RESEARCH METHODS IN

EDUCATION

A Handbook for Theory and Practice

MASINDE W. JOSEPH

First published in January, 2023

Kenpro Publishers, off Magadi Road
P.O. Box 1559 - 00503, Mbagathi - Nairobi, Kenya
Email: infokenpro@gmail.com
http://www.kenpro.org
Tel.: +254 725 788 400

©2023 Dr. Joseph Masinde Wangila
All rights are reserved. No part of this book may be reproduced, stored in retrieval systems or transmitted in electronic or any other form such as photocopying, without written permission from the author or publisher.

Printed by;

Kingdom Scribes Publishers Global
P.O. Box 46-0200
Nairobi, Kenya
Email: Kingdomscribesglobal@gmail.com
Phone: +254768978332

First Edition 2023
Printed in Kenya

All rights reserved worldwide.

Cover Design by Vanli Graphics
Emusioma Building, 2nd Floor,
P.O. Box 1283 – 50100
Kakamega, Kenya
Email: vanligraphics.ent@gmail.com
Tel.: +274 713 960 711

Research Methods in Education

by

Joseph Masinde Wangila

PROLOGUE

Pure felicity reins through my heart, following successful publication of the premier edition of this university level book, a culmination of many years of brainstorming, soul-searching, hard work, critical and creative thinking. The book is a precious tool for all university students at both undergraduate and postgraduate levels, together with their academic staff in the faculty of education and a few other disciplines by extension.

The success story behind this pioneer project has been the high demand for current knowledge on the methods and practices of conducting effective educational research, which is critical in solving a myriad of contemporary issues in the education sector in Kenya and beyond. University dons, students and other educational research enthusiasts are herein presented with easy-to-implement methods and practices that will guide them at every stage of the educational research process.

Being the first of its own kind, my soul would be uplifted if the book spares postgraduate students and their university supervisors the blushes of a failed research methodology. To this end, I look forward to getting positive feedback that would help yield finer versions of this book in the near future.

The book taps from classic scholarly works of the previous century, bringing in new postmodern 21st Century perspectives and insights on how selected methods and practices in educational research can be easily and conveniently implemented in the wake of advanced technology in the world today, the Kenya Vision 2030 and Kenya's Sustainable Development Goals (SDG's).

DR. JOSEPH MASINDE WANGILA - PhD, MSc, B. Ed (Sci.)
Murang'a University of Technology
© 2023

DEDICATION

In the form of this book is a special dedication to my beloved parents, retired Mwalimu Patrick Wangila and retired Madam Dorcas Nekesa, for their strong belief in the weaponry of very good education, which they selflessly accorded me, a legacy that will stand the test of time.

ACKNOWLEDGEMENT

Special thanks goes to the professors, doctors and research experts, who participated in one way or another towards review, critique and professional advisory about the contents of this book. I will forever be thankful to below mentioned;

Prof. Martin Wanjala	Masinde Muliro University of Science and Technology
Prof. Maurice Amadalo	Kaimosi Friends University
Prof. Stanley Mutsotso	Kibabii University
Prof. Catherine Aurah	Masinde Muliro University of Science and Technology
Prof. Robert Kati	Kibabii University
Dr. Peace Agufana	Murang'a University of Technology
Dr. Ronald Kikechi	Mount Kenya University
Dr. Francis Mwangi	Kenyatta University
Dr. Wesley Omwoyo	Maasai Mara University
Dr. Judith Ngome	Teachers Service Commission
Dr. Alphayo Ocholla	Kenya Institute of Curriculum Development

PREFACE

The book comprises nine chapters that all examine the major aspects of educational research. Chapter one defines educational research, outlines the characteristics of educational research, gives reasons for conducting educational research, outlines the educational research process, discusses the taxonomy of educational research, illustrates how to write an educational research report, describes how to state the educational research problem and finally discusses theoretical and conceptual frameworks for educational research.

Chapter two of this book gives the definition of literature review, highlights the importance of literature review, lists the various components of the literature review section in a research report, outlines the steps to follow while reviewing literature, discusses gaps of a literature review, how and where to identify them. All these are important aspects that must be mastered by any researcher, which makes the topic worth your careful readership.

Chapter three defines a research design, examines the structure and style of the research design section, enumerates the characteristics of a good research design, discusses the characteristics and types of research designs, highlights on the validity of research designs, and examines threats to validity of research designs and how to counter them in educational research.

The fourth chapter talks about target population in detail, definition of sampling, reasons for sampling, commonly used sampling techniques in educational research, choosing the right sampling method, the sampling frame, sample space, determination of sample size in educational research and finally the ethical issues in sampling and research in education.

Chapter five gives the definition of data collection, highlights primary and secondary data sources, examines the importance of data collection then discusses the tools and instruments for data collection in educational research.

Chapter six is defines validity of a research instrument, types of validity of instruments, assessment of validity, methods of assessing reliability of research instruments and factors that affect reliability of educational research instruments.

Chapter seven presents the definition of data analysis, reasons for data analysis, and steps in data analysis process, types of data and the various ambiguities in classifying data variables, types of data analysis and how to report findings in scholarly manner.

Chapter eight talks about the definition of data presentation, methods of data presentation, importance of data presentation, scholarly writing of educational research findings, structure and style of the results section, interpretation and discussion of research findings.

Chapter nine discusses the rationale for suggestions for further research, recommendations in educational research and how to write both sections in a dissertation or research paper. As you read through the nine chapters of this book, I strongly urge you to share the book's insight with your colleagues in all academia, for all important goal of ameliorating the quality of education in Kenya and beyond. Enjoy your readership.

DR. JOSEPH MASINDE WANGILA - PhD, MSc, B. Ed (Sci.)

TABLE OF CONTENTS

Prologue iv

Dedication v

Acknowledgement vi

Preface vii

List of Tables xii

List of Figures xiii

CHAPTER 1: ..... - 1 -

INTRODUCTION TO EDUCATIONAL RESEARCH ..... **- 1 -**

Definition of Educational Research ..... - 1 -

Characteristics of Educational Research ..... - 1 -

Reasons for Conducting Educational Research ..... - 2 -

The Educational Research Process ..... - 3 -

Taxonomy of Educational Research ..... - 8 -

Writing a Research Proposal/Report in Education ..... - 13 -

Statement of the Problem in Educational Research 18

Conceptual Framework for Educational Research 21

Theoretical Framework for Educational Research 26

CHAPTER 2: **29**

LITERATURE REVIEW IN EDUCATIONAL RESEARCH **29**

Definition of Literature Review 29

Importance of Literature Review in Educational Research 29

Components of a Good Literature Review Section 30

Key Steps in Literature Review 30

Gaps in Literature Review 34

Difference between Gap and Problem in Educational Research 35
Identifying Gaps in Existing Literature 35
Where to Find Research Gaps in Scholarly Work 37
CHAPTER 3: 38
RESEARCH DESIGNS IN EDUCATIONAL RESEARCH 38
Definition of a Research Design 38
Structure and Style of the Research Design Section 38
Characteristics of a Good Research Design 39
Classification of Research Designs 39
Validity of Research Designs 54
CHAPTER 4: 61
SAMPLE SIZE AND SAMPLING IN EDUCATIONAL RESEARCH 61
Target Population 61
What is Sampling? 62
Reasons for Sampling in Educational Research 62
Commonly Used Sampling Methods in Educational Research 63
The Research Sampling Frame 73
Sample Space 74
Determination of Sample Size in Educational Research 75
Ethical Issues in Sampling and Research in Education 79
CHAPTER 5: 81
DATA COLLECTION IN EDUCATIONAL REEARCH 81
Definition of Data Collection 81
Primary Data 81
Secondary Data 83

Importance of Data Collection 84
Methods and Instruments for Data Collection 85
CHAPTER 6: 102
VALIDITY AND RELIABILITY OF EDUCATIONAL RESEARCH INSTRUMENTS 102
What is Validity? 102
Types of Validity of Research Instruments 103
Assessment of Validity of Research Instruments 105
Reliability of Educational Research Instruments 106
Methods of Assessing Reliability of Research Instruments 106
Factors that Affect Reliability of an Instrument 115
Trustworthiness of Qualitative Research Instruments 116
CHAPTER 7: 118
DATA ANALYSIS IN EDUCATIONAL RESEARCH 118
Definition of Data Analysis 118
Reasons for Data Analysis 118
Preparing Data for Analysis 118
Types of Data 119
Types of Data Analyses 122
Qualitative Data Analysis Methods 123
Triangulation in Educational Research 129
Quantitative Data Analysis Methods 129
Inferential Statistics and Hypothesis Testing 140
Parametric Versus Non-Parametric Inferential Tests 141

CHAPTER 8: 170

PRESENTATION AND DISCUSSION OF EDUCATIONAL RESEARCH FINDINGS 170

Definition of Data Presentation 170

Importance of Data Presentation 170

Methods of Representing Data 171

Interpretation of Research Findings 178

Scholarly Writing of Educational Research Findings 183

Discussion of Findings 187

CHAPTER 9: 191

RECOMMENDATIONS AND SUGGESTIONS FOR FURTHER RESEARCH 191

Research Recommendations 191

How to Write Recommendations in a Research Report 193

Suggestions for Further Research 194

BIBLIOGRAPHY 198

APPENDICES 201

LIST OF TABLES

Table 1: Differences between Basic and Applied Research ..... - 10 -
Table 2: Differences between Qualitative and Quantitative Research ..... - 12 -
Table 3: Description of the Various Types of Research Gaps 35
Table 4: Difference between Quantitative and Qualitative Research Designs 43
Table 5: Primary and Secondary Design Dimensions in Mixed Methods Research 52
Table 6: Steps of Hypothesis Testing 140
Table 7: Sample T-Test Results 144
Table 8: Sample Contingency Table 149
Table 9: Sample Expected Frequencies Table 149
Table 10: Sample Chi-statistics Table 150
Table 11: Sample Model Summary Table for Regression Analysis 157
Table 12: Sample ANOVA Table for Regression Analysis 158
Table 13: Sample Coefficients Table for Regression Analysis 159
Table 14: Example of a Statistical Table in APA Format 177

LIST OF FIGURES

Figure 1: Summary of the Educational Research Process ..... - 7 -
Figure 2: Dummy of a Conceptual Framework 24
Figure 3: Dummy of a Conceptual Framework 25
Figure 4: Example of a Conceptual Framework 25
Figure 5: The Literature Review Process 33
Figure 6: Illustration of the Counter-Balanced Research Design with 4 groups 47
Figure 7: Fair dice – Used for Selecting one from Six Individuals/Groups 65
Figure 8: Fair coins - Used to Select one from Two Individuals/Groups 65
Figure 9: Solitaire Cards used to Select Several from about Fifty 65
Figure 10: Summary of Sampling Methods 72
Figure 11: Approaches to Qualitative Analysis 131
Figure 12: Shape of a Normal Distribution 137
Figure 13: Shape of a Positively Skewed Distribution 137
Figure 14: Shape of a Negatively Skewed Distribution 138
Figure 15: Shape of Mesokurtic, Leptokurtic and Platykurtic Distributions 139
Figure 16: Shape of a Bimodal Distribution 139
Figure 17: Example of a Pictogram 172
Figure 18: Example of a Column Chart 173
Figure 19: Example of a Pie-Chart 174
Figure 20: Example of a Bar Graph 175
Figure 21: Example of a Histogram 176

ABBEVIATIONS AND ACRONYMS

ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
APA	American Psychological Association
CBC	Competency-Based Curriculum
CEMASTEA	Centre for Mathematics, Science and Technology in Africa
DAG	Document Analysis Guide
FGD	Focus Group Discussion
GT	Grounded Theory
HSD	Honestly Significant Difference
ICT	Information and Communication Technology
IPA	Interpretive Phenomenological Analysis
IQ	Intelligence Quotient
IRC	Ideal, Reality, Consequences
KCPE	Kenya Certificate of Primary Education
KCSE	Kenya Certificate of Secondary Education
KESI	Kenya Educational Staff Institute
KICD	Kenya Institute of Curriculum Development
KNEC	Kenya National Examinations Council
PPMCC	Pearson's Product Moment Correlation Coefficient
SD	Standard Deviation
Sig.	Significance
SMART	Specific, Measurable, Achievable and Time-bound
SPSS	Statistical Package for Social Sciences
STEM	Science, Technology, Engineering and Mathematics

CHAPTER 1:

INTRODUCTION TO EDUCATIONAL RESEARCH

This chapter gives a precise definition of educational research, outlines the characteristics of educational research, gives the reasons for conducting educational research, examines the educational research process, discusses the taxonomy of educational research, illustrates how to write a research proposal/report in education, outlines how to state the educational research problem and finally discusses theoretical and conceptual frameworks for educational research.

Definition of Educational Research

The term "*research*" is one of the commonest among scholars in any field of study. It is the process of systematic investigation that entails collection, analysis and interpretation of data in accordance with methodologies set by specific professional fields in order to describe, explain, predict and control a phenomenon.

