Tests

Non-parametric techniques are used when the data depart from the normal distribution and the measures can be nominal or ordinal. Since non-parametric statistics do not require that the scores come from a

population with a normal distribution, they are also called distribution-free tests. Examples of this include Chi-Square, Sign test, Mann-Whitney U test, Median test, Wilcoxon Matched Pairs Signed Rank test, etc.

(1) The Chi-Square Test

Chi-Square is used to test for differences among sample frequency (nominal) data represented as frequency counts or percentages typically presented in a contingency table or cross-tabulation. The analyst collects frequency data indicating the number of subjects in each category, and arranges the data in a table format as below

Table 2: Contingency Table of Age and Opinion Responses to a Questionnaire

Age	Agree	Disagree	No Opinion	Row Total
Under 25	5	22	6	33
25-40	5	16	6	27
Over 40	30	5	5	40
Column Total	40	43	17	100

The X^2 analysis assesses the difference between the frequency expected (f_e), i.e. the theoretical relative frequency, and the actual frequencies observed (f_o). The larger the observed discrepancy is in comparison to the expected frequency, the larger the X^2 statistic and the more likely the observed differences are statistically significant. The expected frequency is computed by multiplying the row total by the column total for each cell, and dividing the product by the total number of cases in the table.

Applying the Chi-square formula which is

$$\sum \left[\frac{(f_o - f_e)^2}{f_e} \right]$$

and computing the degrees of freedom ($df=(r-1)(c-1)$), the X^2 value can be read off in any standard table of X^2 values at a chosen probability level. For example, from the above contingency table, the computed X^2 value of 36.0 is higher than the table value of 9.49 at the 0.5 level of significance, and one can therefore be certain with 95% probability that the differences among the groups is due to something other than chance. In other words, that the sample respondents do have different views on the subject matter.

It is necessary to state that although the X^2 does not require normality of population distribution, it does assume the following: (1) that the subjects of each group are randomly and independently selected; (2) that the groups are independent; (3) that each observation fits into only one category, and (4) the sample size must be fairly large such that no expected frequency is less than 5 for row or column.

SELF-ASSESSMENT EXERCISE

- i. Demonstrate the computation and application of the chi-square test. What are its assumptions?

Other non-parametric tests

In situations involving two or more related samples, or two or more independent samples, the following non-parametric tests, amongst others, are also often employed:

- (a) Wilcoxon Matched Pairs Signed Rank test:

This is in fact the non-parametric equivalent of the correlated t-test, and is used for testing differences between two or more related samples which have been measured at the interval level, or repeated measurements on a single sample. Its rationale is that researchers using this technique can determine whether one member of a matched pair of observations exceeds the other, and by how much. Under the null hypothesis, the probability is that the number and magnitude of differences favouring one group will be about the same as those favouring the other group.

The test is based on the magnitude of the difference between the pairs of observations. The distributional assumption in the correlated t-test that the differences follow a normal distribution is avoided in the Wilcoxon signed rank test because the test is based on the rank order of the differences rather than the actual value of the differences.

The procedure for calculating the Wilcoxon is to compute the difference between each pair, and rank the differences irrespective of the algebraic sign from the lowest to the largest, after which the signs are restored to the rank of the differences. The statistic T (different from the standard score T) is obtained by summing the smaller number of the like-sign ranks. If the calculated T-value for the normal 6 – 25 pairs in the Table is less than 14, then the null hypothesis is rejected. If the null hypothesis were true and there was no difference, then we would expect the

rank sum for positive and negative ranks to be the same. The smaller the T-value is, the more significant it is likely to be

(b) The Mann-Witney U-test

The Mann-Witney test is a non-parametric test for assessing whether two independent samples of observations measured at least at the ordinal level, come from the same distribution. The null hypothesis here is that the two samples are drawn from one single population, and therefore their probability distributions are equal.

The test involves the calculation of a statistic called U whose distribution under the null hypothesis is known. The value of U is computed, after the combined ranking, by concentrating on the lower ranked group and counting the number of ranks of the higher group which fall below the lower ranked group. To be significant at the given level of significance, the obtained U must be equal to or less than the table value for either one-tailed or two-tailed test. One simple way of computing the U is to use the formula

$$U = n_1 n_2 + n_1 \frac{(n_1 + 1)}{2} - R$$

Where R = the sum of ranks assigned to the group with sample size n_1

SELF-ASSESSMENT EXERCISE

- i. Show how the U statistic in the Mann-Whitney U Test is computed and interpreted.

4.0 CONCLUSION

We have now examined the major aspects of inferential data analysis in terms of the common parametric and non-parametric techniques used in drawing the required inferences about a population from the population sample. These will clearly aid you in your quest to test the various hypotheses or questions you may have stated for study.

5.0 SUMMARY

Research usually terminates in the analysis of data in which relationships or data that support or conflict our original hypotheses are subjected to tests of significance to help us to accept or reject, with some degree of confidence, the conclusions to be drawn. If this is successfully done, we can then rest assured that our research effort has been worth the while.

6.0 TUTOR-MARKED ASSIGNMENT

- i. How is the F-ratio in ANOVA calculated and applied in educational research studies?
- ii. Show how the Wilcoxon Matched Pairs Signed Rank test is the non-parametric equivalent of the correlated t-test for independent samples.

7.0 REFERENCES/FURTHER READING

Badmus, G.A. (1996). Treatment, Interpretation and Use of Educational Performance Scores from multiple tests. In G.A. Badmus and P. Odor (eds.), *Challenges of Managing Educational Assessment in Nigeria*. NABTEB, Benin City and Kaduna: Atman Publishers.

Cronbach L.J. (1970). *Essentials of Psychological Testing*. Harper and Row, New York.

Ebel R.L. (1979). *Essentials of Educational Measurement*, 3rd ed. Prentice-Hall, Englewood Cliffs, New Jersey.

Ferguson, G.A. (1981). *Statistical Analysis in Psychology and Education*. Singapore: McGraw-Hill International Publishing Co.

Guilford, J.P. (1965), *Fundamental Statistic in Psychology and Education*. New York: McGraw-Hill Book Company.

Okoh, N. (1980): *Bilingualism and Divergent Thinking among Nigerian and Welsh Schoolchildren*. *Journal of Social Psychology*, 110, Massachusetts, U.S.A

Thorndike R.L, Hagen E. (1986). *Measurement and Evaluation in Psychology and Education*, 4th ed. Wiley, New York.