Educational Research refers to the systematic collection and analysis of data in an educational context, setting or field with the aim of solving issues concerning daily teaching and learning. Examples of fields in education may include but not limited to; educational foundations, educational psychology, educational technology, technology education, curriculum studies, early childhood education, special education and several others.

Characteristics of Educational Research

Education research examines various aspects of instruction, human behaviour, interactions, organizations and institutions that shape learning outcomes etc. The following are characteristics of educational research that makes it unique from other research disciplines;

i. It involves collection of data either from primary sources or secondary sources in learning institutions
ii. It is either empirical or observation-based
iii. It demands accurate and precise observations

iv. It emphasizes on developing generalizations, theories, principles etc. which may facilitate understanding as well as prediction of learners' behaviour
v. It requires expertise in designing the research, constructing data collection instruments, collecting data, analyzing the data and interpreting the results
vi. It tends to find an objective, unbiased solution to educational problems and makes efforts to validate the procedures involved
vii. It is well directed, deliberate, and unhurried activity, often involving refinement of the problems or research questions
viii. It is conducted with well-defined procedures and in an organized manner.
ix. Findings are reported in organized, pre-defined format

Reasons for Conducting Educational Research

The primary purpose of educational research is to expand the existing body of knowledge in the various sub-disciplines in education by providing solutions to different problems in pedagogy while improving teaching and learning practices. Educational researchers also seek answers to pertinent questions surrounding learner motivation, attitudes, academic performance and classroom management. The following is a summary of the main reasons why educational research is done;

i. **Generation of knowledge** - educational research helps us verify existing educational knowledge and generate new theories. For example, various theories of learning explain how human beings learn, Behaviourism theory of learning explains how learning habits can be formed, information-processing theory explains how leaners process various types of sensation and how they make meaning of them, together with the role memory. Such research leads to generation of knowledge and must therefore be planned and executed in a carefully controlled manner.

ii. **Problem-solving -** educational research attempts to solve existing problems related to education. These problems may be faced by students, teachers, school administrators, parents or any other stakeholders in the educational sector. Educational researchers are frequently confronted with a variety of issues e.g. what is the best method of teaching

abstract concepts? How can children learn scientific skills more efficiently? How can scientific creativity be fostered among children? Just to mention a few. Answers to such critical questions are realistic ways of promoting learning and teaching. In each of the problems, the researcher's main objective is to come up with a practical solution, so that the same can be applied by policy makers to solve the educational problems.

iii. **Course of action in specific situations -** educational research also demands not only finding out solutions of the problem but also expects that solutions be applied immediately in order to avert a crisis. By applying the solution, researchers should bear in mind the context specificity of the solution, because the nature of the problem(s) may be the same but solutions may not be applicable in other situations. E.g. online teaching may solve the issue of insufficient classrooms in universities, but not in primary or secondary schools in Kenya.

iv. **Training for future researchers -** another main purpose of educational research is to provide knowledge and skills to build the capabilities of young researchers. As we already know, educational research is an activity that is not only scientific but also requires an inquisitive mind, critical thinking ability and most importantly carried out in a systematic and organized manner. Prospective researchers must therefore be well-trained.

The Educational Research Process

Research being a systematic and organized process involves a number of steps. Research in the field of education generally follows the following 8 major steps;

1. **Identifying the problem -** the first major step in the process of educational research is to identify and formulate the problem. The research problem may be an issue that is currently facing the education sector in one way or another, which the researcher wants to investigate e.g. consistent poor performance by students in a given subject and the reasons are unknown, negative attitude to Mathematics among secondary school students, increasing cases of school dropout outs among others. The problem stated must however be in the researcher's area of interest or specialization in the case of postgraduate students.

2. **Formulating objectives -** research objectives are the goals that are set to guide the whole research. The researcher must clearly specify what he or she wants to research when formulating objectives. A properly stated objective informs readers about the scope and extent of the research. Each objective should be accompanied by a research question or null hypothesis or both, depending on the complexity and type of research. A good research objective should have two measurable variables; one independent and one dependent, starting with the former.

   A hypothesis on the other hand is a statement of the researcher's educated guess on how the independent and dependent variables are related. In educational research, all hypotheses must be stated in null form i.e. assuming no association/relationship/effect/influence between the independent and dependent variables. Formulating the hypothesis helps one to decide key research methodologies like the research design, data collection instruments and data analysis techniques.

   A research question on the other hand is the question that the study seeks to answer at the end of the study. The question is derived from the research objective, in a manner such that the answer in not a simple yes or no.

   For example, an educational researcher seeks to find out if online teaching improves students' achievement in Mathematics. The following is an example of a research objective and its corresponding hypothesis and research question:

Objective:	To determine the effect of online teaching on students' achievement in Mathematics
Hypothesis:	Online teaching has no statistically significant effect on students' achievement in Mathematics
Research question:	How does online teaching affect students' achievement in Mathematics?

3. **Literature review -** this is the systematic reading of what other researchers have done previously in relation to your research problem. Most likely, you won't be the first to investigate that problem and neither will you be the last. Literature review therefore helps you to know what has been done so far and what has been found out. Without extensive literature review, you might find yourself investigating a problem that was solved long time ago! For organized and effective literature review, one must be guided by the research objectives. This explains why literature review follows formulation of objectives in the research process and not the other way round.

4. **Designing the study -** this is another very important step in the research process, as it involves deciding the research methodology. At this stage, the researcher has to carefully decide which methods are best suited to execute their study and successfully achieve the research objectives. The methods include; research design, target population, sample size, sampling techniques, data collection instruments, piloting, data analysis procedures and ethical considerations.

   This decision is best arrived at by consulting the research objectives and hypothesis e.g. since the study will investigate the factors leading to student unrest in schools within County A, it must be a qualitative research as it is only interested in getting explanations from students about the issues that caused unrest, interviews will be used to collect data to give room for elaborations if any. It must be done in schools in County A because that is where we are likely to get respondents with a clear picture of what happened or who actively participated in the unrest activities.

5. **Piloting the study -** this involves trying out the data collection instruments in order to establish if they are ideal. A good research instrument should be valid and reliable and therefore both of these measures must be assessed at the piloting stage and any issues emanating from the pilot study addressed before proceeding to actual field work.

6. **Collecting the data -** the type of instrument used for data collection determines the type of data to be collected and by extension the appropriate method of collecting it. In this step, the researcher will personally conduct interviews, administer achievement tests, moderate focus group discussions and so on. The researcher may also seek the help of research assistants if

the data needed is voluminous. The research assistants must be trained to ensure data collection is objective and reliable.

7. **Manipulating the data** - this step entails analyzing and interpreting the raw data as captured by research instruments. The researcher must sort, arrange and organize the data, then make relevant statistical computations. The data analysis outputs may be in a form that is not easily deciphered by a layman e.g. box plots, Q-Q plots, graphs, tables, charts etc. The researcher at this stage must therefore interpret the data into information that makes sense to everyone. Strict adherence to research objectives is of the essence here, so as to give comprehensive answers to the research questions.

8. **Dissemination of findings -** after data analysis and interpretation, the researcher should discuss it and make a conclusion basing on his or her research objectives. Research findings should then be shared widely, so that the scholarly world, policy makers and other stakeholders utilize the 'good news' in solving the problem that led to the research in the first place. This can be done through journal publications, conference presentations or writing a thesis. All these are examples of research reports, as they include all the information related to the research, answering the what was done, why was it done, how was it done, and what was found out kind of questions.

The following figure gives a summary of the steps in the educational research process:

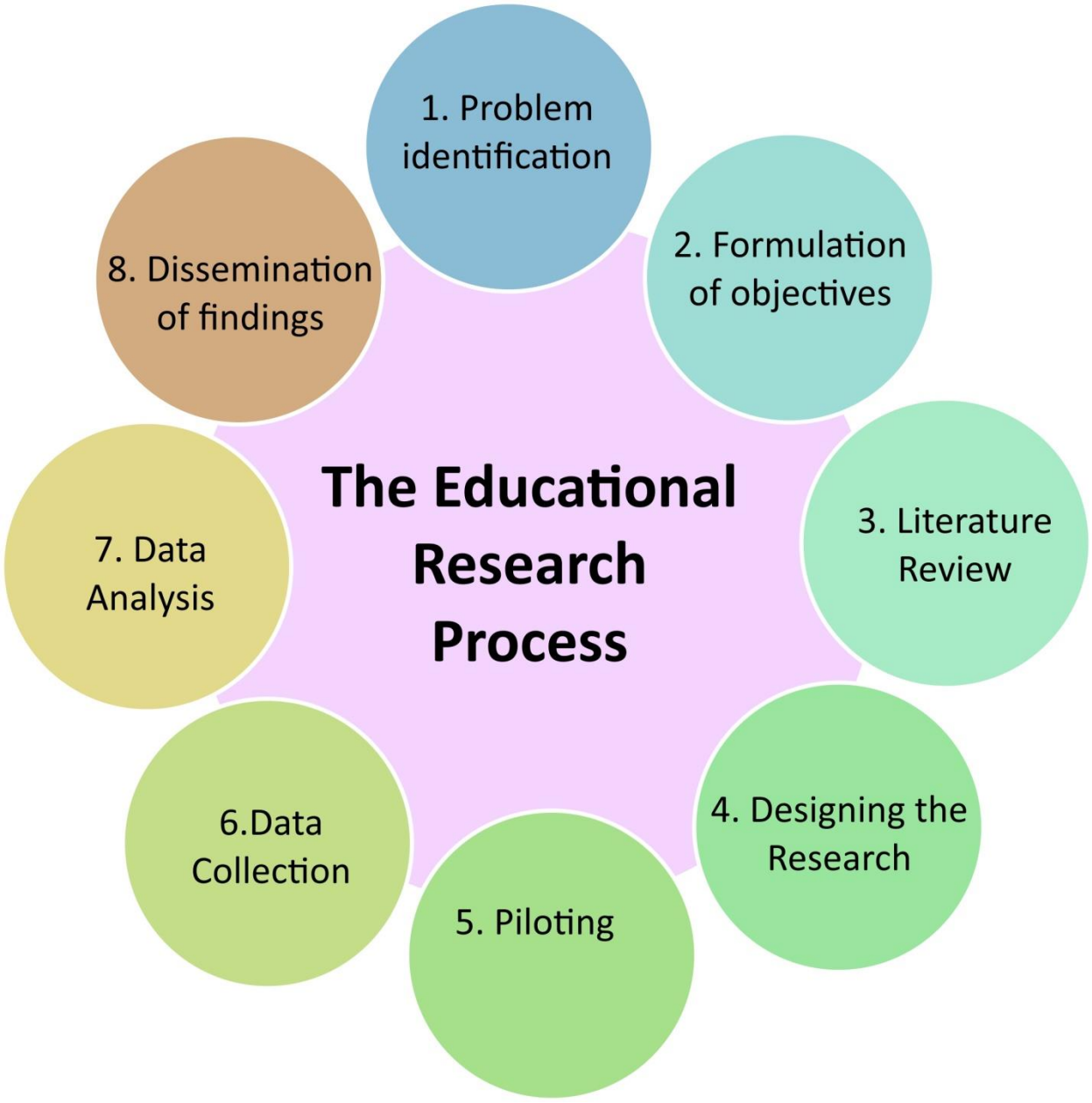


Figure 1: Summary of the Educational Research Process

Taxonomy of Educational Research

Educational research per ce is broad and can be classified in a number of ways. It is important for novice researchers to be conversant of the different taxonomic units of educational research, as this has implications on the methodology that are used to execute it. The various types of educational research can be placed into several categories depending on *purpose* of the research or *type of data* involved. The next two sections examine each of those taxonomies in detail.

Purpose-Based Classification of Educational Research

Here we will discuss the first three types of research basing on their purpose or goals. As earlier mentioned, educational research is all about generation of new knowledge and using that knowledge to ameliorate academic standards. To this end, educational research falls into three categories: (i) Fundamental/Basic Research (ii) Applied Research and (iii) Action Research. Let us discuss each one of them in detail;

i. Basic Research

Basic research is also called *fundamental research*. It is generally carried out for purposes of generation of knowledge and has no immediate practical application of its findings. It always aims at generation of new knowledge in form of theories, principles, generalizations and laws about natural phenomena in academic institutions. These researches are normally done in controlled situations like a laboratory or laboratory-like conditions. Basic research is done to generate information that applied researchers may use to invent or design various machines, software, equipment or devices that improve academic standards.

Basic or fundamental research always builds on, modifies or corrects the existing knowledge. Since educational research has immediate utility and education being inter- and multi-disciplinary area of study, researchers therein are more applied in nature. In the same vein, studies in other fields like social sciences, psychology etc. are closely related in many respects.

Basic research is also carried out to understand different concepts in these allied areas. Examples include; learning theories that explain how learning takes place in human beings, basic human

intelligence and how it functions, understanding the personality of human beings and categories of people based on their personality. All these theories from the allied areas of study help us to understand the social behavior of human beings. Scientists doing basic research normally take different perspectives/angles/routes to look at the main issue.

ii. Applied Research

The second type of research basing on its purpose is applied research. As previously mentioned, applied research is done by scientists or researchers to solve the immediate or current problems. Research that is currently being conducted by scientists all over the world to develop vaccines for controlling the novel coronavirus is a good example of applied research. Applied research is done with the aim of solving a practical problem, like the many covid-19 related deaths ravaging the entire world since the year 2019.

The major difference between basic and applied research is that in applied research, problems are solved in natural settings rather than in laboratory conditions. In applied research, researchers or scientists use the existing laws, theories, and principles of basic research to solve their immediate or practical problems in real-life situations. Take for example, using the intelligence theories, one can identify children with various types of intelligence (numeracy, verbal, reasoning, kinesthetic, spatial etc.) and develop educational programs based on their type of intelligence for their educational enrichment.

Applied researches are in a way used to test and verify existing theories, laws or principles generated through basic research. Thus, applied research is generally carried out to improve a product or process. In this regard, applied research helps basic researchers to nullify/review their theories or formulate new ones. Applied researchers follow the same steps as those for conducting basic research.

As already stated, educational research is more of applied in nature, as it tries to answer questions on what, how, and why. Practitioners of educational research like teachers, educational administrators and education officers carry out applied research for the dual purpose of addressing professional problems and improving academic standards in learning institutions.

The following table outlines the major differences between basic and applied research:

Table 1: Differences between Basic and Applied Research

Basic Research	Applied Research
Aims at generating new knowledge	Aims at solving real-life problems
Is purely theoretical	Has practical aspects
Develops theories, laws, principles etc.	Uses existing theories in its methodology
Done even when there's no problem	There must be a problem
Is exploratory by nature	Is explanatory

iii. Action Research

Action research is diagnostic in nature, as it intends to improve educational practices of teachers through a systematic inquiry. In essence, action research is associated with hands-on solutions to problems affecting teachers, students, school environments, education curriculum, assessment etc. This type of research empowers teachers to introspect about their teaching methods, teaching styles and activities that are normally executed on a day-to-day basis in relation to the standard practice. It also provides an opportunity for teachers to learn and grow in their profession. The main difference between action research and other types of research is that solutions in basic research are highly localized, contextual and specific to the problem hence cannot be generalized to other circumstances. Action research usually combines both theory and practice in real educational settings.

Data-Based Classification of Educational Research

Researchers in education normally want to address three basic questions related to their research; (i) how can the educational issue be addressed? (ii) what exactly happens when the issue arises? (iii) what caused the issue? Based on the nature of educational research problems and the above-raised questions, the research can be classified as either qualitative or quantitative. Thus, educational research can also be placed into these two categories depending on the type of data involved.

i. Quantitative Research

Educational research can be classified as quantitative if the variable under investigation can be measured precisely with the help of structured instruments that yield continuous data i.e. data in numerical form. Such data can be analyzed statistically. For example, if we want to measure the achievement of students in any school subject like English, Kiswahili and Mathematics, through structured research instruments like standardized achievement tests. Resultant data will be students' scores in form of numbers and may therefore allow for statistical computations like mean, median, standard deviation and more complex inferential statistics. This qualifies it to be classified under quantitative research.

ii. Qualitative Research

Qualitative research involves collection of discrete data i.e. data in form of categories, which cannot be expressed in number form. Qualitative research is usually an in-depth understanding of human behaviour and the reasons that govern leaners' behavior. Unlike quantitative research, qualitative research focuses on the reasons behind various aspects of behavior. It simply investigates the why and how of decision-making, as opposed to the what, where, and when of quantitative research. Qualitative research is therefore done on smaller, purposively selected samples rather than large random samples. Resultant data is organized into patterns, trends and themes.

iii. Mixed Methods Research

It is increasingly becoming clear among educational researchers that using only qualitative or quantitative data to back up your findings leaves many unanswered questions. For this reason, it is imperative to blend the two. Mixed methods research is one that involves collection of both qualitative and quantitative data at one or more stages of the whole process. By conducting mixed methods research, the questions that arise from one phase of data collection are answered in the other phase.

The following table gives the major differences between qualitative and quantitative research basing on 5 criteria.

Table 2: Differences between Qualitative and Quantitative Research

Criteria	Quantitative	Qualitative
1. Purpose	It is interested in studying cause-effect relationships	It studies a phenomenon in detail as it is
2. Design	Has a fixed research design which is developed before the study is done	If flexible and may evolve as the study progresses
3. Approach	Adopts a deductive approach, tests theories	Adopts an inductive approach, may generate theories
4. Sample	Uses large sample sizes, which must be representative of population	Uses smaller sample sizes, which need not be representative of population
5. Type of data and data analysis	It involves continuous data, which is analyzed statistically using numbers	Involves categorical data, which is analyzed thematically using words

Writing a Research Proposal/Report in Education

A *research proposal* is an official document that details everything that a researcher to do in their impending study i.e. the what, how, when. A Research report on the other hand is written after the study has been conducted. It has all sections that were in the proposal, with addition of the findings. Whether it is for purposes of fulfilling academic requirements of a university degree or for addressing a real-life problem, a proposal /thesis for educational research follows a format that is prescribed by the body or institutions to which the research is affiliated to e.g. the university, research institution or research journal. Nevertheless, a few sections are common in all educational research disciplines and will be outlined in this section. The following format is hereby recommended;

1.0 CHAPTER 1: INTRODUCTION
1.1 Overview of chapter one
1.2 Background of the study
1.3 Statement of the problem
1.4 Purpose of the study
1.4.1 Objectives
1.4.2 Hypotheses/Research questions
1.5 Assumptions of the study
1.6 Scope of the study
1.7 Limitations & Delimitations of the study
1.8 Significance/justification of the study
1.9 Conceptual/Theoretical Framework
2.0 CHAPTER 2: LITERATURE REVIEW
2.1 Overview of Chapter Two
2.2 Literature on first objective
2.3 Literature on second objective
2.4 Literature on third objective
2.5 Summary of Literature Reviewed
2.6 Gaps in Literature Reviewed
3.0 CHAPTER 3: RESEARCH METHODOLOGY
3.1 Overview of chapter three
3.2 Research Design
3.3 Research Location
3.4 Target Population
3.5 Sample Size and Sampling Procedures
3.6 Research Instruments
3.7 Piloting
3.7.1 Validity of Instruments
3.7.2 Reliability of Instruments
3.8 Data Collection Procedures
3.9 Data Analysis Procedures
3.10 Ethical Considerations
4.0 CHAPTER 4: THE RESEARCH FINDINGS
4.1 Overview of Chapter Four
4.2 Demographic Information about Respondents
4.3 Preliminary Report
4.5 Results of Objective One (Presentation, Interpretation, Discussion)
4.6 Results of Objective Two (Presentation, Interpretation, Discussion)
4.7 Results of Objective Three (Presentation, Interpretation, Discussion)
5.0 CHAPTER 5: SUMMARY, CONCLUSION & RECOMMENDATIONS
5.1 Overview of Chapter Five
5.2 Summary of the Study
5.3 Summary of the Findings
5.4 Conclusion (from the Findings)
5.5 Recommendations (for policy action)
5.6 Suggestions (for further research)

REFERENCES

APPENDICES

Note that there might be slight variations from one university to another regarding the order in which some subsections of each chapter appear. However, the content and style of writing each section remains the same regardless of the institution, faculty or department. For purposes of staying true to the philosophy behind the writing of this book, only the major sections of a thesis will be discussed in the remaining sections of this chapter.

Background of the Study

The research background is one of the most important components of a research proposal or dissertation. Quality of background section determines whether the reader will be interested in the rest of the document or not. Careful attention should therefore be given while writing the background section to make it appealing to the reader. This will ensure that the audience reads the entire research paper or thesis. A well-written background will also provide your study with a context and prompt the readers to read the rest of your document.

Unfortunately, most researchers struggle to write a good background to their study. Another common problem that inexperienced researchers and postgraduate students encounter is distinguishing between the background and literature review, both of which are distinct and critical aspects of any research paper or thesis. The two terms cannot be used interchangeably, as they serve clearly defined roles that are of equal importance. This section highlights the basics of writing the background section of a research proposal, thesis or report and explains how it differs from literature review section.

Components of the Background Section and how to Write it

The background of a study is the first section of the research paper, proposal or thesis and establishes the context underlying the research. Background of the study is important because it provides the reader with a better understanding of what led to the research and epitomizes the problem statement. The background forms the crux of the study, because it introduces an unaware audience to the research and its importance in a clear and logical manner. At times, the background may even explore whether the study builds on or refutes findings from previous studies.

Any relevant information that readers need to know before delving into the research nitty-gritty should be made available to them in the background section. The length and detail presented in

the background varies depending on the complexity and novelty of the research topic. At times, a simple background suffices, even if the study is complex. Before writing and adding details in the background, take note of these additional points;

- ✧ Start with a strong beginning - begin the background by defining the research topic and then identify the target audience.
- ✧ Cover key components – clearly explain all theories, concepts, key terms, and ideas that may be unfamiliar to the target audience thoroughly.
- ✧ Take note of important prerequisites - go through the relevant literature in detail, taking notes while reading and citing the sources.
- ✧ Maintain a balance - make sure that the background is focused on important details, but also appeals to a broader audience. For instance, a topic on educational psychology can be written in a matter that also attracts readership from science educators.
- ✧ Include historical data - current issues largely originate from historical events or findings. If the research borrows information from a historical context, add relevant data in the background, clearly illustrating how the problem has evolved over the years. Such statistics makes the problem concrete.
- ✧ Explain novelty - if the research topic is unique or novel, provide an explanation that helps to understand the research better e.g. covid-19 is just one of the many types of corona-viruses that have been in existence for ages.
- ✧ Increase engagement - to make the background engaging, build a story around the central theme of the research e.g. “online teaching is a relatively new approach to instruction. However, unlike other technology-based methods of instruction….” This makes the problem unique.

The following are common mistakes that researchers make while writing their background sections. Suggestions on how to avoid them are also presented;

- Ambiguity - don't be ambiguous. While writing the background section of your research proposal, assume that the reader does not understand any intricate details about your research topic. Writing in a way that does not convey the message to the readers defeats the rationale of the background section in a proposal, thesis or paper. You should therefore express yourself keeping in mind that the reader might be having no clue about your research or problem.
- Unrelated themes - steer clear from topics that are not related to the variables of your research. Instead, center your writing around the pivotal aspects of your research topic by highlighting the gaps in the literature, stating the novelty of the study and need to conduct the study.
- Poor organization - do not place information without a structure. Make sure that the background content is presented in a chronological manner that flows well. Not discussing the themes in a chronological manner, in paragraphs that flow well, can confuse the reader about the progress in the field, so try and organize your paragraphs carefully to urge the audience to reader further.
- Too long - don't write a background that is too long or too short. Instead, focus on including all the important details but write concisely, in three or four pages.

As earlier mentioned, background is an indispensable part of a research proposal, thesis or paper. This is because when you are doing research, you are focusing on a single issue or phenomenon. However, the phenomenon occurs within a context. The dynamics of the context includes the history of the problem. To put it another way, research mostly works on gaps in human knowledge, something that is missing, hence the word 'search' in research. To identify the gap, one needs to know what borders on the gap, to inform you what is missing.

In the background section, the author usually outlines the historical developments in the literature that led to the current topic of research concisely. If the study is interdisciplinary, it should describe how different disciplines are connected and what aspects of each discipline will be studied. Additionally, authors should briefly highlight the main developments of their research topic and

identify the main gaps that need to be addressed. In other words, this section should give an overview of your study. The section should be structured as follows;

- ✓ What is known about the broad topic?
- ✓ What are the gaps or missing links that need to be addressed?
- ✓ What is the significance of addressing those gaps?
- ✓ What are the rationale and hypothesis of your study?

The background section should therefore provide general information about the topic of your research and emphasize the main aims. Ensure that you only discuss the main and relevant aspects of the studies that have led to your aims. Do not elaborate on them as this should be done in the literature review section.

The background section should discuss your insights in a chronological manner to accentuate the progress in the field and the missing points that need to be addressed. The background should be written as a summary of your interpretation of previous research and what your study proposes to accomplish.

As the background includes a lot of information, it can become a long drag, causing the readers to lose interest. To ensure that your background is engaging, you should try to build a story around the central theme or variables of your research. Ensure that the story adheres to the core idea and does not digress into a broad literature review. Each idea should lead to the next, so that readers are able to grasp the story by themselves and identify the gaps that your study is going to address.

Difference between Background and Literature Review Sections

Many authors find it difficult to fathom the difference between literature review and the study background. The literature review section should follow the background section, as the second section of your manuscript or thesis. This section basically supports the background section by providing evidence for the proposed hypothesis. This section should be more comprehensive and thoroughly describe all the studies that have been mentioned in the background section. It should

also elaborate on all studies that form evidence for the present study and discuss the common trends.

To write this section, you will need to do a thorough literature search on different studies that relate to the broad topic of your research. This will introduce to the readers the area of your research. Following this, you should present a more focused cross-section of the specific studies associated with the specific objectives of your study. It would be ideal to organize them thematically and discuss them chronologically so that readers are aware of the evolution and progress in the field. In other words, separate themes should be discussed chronologically, to highlight how research in those fields has progressed over time. This will highlight what has been done and future directions that need to be taken.

Statement of the Problem in Educational Research

As earlier mentioned, research is a systematic investigative process employed to increase or revise current knowledge by discovering new facts. It can be divided into two general categories; (i) basic research, which is inquiry aimed at increasing scientific knowledge, and (ii) applied research, which is effort aimed at using basic research for solving problems or developing new processes, products, or techniques. One of the most important steps in any research is to identify and delineate the research problem i.e. what the researcher wants to address and what questions he/she wishes to answer.

A research problem in education may be defined as an issue of concern or a gap in the existing knowledge, or a deviation in the norm or standard that points to the need for further understanding and investigation. Although many problems turn out to have several solutions, difficulties arise where such means are neither obvious nor immediately available. This then necessitates some research to reach a viable solution.

Importance of Stating the Research Problem

The ultimate goal of stating the problem is to transform a generalized problem (something that bothers you; a perceived lack) into a targeted, well-defined problem; one that can be resolved through systematic investigation and careful decision-making.

Writing a statement of the problem should help you clearly identify the purpose of the proposed project. Often, statement of the problem will also serve as the basis for the introductory section of your final proposal, directing your reader's attention quickly to the issues that your proposed project will address, thereby providing the reader with a concise statement of the proposed project itself. A good statement of problem needs not be long and elaborate, as most of the material may well be already in the background section. One page or less is therefore enough for a good statement of problem, as long as it has all the important parts that are discussed in the next section.

Features of a Good Statement of the Problem Section

A well-drafted problem has the following 7 important qualities;

- ✓ It should address a gap in educational knowledge.
- ✓ It should be significant enough to contribute to existing body of knowledge
- ✓ It should lead to further research
- ✓ The problem should render itself to investigation through collection of data
- ✓ It should be of interest to the researcher and suit his/her skills, time, resources and area of specialization.
- ✓ The approach towards solving the problem should be ethical.
- ✓ Above all, *it should raise critical questions which cannot be answered unless your research is carried out.*

Format for Writing the Statement of the Problem Section

A persuasive statement of problem is usually written in three parts, which can easily be remembered by the acronym IRC (Ideal, Reality, Consequences);

- ➢ Part A (The ideal) - This is the first part of problem statement. It describes a desired goal or ideal situation; explains how things should be.

- ➢ Part B (The reality) - This immediately follows the ideal. It is the penultimate part of problem statement, which describes a condition that prevents the goal, state, or value in Part A, from being achieved or realized at this time. It lso explains how the current situation falls short of the goal or ideal.

- Part C (The consequences) - This is the last part of problem statement. It identifies the way you propose to improve the current situation and move it closer to the goal or ideal. i.e. What will happen if the problem is not addressed?

Here is an example of a problem statement in IRC format:

Part A:

The Higher Education Loans Board (HELB) is dedicated to allocating loans to university students. These funds are made available in order to enable them pay their school fees and other needs at the university. (Provide relevant statistics and quote)

Part B:

One of the main focuses of the HELB is consistency. Unfortunately, consistency in allocating funds to the next generation of recipients requires prior knowledge of previous allocations and established practices. The current continuous disbursement method does not allow for adequate analysis of previous disbursements before a current disbursement is done.

Part C:

Continuing with this current disbursement method prevents consistency and causes decisions to become grossly political, which in turn inhibits the achievement of the goals of the funds. Developing a more informed disbursement system could help better implement the consistency focus of the HELB and at the same time help to better monitor and evaluate its funds. This proposed research aspires to explore options for a new funds disbursement system that would focus on consistency. To do this, the researcher will carry out a full stakeholder analysis and use it to propose appropriate policy interventions by the Kenya Kwanza administration.

Another sample statement of the problem:

It is apparent that of all the science subjects being offered at secondary school level in Kenya currently, Chemistry is the worst performed, if the analyses of previous KCSE results provided by KNEC are anything to go by. The subject has always ranked poorly in KCSE for the last five consecutive years, especially when compared to Biology, which has a similar students' enrollment. Reasons leading to this predicament are well documented and among them are students' low self-efficacy, low motivation and negative attitude towards abstract topics. The widespread use of and overreliance on Conventional Instructional Strategies in abstract topics has only made matters worse.

Electrochemistry, a very important topic with a wide range of applications in industry, is one such topic, no wonder it is frequently tested in the national examinations, with high weightage given to its question items in the theory papers.

Poor performance in Electrochemistry question items should therefore be everyone's concern because it will eventually compromise production and development. However, the Conventional Instructional Strategies, which have been criticized for being mainly teacher-centered, continue to be used to teach it. This therefore means students will continue to perform poorly in the subject at national level unless research about more effective student-centered and metacognitive instructional strategies is done is the country and its recommendations adopted.

Software-Oriented Concept Mapping is one good example of a student-centered instructional strategy, which incorporates technology to promote creativity in the way learners think about, visualize and relate different concepts in a topic or subject. Elsewhere, this strategy has been found to impact positively on several aspects of students' academic performance but research about its use in the Kenyan context is scanty at the moment and hence this study couldn't have been done at a better time. In fact, there is no study known to the researcher so far, about its application anywhere in Kakamega County. It is on these premises that the current study will be carried out.

Conceptual Framework for Educational Research

A conceptual framework is a figure that illustrates the expected relationship between the research variables from the researcher's point of view. It identifies the relevant indicators, characteristics or properties of each variable and maps out how they come together to draw coherent conclusions about the investigation.

How to Develop a Conceptual Framework

Conceptual frameworks can be written or visual and are generally developed based on review of existing studies relevant to your topic. The following steps should be followed in order to come up with a good conceptual framework;

Step 1: Choose your research question - Your research question guides your work by determining exactly what you want to find out, giving your research process a clear focus.

Example: Research question

Let's say you want to study whether students who study more hours get higher test scores. To investigate this question, you can use methods such as an experiment or a survey to test the relationship between variables.

However, before you start collecting your data, consider constructing a conceptual framework. This will help you map out which variables you will measure and how you expect them to relate to one another.

Step 2: Select your independent and dependent variables - In order to move forward with your research question and test a cause-and-effect relationship, you must first identify at least two key variables; your independent and dependent variables.

Example of Variables:

Following the previous example, the expected cause, "hours of study," is the independent variable. The expected effect, "exam score," is the dependent variable. In other words, you suspect that "exam score" depends on "hours of study." Thus, your hypothesis will be that the more hours a student studies, the better they will do on the exam.

Note that causal relationships often involve several independent variables that affect the dependent variable. For the purpose of this example, we'll work with just one independent variable, "hours of study".

Step 3: Visualize your cause-and-effect relationship - Now that you've figured out your research question and variables, the next step in designing your conceptual framework is visualizing your expected cause-and-effect relationship. We demonstrate this, using basic design components - boxes and arrows. Here, each variable appears in a box. To indicate a causal relationship, each arrow should start from the independent variable (the cause) and point to the dependent variable (the effect).

Step 4: Identify the intervening variables - It's crucial to identify other factors that may possibly influence the relationship between your independent and dependent variables early in your research process. Such factors are collectively called *intervening variables*. There are three main types of intervening variables. These include; *moderating*, *mediating*, and *control* variables.

Moderating Variables

Moderating variables (or moderators) are intervening variables that alter the effect of an independent variable on the dependent variable. In other words, they change the "effect" component of the cause-and-effect relationship. The following is an example of a moderator:

> *We expect that the number of hours a student studies is related to their exam score i.e. the more you prepare, the higher your score will be. Let's add the moderator "IQ." Here, a student's IQ level can change the effect that the variable "hours of study" has on the exam score. The higher the IQ, the fewer hours of study are needed to do well on the exam. Obviously, we expect that the "IQ" moderator moderates the effect that the number of study hours has on the exam score.*

Mediating Variables

Mediating variables are intervening variables that link the independent and dependent variables, allowing the relationship between them to be better explained. An example of a mediator is as follows:

> *The mediating variable of "number of practice problems completed" comes between the independent and dependent variables. Hours of study impacts the number of practice problems, which in turn impacts the exam score. In this case, the mediator helps explain why studying more hours leads to a higher exam score. The more hours a student studies, the more practice problems they will complete; the more practice problems completed, the higher the student's exam score will be. Keep in mind that mediating variables can be difficult to interpret. Take care when drawing conclusions from them.*

Difference between Moderator and Mediator Variables

It's important not to confuse moderating and mediating variables. To remember the difference, you can think of them in relation to the independent variable: A moderating variable is not affected by the independent variable, even though it affects the dependent variable. For example, no matter how many hours you study (the independent variable), your I.Q will not rise. On the other hand, a mediating variable is affected by the independent variable, which in turn, affects the dependent variable. It therefore bridges the two variables and helps explain the relationship between them.

Control Variables

These are intervening variables that are held constant so that they don't interfere with the results. On another day, they could as well be your independent variables. Control variables must also be taken into account even though you aren't interested in measuring them for your study. It is crucial to be aware of as many of them as possible, as this has implications on the way you'll interpret your research findings. The following is an example of Control Variable:

It is very much possible that if a student feels ill, they will most likely get a lower score in the exam as compared to those who were okay health wise. However, we are not interested in measuring health status of learners a part of our research. This makes "health" a good candidate for a control variable as it still impacts our results, but we aren't interested in studying it.

Adding the variable "health status" to our conceptual framework, but decide to keep it constant means we'll *only admit participants who are in good health* on the day of the exam (data collection). The next two figures show a dummy of a conceptual framework and an example from the just described situation:

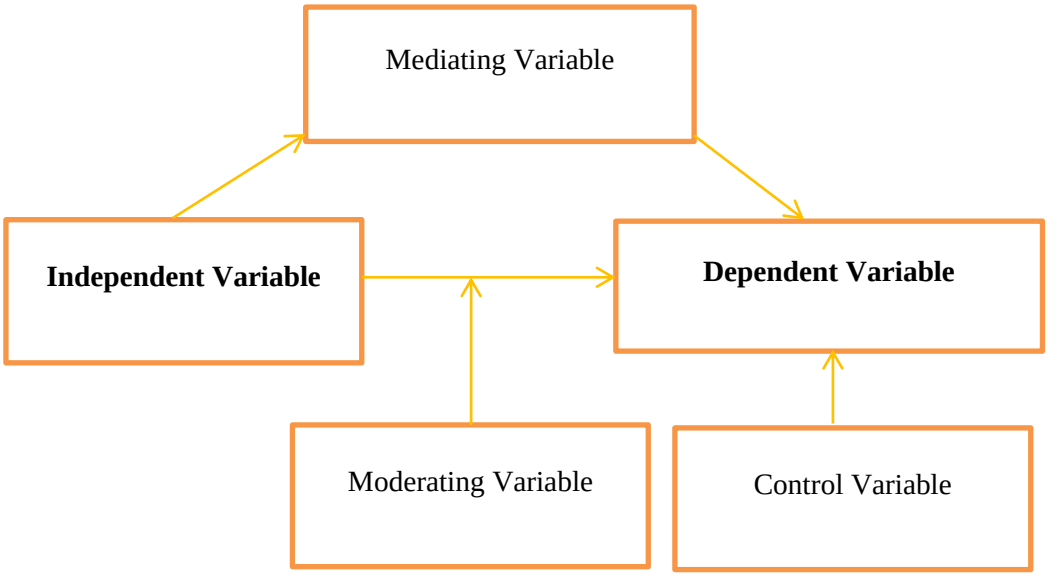


Figure 3: Dummy of a Conceptual Framework

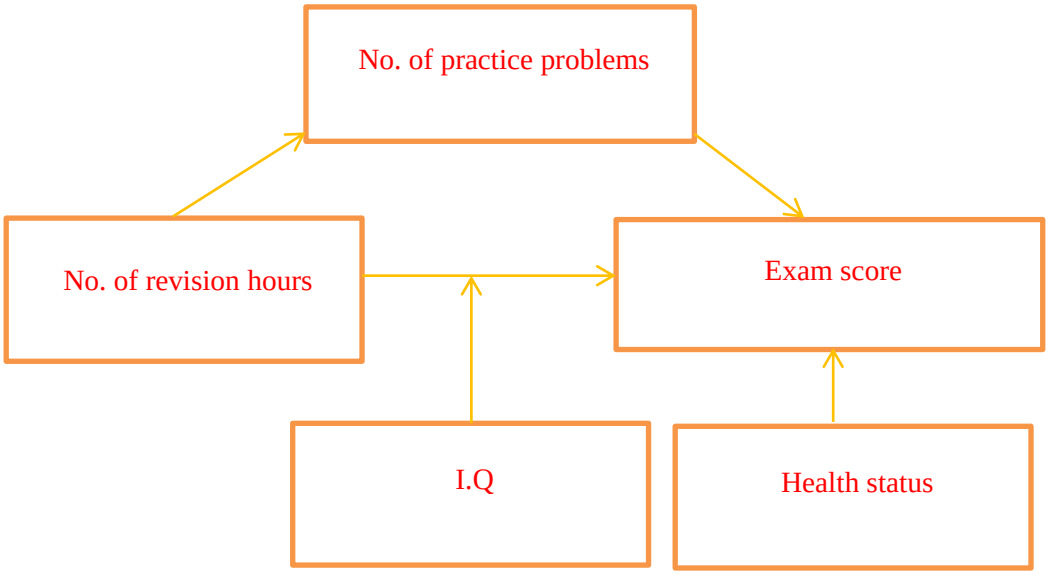


Figure 4: Example of a Conceptual Framework

Theoretical Framework for Educational Research

A theoretical framework is a foundational review of existing theories that serve as a road-map for developing the arguments you will use in your own research. Theories are developed by educational researchers to explain educational phenomena, draw connections, and make predictions.

In a theoretical framework, you should explain the existing theory that supports your research, showing that your work is based on well-known ideas. In other words, your theoretical framework justifies and contextualizes your later research, and it's a crucial first step for your research paper, thesis, or dissertation. A well-rounded theoretical framework sets you up for success later on in your research and writing process.

Importance of Theoretical Framework in Educational Research

Before embarking on your own research, it's crucial to familiarize yourself with the theories and models that other researchers have already developed. The theoretical framework is your opportunity to present and explain what you've learned and situated within your current research topic. There's a good chance that many different theories relevant to your topic already exist, especially if the topic is broad. Don't insist on the most recent one at the expense of the most relevant to your research problem. Theories rarely expire!

In your theoretical framework, you need to evaluate, compare, and select the most relevant one. By mapping your research within a clearly defined educational field, you make the reader aware of all assumptions that inform your approach, showing the rationale behind your choices for later sections, like methodology and discussion. This part of your work therefore lays the foundations that will support your data analysis, helping you interpret your results and make broader generalizations. The same research topic can be approached very differently depending on which theoretical approach you take, even within the same field. In educational psychology for instance, a behaviorist approach to learner's depression would involve different research methods and assumptions than a psychoanalytic approach. In this case, you should explain the theory you're using in order to support your interpretations and explanations of the data you collected.

How to Write a Theoretical Framework in Educational Research

Note that a theoretical framework can sometimes be integrated into the literature review chapter, but it can also be included in the latter sections of chapter one in a proposal or thesis. There are different justifications for each of these and as such, there are no fixed rules for structuring your theoretical framework, but it's best to double-check with your education department or faculty to make sure you adhere to set guidelines. To create your own theoretical framework, you can follow these three steps; (i) identify your key concepts, (ii) evaluate and explain relevant theory, (iii) show how your research fits into existing theory. Let us examine each of these steps;

- *Identifying your key concepts* - the first step is to pick out the key terms or variales from your problem statement and research questions. Concepts often have multiple definitions, so your theoretical framework should also clearly define what you mean by each term.

Example of problem statement and research questions;

> *A university is struggling with the problem that many of their students who attend classes online perform dismally in their end-of-semester examinations as compared to those who do the same face-to-face. This has become a big issue for the university, which has few lecture halls and consequently a small student population. Therefore, the university wants to increase the number of students in their institution without straining the few and over-stretched facilities they have, by offering some of their courses online. They believe that good academic performance by online students will play a major role in achieving their goal of increased student enrolment.*

To investigate this formulated problem, the researcher could identify and plan to focus on the following problem, objective and research questions;

Problem = Many online students perform poorly in their end-of-semester examinations.

Objective = to determine the effect of online teaching on students' academic performance.

Research question = How does online teaching affect students' academic performance?

The key terms are "online teaching" and "academic performance", which are central to this study, along with their relationship to the likelihood that a student enrolment might consequently increase. The theoretical framework should define these concepts and discuss theories about the relationship between them.

- *Evaluating and explaining relevant theories* - by conducting a thorough literature review, you can determine how other researchers have defined these key concepts and drawn connections between them. As you write your theoretical framework, the aim must be to compare and critically evaluate the approaches that different authors have taken.

  After discussing different models and theories, you can establish the definitions that best fit your research and justify them. You may even combine theories from different fields to build your own unique framework if this better suits your topic. Make sure to at least briefly mention each of the most important theories related to your key concepts. If there is a well-established theory that you don't want to apply to your own research, explain why it isn't suitable for your purposes. This will prevent critics from making comments like "why didn't you use theory X?"

- *Showing how your research fits into existing research* - apart from summarizing and discussing existing theories, your theoretical framework should also show how your study will make use of these ideas and take them a step further. You might aim to do one or more of the following;

  a. Test whether a theory holds in a specific, previously unexamined context

  b. Use an existing theory as a basis for interpreting your results

  c. Critique or challenge a theory

  d. Combine different theories in a new or unique way

CHAPTER 2:

LITERATURE REVIEW IN EDUCATIONAL RESEARCH

The second chapter of this book gives the definition of literature review, highlights the importance of literature review, lists the various components of the literature review section in a research report, outlines the steps to follow while reviewing literature, discusses gaps of a literature review, how and where to identify them. All these are important aspects that must be mastered by any researcher, which makes the topic worth your careful readership.

Definition of Literature Review

Literature review is a dissection of previously published works that are related to your research topic. The term Literature refers to a full scholarly paper or a section of a scholarly work such as a book. Either way, literature review is supposed to provide the researcher with a general picture of existing knowledge in the topic under investigation. Good literature review helps one to properly formulate research questions and select appropriate research methodology. To be precise, literature review serves to situate the current study within the body of other relevant existing research works, to provide context for the reader.

Literature review also presents the current knowledge including substantive findings as well as theoretical and methodological contributions to a particular topic. To this end, Literature reviews are secondary sources of information and do not report new or original experimental work. Most often associated with academic-oriented literature, such reviews are found in academic journals and should not to be confused with book reviews, which may also appear in the same publication. Literature reviews form a basis for research in nearly every academic field.

Importance of Literature Review in Educational Research

Without carrying out a comprehensive literature review, your research stands the risk of running into technical hitches. The following are the reasons why literature review should be done;

i. Provides foundation of knowledge on the research topic
ii. Identifies well-researched areas to prevent duplication and give credit to other researchers

iii. Identifies inconsistencies, gaps, conflicts and open questions left by other researchers
iv. Identifies the need for additional research, which is also a justification for your research
v. Identifies the relationship between previous works and its contribution to your research
vi. It brings clarity and focus to your research problem, helping refine your topic to a specific area
vii. It gives insight on the best research methods to use in your study
viii. It broadens your knowledge base in the research topic

Components of a Good Literature Review Section

A comprehensive and well-written literature review section in a thesis or paper should comprise of the following in every individual paper read;

- ✓ A brief description of the publication (author, country, year, topic)
- ✓ A summary of the publication's main points (objectives, methodology, findings)
- ✓ A discussion of gaps in research (what was left out due to limitations)
- ✓ An evaluation of the publication's contribution to your topic or discipline.

Key Steps in Literature Review

To ensure your literature review process is systematic and well-organized, the following steps should be followed;

1. *Narrow down your topic* - consider your specific area of study. Think about what interests you and what attracts other researchers in your field. Limit your scope to a smaller topic area e.g. reading on use of software-oriented concept mapping instead of looking at computer assisted instruction which is very wide.

2. *Search for relevant literature* - define your source selection criteria e.g. articles published within a specific time range, focusing on a specific geographic region, or using a specific methodology. Use keywords to search a library or online repository. Reference lists of recent articles and reviews can lead to other useful papers. Be sure to include any studies contrary to your point of view. This initiates, livens up and enriches discussion.

3. *Read and critique* - the selected articles should be thoroughly read and critiqued. The researcher should evaluate and synthesize the studies' findings and conclusions. Take note of assumptions most researchers have made, their methodologies, testing procedures, sampling, instruments, theories and main findings.

4. *Organize the selected papers* - look for patterns and develop sub-themes that stand out from published works. Note the following; findings that are common or contested, apparent trends in the findings, the most influential theories. A useful tip if your literature review is extensive is to find a large table surface and on it, place notes or fill cards to help organize all your findings into categories.

   You may move them around if you realize that they fit better under different sub-themes or you need to develop themes and sub-themes that reflect the major variables you identified.

5. *Compile the literature* - follow the organizational structure you developed above, including the headings and sub-headings you constructed. Ensure each section flows logically with the one before and after. Structure your sections by themes or subtopics, not by individual theorists or researchers. In case you realize that each paragraph begins with a researcher's name, it might indicate that instead of evaluating and comparing the searched literature from an analytical point of view, you have simply described what research has been done. You need to prioritize analysis over description. For example, look at the following two passages:

Researcher 1:

> *Mbacho (2021) concludes that personal privacy in their living quarters is the most important factor in nursing home residents' perception of their autonomy. He suggests that the physical environment in the more public spaces of the building did not have much impact on their perceptions. Neither the layout of the building nor the activities available seem to make much difference. Makori (2021) make the claim that the need to control one's environment is a fundamental need of life, and suggest that the approach of most institutions, which is to provide total care, may be as bad as no care at all. If people have no choices or think that they have none, they become depressed.*

Now compare with the next researcher.

Researcher 2:

> *After studying residents and staff from two colleges in Nairobi, Mbacho (2021) came to the conclusion that except for the amount of personal privacy available to residents, the physical environment of these institutions had minimal if any effect on their perceptions of control (autonomy). However, Makori (2022) found that availability of private areas is not the only aspect of the physical environment that determines residents' autonomy. He interviewed 115 residents from 32 different nursing homes in Bomet, which are known to have different levels of autonomy. He found that physical structures, such as standardized furniture, heating that could not be individually regulated, and non-possession of a house key for residents limited their feelings of independence. Moreover, Mbogani (2022), who interviewed 225 residents from various nursing homes in Kakamega, substantiates the claim that characteristics of institutional environment such as the extent of resources in the facility, as well as its location, are features that residents have indicated to be of great importance to their independence.*

It can easily be noted from the two examples of that researcher 1 merely describes the literature, whereas researcher 2 takes a more analytical and evaluative approach by comparing and contrasting the literature under review. You can also see that this evaluative approach is well signaled by linguistic markers that indicate logical connections (words such as "however", "moreover") and phrases like "substantiates the claim that…" which all indicate supporting evidence and Student 2's ability to synthesize points.

6. *Review and cite your sources* - look at the sentences in each paragraph to ensure they indicate the main points of your literature review sources. Make an outline of each section of the paper and decide whether you need to add information, delete irrelevant information, or re-structure sections. You may read your work out loud to be able to identify where you need punctuation marks to signal pauses or divisions within sentences where you have made grammatical errors or where sentences are unclear.

   In educational research, it is important that literature be quite recent. Make sure that all citations and references are correct and that you are referencing in APA format, which is the most common in education. Also check to make sure that you have not plagiarized,

either by failing to cite the source of information or by using words quoted directly from a source.

The following figure gives a summary of the main steps in conducting literature review for educational research:

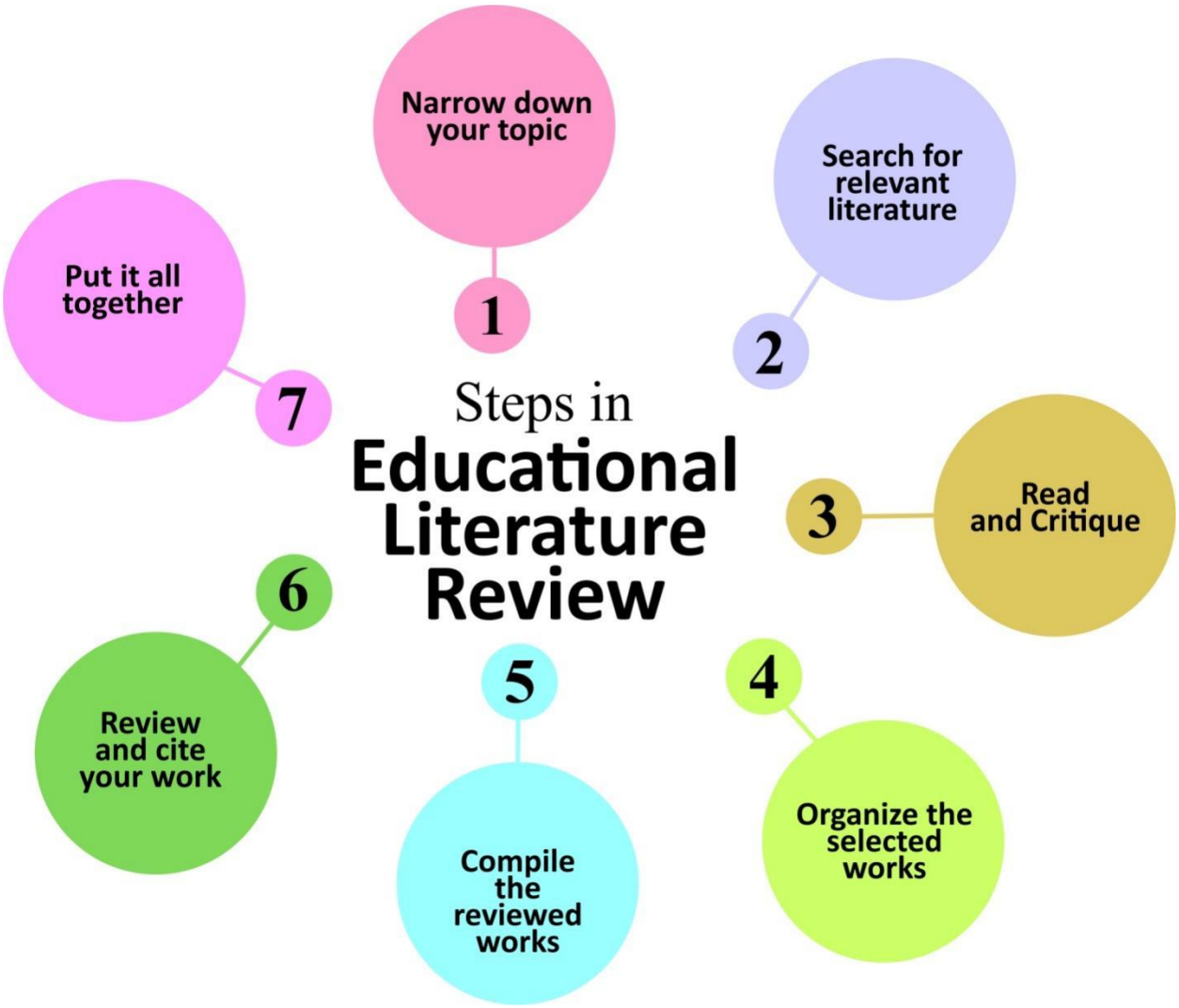


Figure 5: The Literature Review Process

Gaps in Literature Review

If you are a researcher, the first challenge towards starting your research is to identify an interesting and researchable topic. Most researchers come up with a great research problem while others, especially postgraduate students may struggle to get an acceptable research problem. The best solution to this issue lies in the research gap. A research gap is an area where not enough research has been done, or whatever has been done so far may have been overtaken by recent events.

Gaps in the literature are therefore the missing pieces of information in existing literature. They include areas that have room for further research because they are not sufficiently explored, or misreported. It is a research question that remains unanswered, unresolved or inadequately addressed by previous studies. Examples of gaps could be on: -

- ✧ Population or sample size
- ✧ Study location
- ✧ Research design
- ✧ Data collection / analysis procedures
- ✧ Research variables or conditions

Research gaps fall into six different categories; knowledge gap, evidence gap, methodology gap, empirical gap, theoretical gap and population gap. The following table sheds light on each of these;

Table 3: Description of the Various Types of Research Gaps

Type of Gap	Description
Knowledge gap	Expected research findings do not exist
Evidence gap	Results from similar previous studies that used same methodology but results contradict one another
Methodological gap	One type of research design or method was used in previous studies yet there many others that could work well in the study
Empirical gap	Research findings from previous studies have not been verified
Theoretical gap	A new theory needs to be applied to the study problem to generate new insights
Population gap	The sample used was not stated or was not sufficient to adequately represent the target population

Difference between Gap and Problem in Educational Research

Novice researchers find it challenging to differentiate a research gap from a research problem. Simply put, a research gap is the research problem discovered by the researcher after a detailed literature review. The problem statement on the other hand is a constructed paragraph that is derived from a research gap and it describes what cannot happen unless your research is carried out.

Identifying Gaps in Existing Literature

If you do not find any articles in your literature search that are related to your study, this may already indicate a gap. If you do find articles, the deal is to find a gap that will contribute to new knowledge if explored. Authors of previous research works signal that there is a gap using words such as;

- …… has not been clarified, studied, reported, or elucidated
- Further research is required or needed on …..
- …... is not well reported
- Suggestions for further research include ……..
- The key question that remains unanswered is …….
- It is important to re-look or re-think about …..
- …... is poorly understood or is still unknown
- Research on ….. is scanty
- The question on …….. remains answered

Conducting a comprehensive literature review is the game changer for identifying gaps in existing literature. As you search for journal articles, you will need to read critically across the breadth of the literature to identify gaps using the tips listed above. Your goal should be to find a 'space' or 'opening' for contributing new knowledge. To identify a research gap, the following steps should be followed;

Conduct literature review: Gathering a broad range of research articles on your topic is crucial. Look for articles, books, and other sources that address the research objective you want to investigate. You may want to look for research that approaches the topic from a variety of methods - qualitative, quantitative or mixed methods.

Analyze the literature. As you read the literature, analyze the research methods used, the findings and conclusions made in each study. i.e. who, what, when, where and how - about the population, setting, conditions, variables, methods, analysis and results. Note any inconsistencies, contradictions or limitations as these give rise to possible 'gap' or opening for your potential research and contribution to the topic.

Identify a research question and define the gap therein. Remember that researchers may not explicitly state that a gap in the literature exists, so you may need to thoroughly review and assess the research to make that determination by yourself. You can use organizational tools such as charts or Venn diagrams to map out the research you find from scholarly articles. These methods may be helpful to organize what information you have found and what is shared among the literature, as well as to identify what areas may be missing in the research. You may also want to

use search words such as “literature gap” or “future research” along with your subject keywords to pinpoint articles that include these types of statements. To this end, you need to identify a research question that has not been sufficiently addressed and explain why you think it has not addressed the problem you have identified.

Where to Find Research Gaps in Scholarly Work

As you begin to gather the literature, you may want to critically read for what has or has not been done before. The following are some of sources of research gaps in the searched literature;

- ✧ Use the discussion, limitations and suggestions for future research sections of the articles or dissertations to understand what the researchers have found and where they point out future research areas.
- ✧ Several educational research areas can be found on digital platforms like educational social media groups, research institution websites, educational magazines like *mwalimu weekly*, education weekly and education journals. These sources give an idea of that could be picked up or modified from their scope and limitations.
- ✧ A few education-based organizations in Kenya like CEMASTEA, KICD, KNEC and KESI focus a lot on problem-solving areas in the education sector. They also keep writing and mentioning pertinent academic issues in their newsletters or they put it in public domain via different platforms. Such articles are excellent sources of real educational issues from which a research gap could be identified and followed up.
- ✧ The most cited research articles or research papers may give a clue on the inspiring and motivating research gaps for you. These help with the understanding of current topics among other educational researchers and identify gaps in their research work.

CHAPTER 3:

RESEARCH DESIGNS IN EDUCATIONAL RESEARCH

This chapter gives the definition of a research design, examines the structure and style of the research design section in educational research, enumerates the characteristics of a good research design, discusses the characteristics and types of research designs, highlights on the validity of research designs, examines threats to validity of research designs and how to counter them in educational research.

Before beginning your research, you need to have a blue print, just the way a person who intends to build a good house must have an architectural plan, which at a glance gives details of the number of rooms, dimensions and layout of each room, location of key installations, type and quality and of materials needed etc. equating research to building a house, the house blue print corresponds to the research design.

Definition of a Research Design

A research design refers to the overall strategy or plan for integrating the different components of a study in a coherent and logical way, to ensure the research purpose is effectively executed. It constitutes the blueprint for the data collection, data analysis and generalization of findings. Note that your research problem determines the type of design you should use, not the other way around!

Structure and Style of the Research Design Section

The function of a research design is to ensure that any evidence obtained from research data enables you to effectively address the research problem as unambiguously as possible. Research in education entails obtaining evidence relevant to the research problem generally and specifying the type of evidence needed to test the underlying theory, to evaluate a program or to accurately describe a phenomenon.

However, some researchers often begin their investigations far too early, before they think critically about what information is required to answer their research questions. Without attending to these design issues beforehand, conclusions drawn from the findings risk being weak or

unconvincing and may not adequately address the problem that necessitated the study. In case of a funded research, imagine how disappointed the donor would be! Given this, the length and complexity of research designs can vary considerably, but the bottom-line is that any good research design will do the following things: -

- ✓ Address the research problem clearly
- ✓ Address the research questions exhaustively
- ✓ Yield data that will adequately support the hypotheses
- ✓ Provide for circumventing possible threats to external and internal validity

Characteristics of a Good Research Design

A good research design must have the following four characteristics;

(i) Reliability - a good research design must produce consistent findings when replicated under similar research circumstances by an independent researcher.

(ii) Validity - a good research design must yield data that fully addresses the research objectives.

(iii) Neutrality - When setting up your study, you may have to make assumptions about the data you expect to collect from the field. The results projected in the research should be free from bias and neutral. This calls for good understanding of opinions about the final evaluated scores and conclusions from multiple individuals and consider those who might agree, disagree or question the derived results.

(iv) Generalizability - The outcome of your design should apply to the target population and not just a restricted sample. A generalized design implies that your research can be conducted on any part of a population with similar levels of accuracy.

Note that all the four factors affect the way respondents answer the research questions and so, all the mentioned characteristics should be carefully taken care of by a good research design.

Classification of Research Designs

Research designs can be classified basing on several criteria e.g. by type of data, by grouping of respondents, by methodology just to mention a few. In educational research however, research

designs are usually classified according to methodology adopted i.e. qualitative, quantitative and mixed methods.

Quantitative research designs entail collection of quantitative (continuous) data and using it to address the research objectives, where inferences are made about various parameters that describe the population of the study. These Research Designs examine various interval or ratio variables while including numbers to analyze its findings and come up with a verifiable evidence-based conclusion. Quantitative research instruments are used to collect data, which is thereafter presented by the use of bar graphs, tables, pie charts and histograms.

Qualitative research designs on the other hand are quite the contrary, as they are explanatory in nature and always seek answers to "What" and "How". They mainly focus on why a specific theory exists and what would be the respondent's answer to it. This therefore allows the researcher to draw a conclusion with elaborate findings.

Mixed methods Research designs are blend between the two. Let us now discuss each of these broad categories in detail.

Qualitative Research Designs

To perform qualitative research, you must carefully choose at least one research design approach that fits your topic. It is not uncommon for a researcher to employ more than one design or approach throughout their study. Here are five common qualitative designs in educational research:

i. Historical Study Design

A historical study is the ideal choice for studies that involve extensive examination of the past, including people, events and documents. The purpose of historical study designs is to draw conclusions about the present and future, based on research conducted in the past. This design relies on adequate interview sources and historical documents. It is essential to validate the accuracy of the data and find primary sources. Depending on the goal of the researcher, this form of study may result in a biography, which is why the term "historical study" is sometimes used interchangeably with "biographical study."

ii. Phenomenology

Phenomenology is a wide-ranging form of study. In this research design, the researcher seeks to gather information that explains how individuals experience a phenomenon and how they feel about it. This model recognizes that there is no single objective reality. Instead, everyone experiences events differently. The outcome is described from the point of view of the participants. However, the researcher is still able to derive a set of findings that can be used to identify themes surrounding the phenomena under investigation.

The goal of phenomenological design is to describe the meaning that experiences hold for each respondent. This type of research is used to study topics in which there is little knowledge. Respondents are asked to describe their experiences as they perceive them. They may write about their experiences, but information is generally obtained through interviews. To understand the lived experience from the vantage point of the subject, the researcher must take into account their own beliefs and feeling. The researcher must first identify what they expect to discover and then deliberately put aside these ideas. This process is called *bracketing*.

It is only when the researcher puts aside their own ideas about the phenomenon that it may be possible to see the experience from the eyes of the person who has lived the experience. Phenomenological research would ask a question such as, “What is it like for a mother to live with an autistic child?” The researcher might perceive that they would feel burdened. These feelings would need to be identified and then put aside to listen to what the mother is saying about how she is living through this experience of having to give extra care to their child. It is possible that this mother has discovered tricks for making their autistic child’s life better.

iii. Grounded Theory Design

The purpose of grounded theory is to develop a theory surrounding a social issue in educational institutions. This theory seeks not only to identify problems in education, but also to define how people deal with those problems. Grounded theory is unique among qualitative research designs or approaches, because it depends solely on the data gleaned through the research process. Often, the initial research question is progressively reformed and refined as more information is gathered on the topic. In this sense, the participants help shape the study.

The unit of analysis in grounded theory design is a specific event or incident, not individual behaviors. Researchers follow a systematic set of procedures and a simultaneous approach to data collection and analysis. Grounded theory is most often used to generate rich explanations for complex actions, processes and transitions. The primary mode of data collection is one-on-one participant interviews. Sample sizes tend to range from 20 to 30 individuals, sampled purposively. However, sample sizes can be larger or smaller, depending on data saturation. **Data saturation** is the point in the qualitative research data collection process when no new information is being discovered. Researchers use a constant comparative approach in which previously collected data are analyzed during the same time frame as new data are being collected. This allows the researchers to determine when new information is no longer being gleaned from data collection and analysis i.e. that data saturation has been reached, in order to conclude the data collection phase of the study.

iv. Ethnography Design

Ethnography is the investigation of a specific group of people within their cultural context. Researchers pursuing this study format will immerse themselves into the culture they are researching. The researcher (ethnographer) must be deeply immersed in the social culture over an extended period of time (usually 8 months to 2 years) and should engage, observe and record the daily life of the studied culture and its participants within their natural setting. The information is then presented from their perspective.

Ultimately, this study aims at understanding group culture. Ethnographers interview people who are most knowledgeable about the culture e.g. the longest serving inmate in a prison, a class teacher who has taught a student from form one to form four, a human resource officer in a company, a retired head teacher etc. These people are called key informants. Data collection and analysis occur simultaneously when using ethnography design. As understanding of the data occurs, new questions and insights emerge.

The researcher must therefore take extensive field notes and narrate his or her experience in descriptive detail, so that readers may experience the same culture as the researcher. In this research design, the researcher has two roles; first is to use his or her unique knowledge and engagement to generate insights and secondly to convince the scientific community of the trans-

situational nature of the studied culture. The ultimate goal of ethnography is to develop cultural theories.

v. Case Study Design

Case studies are some of the most common qualitative designs in educational research. They are used to examine a person, group, community or institution. Researchers often use a bounded theory approach that confines the case study in terms of time or space. To conduct a case study, the researcher may tap from multiple sources of data, such as observation, interviews and documents. All participants chosen must share a unifying factor, which means they all must have a connection to the research question or subject being studied. After collecting data, the researcher will analyze it to identify common or prominent themes.

Case study research is technically an intensive longitudinal study of a phenomenon at one or more research sites for the purpose of deriving detailed, contextualized inferences and understanding the dynamic process underlying a phenomenon of interest. Case research is therefore a unique research design in the sense that it can be used in an interpretive manner to build theories or in a positivist manner to test theories. Furthermore, the case researcher is a neutral but direct observer in the social setting, rather than an active participant. Drawing meaningful inferences from case research depends heavily on the observational skills and integrative abilities of the researcher.

In summary, the following are the major differences between the two broad categories of research designs:

Table 4: Difference between Quantitative and Qualitative Research Designs

Quantitative Research Designs	Qualitative Research Designs
1. Involve continuous data/variables	Involve discrete data/variables
2. Ideas are carefully converted into hypotheses and tested statistically to come up with a verifiable conclusion	Ideas are converted to research questions which are answered using respondents' views about the research problem

3. Involve use of intricate mathematical and statistical analysis to manipulate the data to come up with inferences	Data is only summarized, classified or categorized, to identify patterns and trends narrated in themes as they emerge
4. Involves a large sample size in order for the findings to be generalizable to the whole population	Few knowledgeable people are needed, their responses are adequate to answer research questions
5. Fixed responses are used (for questionnaires)	Are open-ended inquiries
6. Key terms; testing, measurement, objectivity, replicability, objectivity	Key terms; understanding, context, complexity, subjectivity

Quantitative Research Designs

Quantitative research designs can be classified into two broad categories; experimental and non-experimental. This section digs deep into each of these.

Experimental Research Designs

Experimental research is a scientific approach to research, where one or more independent variables are manipulated and applied to one or more dependent variables to measure their effect on the latter. The effect of the independent variables on the dependent variables is usually observed and recorded over some time, to aid researchers in drawing a reasonable conclusion regarding the relationship between these two variable types. Experimental research designs are widely used in education. They work by comparing two or more groups with a straightforward logic, which may at times be difficult to execute.

Mostly related to laboratory-like test procedures, experimental research designs involve collecting quantitative data and performing statistical analysis on them during research. This therefore makes them subsets of quantitative research methods. In education, experimental research designs are generally used to investigate the various standards of an academic institution.

These may include but not limited to; testing students' achievement in various topics, coming up with better instructional methods or teaching aids and the implementation of new programs that will aid effective student learning processes.

Experimental research designs are of different types, depending on the way the researcher assigns subjects (respondents) into different conditions or groups. There are of 3 broad types namely; pre-experimental, quasi-experimental and true experimental research designs.

To make it easier to explain each of these, the following notations will be used to illustrate each particular design:

X = Treatment (intervention)
O = Observation (measurement)
R = Random assignment to a group
C = Control (no intervention)

A. Pre-Experimental Research Designs

In pre-experimental research designs, either a group or various dependent groups are observed for the effect of applying an independent variable, which is presumed to cause the change in the dependent variable. It is the simplest of all experimental research designs, as it has no control group. Although very practical, pre-experimental research designs are deficient in several areas of the true-experimental criteria. The pre-experimental research designs are further divided into three sub-types; one shot case study, one group pre-posttest design and static group comparison design.

The three pre-experimental research designs are discussed hereunder:

i. One Shot Case Study Design - This involves a single group studied only once. A group is introduced to a treatment or condition and then observed for changes which are then attributed to the treatment thus:

X O

ii. One Group Pretest/Posttest Design - This involves presentation of a pretest (O_1), followed by a treatment (X), and then a posttest (O_2), where the difference between O_1 and O_2 is explained by X thus:

$$O_1 \ X \ O_2$$

iii. Static Group Comparison Design - This is a two group design, where one is exposed to a treatment and the results are tested while a control group is not exposed to treatment (placebo) and similarly tested in order to compare the effects of treatment.

$$X \ O_1$$
$$C \ O_2$$

B. Quasi-Experimental Research Designs

The word "quasi" in research means "partial" or "resembling". Quasi-experimental research designs bear some resemblance to the true experimental research designs, but are not the same. In quasi-experiments, the participants are not randomly assigned into groups, and as such, they are used in intact settings where randomization is difficult or impossible. This is very common in educational research, where administrators are unwilling to allow the random selection of students for experimental samples. Some examples of quasi-experimental research design include; the time series, non-equivalent control group design, and the counterbalanced design.

i. **Time Series Design -** this is a quasi-experimental research design where repeated measurements of a given behavior are taken across time at equally spaced intervals. Taking multiple measurements is essential for understanding how any given behavior unfolds over time, and doing so at equal intervals affords a clear investigation of how the dynamics of that behavior manifest at distinct time scales.

ii. **Counter-Balanced Design -** this is a quasi-experimental research design that allows for testing of multiple interventions concurrently. The counterbalanced design is often employed when several interventions are being tested concurrently. After random selection from the population, the study subjects are divided into several independent groups, say grp 1, grp 2 ….. grp n. Thereafter, each group receives all the interventions, say T_1, $T_2 \ldots T_n$, but in a different unique order from the other groups.

An observation of the outcome is taken after each series of interventions in the counterbalanced design (O_1, O_2 ,…. O_n). The counterbalanced design is a powerful

design because each participant serves as their own control and multiple interventions can be tested at the same time. The figure below illustrates this, using 4 groups i.e. n=4 thus:

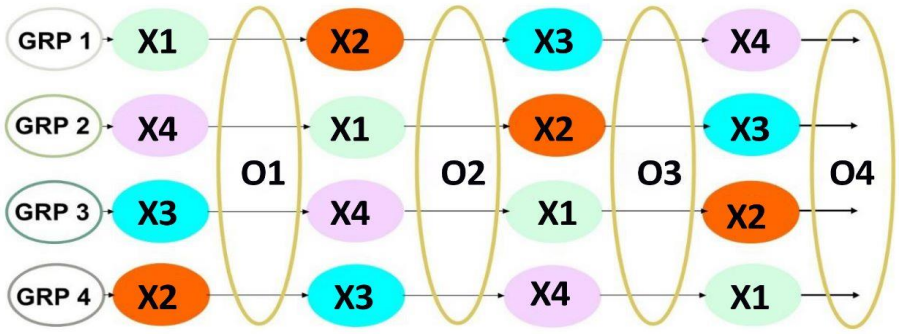


Figure 6: Illustration of the Counter-Balanced Research Design with 4 groups

X = Treatment, O = Measurement

iii. **Non-Equivalent Control Group Design** - This is a type of quasi-experimental research design that is mostly used to detect associations at the intervention level. The non-equivalent control group design is perhaps the most popular of the quasi-experimental research designs. The non-equivalent control group design is very effective while investigating groups that already exist in a given population e.g. class streams. Randomization occurs at the intervention level. Study groups are randomly assigned to either receive the intervention treatment or receive a control treatment e.g. in a 2-streamed school, Form 1N is given treatment (experimental group) while form 1S is not given treatment (control).

After checking to ensure that the intervention groups have similar entry behaviour, a baseline measure of the outcome is then taken from each group. Then, the intervention is given to the treatment group and the control intervention (if any) is given to the control group. A second measure of the outcome is taken and compared to the baseline measure to see if the treatment had a significant effect in comparison to the control. This is considered a mixed-effects type of analysis with both between-subjects

(treatment vs. control) and within-subjects (baseline vs. post) effects being tested for a potential interaction.

It is possible to have a third or maintenance observation of the outcome to see how stable the treatment effects manifest across time with a nonequivalent control group design. The choice to include a maintenance observation gives much more credence to any significant treatment effects found between just two observations. Valid and effective treatments should have lasting and consistent effects in real populations

C. True Experimental Research Designs

This is the crème de la crème of all research designs and are the most strongly recommended designs. There are three types of true experimental research designs;

i. **Pre-test/Post-test with Control Group Design**. This design takes the form illustrated below:

R	O_1	X	O_2
R	O_3	C	O_4

O_1, O_3 = Pretest Observations X = Treatment (intervention)

O_2, O_4 = Post-test Observations C = Control (placebo)

In this research design, subjects are randomly assigned to two different groups; one experimental, one control, both subjected to pre-test and post-test. This design controls for all major threats to validity.

ii. **The Solomon Four-Group Design** - The design is implemented as follows:

R	O_1	X	O_2
R	O_3	C	O_4
R		X	O_5
R		C	O_6

In this design, subjects are randomly assigned to four different groups; 1 experimental with both pre- and posttests, 1 experimental with posttest only and no pretest, 1 control with pre- and post-tests, and 1 control with posttest only without pretest. By using experimental and control groups with and without pre-tests, both the main effects of testing, interaction and treatment are controlled. Generalizability therefore increases and the effect of X is replicated in four different ways. Note that $O_1 = O_3$ and $O_2 = O_4 = O_5 = O_6$. If results reveal that pretests are greater in magnitude than pretests, this implies that the intervention treatment was effective.

iii. **Post-test Only Control Group Design -** This design is as illustrated below:

R	X	O_1
R	C	O_2

It involves two randomly selected groups, one randomly assigned into control group and the other into experimental group. Both are post-tested after their respective treatments. This design can be thought of as the last two groups in the Solomon 4-group design and can be seen as controlling for testing and interaction effects, but unlike Solomon four, this design, it doesn't measure them. Measurement of these effects isn't necessary to the central question of whether or not X has an effect. This design is appropriate in circumstances when pretests are not acceptable or important. Statistical tests for this simple design would be the t-test. However, covariance analysis and blocking on subject variables (prior grades, test scores etc.) can be used to increase the power of the significance test similar to what is provided for by a pretest.

D. Non-Experimental Research Designs

Non-experimental research is the type of research that does not involve manipulation of independent variables. In non-experimental research designs, researchers measure variables as they naturally occur, without any further manipulation. Such research designs are used when the researcher has no specific research question about a causal relationship between 2 different variables, and therefore manipulation of the independent variable is impossible or unnecessary.

In educational research, non-experimental research designs are also used when;

- ✓ Subjects cannot be randomly assigned to groups.
- ✓ The research subject is about a causal relationship but the independent variable cannot be manipulated.
- ✓ When the research question is about one variable rather than a statistical relationship between two variables.
- ✓ The research pertains to a non-causal relationship between variables.
- ✓ Limited information can be accessed about the research subject.

There are 3 main types of non-experimental research designs that are common in educational research; cross-sectional, correlation and observational surveys.

i. **Cross-Sectional Survey Design** - this design involves comparison of two or more preexisting groups of people under the same criteria. This design is classified as non-experimental because the groups are not randomly selected and the independent variable is not manipulated.

   For example:

   > *A certain university may want to reward its first-class bachelor's degree graduands with a master's degree scholarship for their academic excellence. Therefore, each faculty or school places students in the eligible and ineligible group according to their degree classification.*

In this case, the student's class of degree cannot be manipulated to qualify him or her for a scholarship because it is an unethical thing to do. Therefore, the placement is cross-sectional.

ii. **Correlational Survey Design -** this design compares the statistical relationship between two variables. It is considered to be non-experimental because it does not involve manipulation of the independent variables.

   For example:

   > *A researcher may wish to investigate the relationship between the socioeconomic background of a leaner and their academic*

performance. A questionnaire may be given to students to assess their socioeconomic background, and then compare it with their grades at the end of academic year. The researcher will discover whether these two factors are positively correlated, negatively correlated, or have zero correlation at the end of the research.

iii. **Observational Research Design** - Observational or descriptive survey research design focuses on observing the behavior of a research subject in a natural or laboratory setting. It is classified as non-experimental because it does not involve manipulation of independent variables.

For example:

In an investigation of the crowd effect or psychology in a particular group of people, imagine a situation where there are 2 ATM's at a place, and only one of the ATMs is filled with a queue, while the other is abandoned.

The crowd effect infers that majority of newcomers will also abandon the other ATMs. You will notice that each of these non-experimental studies is descriptive in nature. It then suffices to say that descriptive research is an example of non-experimental research design.

Mixed Methods Research Designs

The word "design" has two distinct meanings in the context of mixed methods research. One meaning focuses on the process of design. In this meaning, design is often used as a verb. Someone can be engaged in *designing* a study. Another meaning is that of a product, namely the result of designing. The result of *designing* as a verb is a mixed methods *design* as a noun or "Design". In mixed methods research design, both meanings are relevant.

To obtain a strong design as a product, one needs to carefully consider a number of rules for designing as an activity. Obeying these rules is not a guarantee of a strong design, but it does contribute to it. A mixed methods research design is characterized by the combination of at least one qualitative and one quantitative research component and can therefore be defined as;

"*A research design in which the researcher(s) combines elements of qualitative and quantitative research approaches e.g. use of qualitative and quantitative viewpoints, data collection, analysis*

and inference techniques, all for purposes of in-depth understanding and corroboration"

A mixed methods design as a product has several primary and secondary characteristics that should be considered during the design process as shown in the Table that follows;

Table 5: Primary and Secondary Design Dimensions in Mixed Methods Research

Primary Dimensions	**Secondary Dimensions**
○ Purpose	○ Phenomenon
○ Theoretical drive	○ Social scientific theory
○ Timing (simultaneity and dependence)	○ Ideological drive
○ Point of integration	○ Combination of sampling methods
○ Typological vs. interactive design approach	○ Degree to which the research participants will be similar or different
○ Planned vs. emergent design	○ Degree to which the researchers on the research team will be similar or different
○ Complexity	○ Type of implementation setting

On the basis of these dimensions, mixed methods designs can be classified into a mixed methods typology or taxonomy. In the mixed methods literature, various typologies of mixed methods designs have been proposed. First is that of Creswell and Plano (2011), which includes six major mixed methods research designs. A summary of these designs is as listed hereunder;

(i) **Convergent Parallel Design** - the quantitative and qualitative strands of the research are performed independently, and their results are brought together in the overall interpretation,

(ii) **Explanatory Sequential Design** - a first phase of quantitative data collection and analysis is followed by the collection of qualitative data, which are used to explain the initial quantitative results,

(iii) **Exploratory Sequential Design** - the first phase of qualitative data collection and analysis is followed by the collection of quantitative data to test or generalize the initial qualitative results,

(iv) **Embedded Design** - in a traditional qualitative or quantitative design, a strand of the other type is added to enhance the overall design,

(v) **Transformative Design** - a transformative theoretical framework, e. g. feminism or critical race theory, shapes the interaction, priority, timing and mixing of the qualitative and quantitative strand,

(vi) **Multi-phase design** - more than two phases or both sequential and concurrent strands are combined over a period of time.

It should be noted at this juncture that most of these designs presuppose a specific juxtaposition of qualitative and quantitative components. The last design for instance is a complex type that is required in many mixed methods studies, especially when more than one sampling technique has been used to select participants of a given study. Secondly**,** the following are the adapted definitions of Teddlie and Tashakkori's (2009) five sets of mixed methods research designs;

(i) **Parallel Mixed Designs** - In these designs, one has two or more parallel quantitative and qualitative strands, either with some minimal time lapse or simultaneously. The strand results are integrated into meta-inferences after separate analyses are conducted. Related qualitative analysis and quantitative research questions are answered or aspects of the same mixed research question are addressed.

(ii) **Sequential Mixed Designs** - In these designs, qualitative and quantitative strands occur across chronological phases, and the procedures/questions from the later strand emerge from the previous strand. The research questions are interrelated and may evolve during the study.

(iii) **Conversion Mixed Designs** - In these parallel designs, mixing occurs when one type of data is transformed to the other type and then analyzed, and the additional findings are added to the results. This design answers related aspects of the same research question.

(iv) **Multilevel Mixed Designs** - In these parallel or sequential designs, mixing occurs across multiple levels of analysis, as quantitative and qualitative data are analyzed and integrated to answer related aspects of the same research question or related questions.

(v) **Fully Integrated Mixed Designs** - In these designs, mixing occurs in an interactive manner at all stages of the study. At each stage, one approach affects the formulation of the other and multiple types of implementation processes can occur in the process. For example, rather than including integration only at the findings/results stage, or only across phases in a sequential design, mixing might occur at the conceptualization, methodological, analysis and the inferential stages.

For research in education, the author of this book recommends adding to Teddies and Tashakkori's typology a sixth design type, called *equihybrid* mixed methods design, where any two mixed methods designs are amalgamated in equal proportions. This hybrid design demands that for every research objective formulated, the researcher must also develop one research question and one null hypothesis. This design will have the powerful advantage of enabling an elaborate summary of research findings and attracts researchers who are enthusiastic in both qualitative and quantitative research in equal measure.

Validity of Research Designs

Any research design has its own weaknesses. One weakness of a research design that should be of a researcher's concern is the threats to its external and internal validity.

Internal Validity is the extent to which you can be confident that a cause-and-effect relationship established in a study cannot be explained by other factors, other than the independent variable. In other words, can you reasonably draw a causal link between your treatment and the observation response in an experiment? If the answer is yes, then your research design has a high internal validity, otherwise is has low internal validity. Internal validity makes the conclusions of a causal relationship credible and trustworthy. Without high internal validity, an experiment cannot demonstrate a causal link between two variables.

Example:

> *You want to test the hypothesis that drinking a cup of coffee improves students' memory. You therefore schedule an equal number of college-aged participants for morning and evening sessions at the school science laboratory. For convenience, you assign all morning session participants to the treatment group and all evening session participants to the control group. Once they arrive at the laboratory, the treatment group participants are given a cup of coffee to drink, while control group participants are given water. You also give both groups memory tests.*

After analyzing the results, you find that the treatment group performed better than the control group on the memory test. Can you conclude that drinking a cup of coffee improves memory performance? For your conclusion to be valid, you need to be able to rule out other explanations for the results.

There are three necessary conditions for internal validity, which must all occur to experimentally establish causality between an independent variable A (your treatment variable) and dependent variable B (your response/outcome variable). The three conditions are;

- ✓ Your treatment and response variables change together.
- ✓ Your treatment precedes changes in your response variables
- ✓ No confounding or extraneous factors can explain the results of your study.

In the previous research example, only two out of the three conditions have been met i.e.

- ✓ Drinking coffee and memory performance increased together.
- ✓ Drinking coffee happened before the memory test.
- ✗ Time of day of the sessions is an extraneous factor that can also explain observed outcome of the investigation.

Since you assigned participants to groups based on the school time schedule, the groups were different at the start of the study. Any differences in memory performance may be due to a difference in the time of day. Therefore, you cannot pinpoint with certainty which one between the time of day and drinking a cup of coffee improved memory performance. That implies your study has low internal validity, and you cannot deduce a causal relationship between drinking coffee and memory performance.

Threats to the Internal Validity of a Research Design

The following factors if not taken care of may pose a serious threat to the internal validity of a research design, thereby jeopardizing the whole research process;

- **History** - these are the specific events that occur between the first and second measurements. Some events like natural disasters or political activities may influence the coordination of research participants and how they may perform in the variable being measured.
- **Selection Bias -** these are prejudices which may result involving selection of comparison groups. Random assignment of respondents into a group is a counter-measure against this threat. However, when the sample size is small, randomization may lead to *Simpson Paradox,* a statistical phenomenon where an association between two variables in a population emerges, disappears or reverses when the population is divided into sub-populations.
- **Maturation** - these are processes within subjects, which act as a function of the passage of time i.e. if the research lasts too long, most participants may naturally improve in their performance (dependent variable) regardless of treatment. This means that time should be considered an important variable in experimental research designs. If participants become older or go through a biological (time) change in the course of your research, it may be difficult to prove that the outcome of the study was not influenced by time.
- **Selection and Maturation Interaction -** this is a situation where selection of comparison groups and maturation interact, which may lead to confounding outcomes, and erroneous interpretation that the treatment caused the effect.
- **Pre-Testing** - these are the effects of taking a test before treatment, which affects the outcome of a second test.
- **Instrument Bias** - these are the changes in the data collection instrument which may in turn produce changes in outcomes e.g. one group is given colored copies of a questionnaire, the other black-and-white.
- **Researcher's Bias -** this is when the researcher behaves differently to one group. (This can be in favor of one group and harshly against the other).

- **Statistical Regression** - this is also known as regression to the mean. This threat is caused by selection of subjects on the basis of extreme scores or characteristics.
- **Experimental Mortality** - this is also called attrition. It is the loss of subjects with time due to natural causes. For example, you start a study with 161 subjects and only 95 of them complete it. Those who stayed in the project all the way may be more motivated to learn and thus achieve higher performance on the dependent variable.
- **John Henry Effect** - this is the tendency of a group of persons being investigated to work extra hard once they know they are being compared with another group. John Henry was a worker who tried to outperform a machine under an experimental setting because he was aware that his performance was being compared to that of a machine. Sadly, he died in the process!
- **Repeated Testing** - giving a test to your research participants multiple times may influence research findings. It is only natural that when you give a particular test to someone repeatedly, they'll become accustomed to the test.

Countering the Threats to Internal Validity of a Research Design

The following are some of the measures a researcher should put in place where applicable, in order to ensure the aforementioned threats do not adversely affect the internal validity of a research design used in educational research. Manipulating an experimental design can alter or reduce many threats to internal validity. Here are some ways to achieve this;

- ✓ Include a comparable control group to antagonize all the threats from the treatment group. This counters all threats to single-group studies. If the treatment group and the control group of your study are met with similar threats, the threats should not affect the results of your research.
- ✓ Use a large sample size to eliminate threats in your test. This would ensure your research findings would be sensitive to other variables that might want to appear in the test.
- ✓ You can also eliminate threats in your tests by employing filler tasks in your research. This is because filler tasks will hide the true purpose of your research.
- ✓ Randomly assign participants to sample groups to eliminate regression and selection bias. This will make the groups comparable before you start your test.

- ✓ Make your participants blind to the purpose of your research. This will discourage interaction between your study groups.

External Validity is the extent to which you can generalize the findings of a study to other situations, people, settings or measures. In other words, can you apply the findings of your study to the target population? If the answer is yes, then the research design used has high external validity, otherwise it has low external validity.

The aim of scientific research is to produce generalizable knowledge about the real world. Without high external validity, you cannot apply results from an experimental study to the target population or from a laboratory investigation to the real world. In qualitative studies, external validity is referred to as transferability.

It is interesting to note that there is an inherent trade-off between external and internal validity in the sense that the more applicable you make your study to a broader context, the less you can control for extraneous factors in your study.

Example*:*

> *You want to test the hypothesis that driving reaction times become slower when people pay attention to others talking. In a laboratory setting, you set up a simple computer-based task to measure reaction times. Participants are told to imagine themselves driving around a busy road, say the Thika superhighway in Nairobi, then double-click the mouse whenever they see an orange cat on the screen. For one round, participants listen to a podcast. In the other round, they do not need to listen to anything. After assessing the results, you find that reaction times are much slower when listening to the podcast.*

From this scenario, can you conclude that driving reaction times are slower when people listen to others talking? In this example, it is difficult to generalize such findings to real-life driving conditions. A computer-based task using a mouse does not resemble real-life driving conditions with a steering wheel. Additionally, a static image of an orange cat may not represent common real-life hurdles when driving.

To improve external validity in a laboratory setting, you could use an immersive driving simulator with a steering wheel and foot pedal instead of a computer and mouse. This increases psychological

realism by more closely mirroring the experience of driving in the real world. Alternatively, for higher external validity, you could conduct the experiment using a real driving course.

Threats to External Validity of a Research Design

It is important to identify and counter most threats to external validity of a research design for any plausible findings in educational research. A number of them are mentioned in this section;

- *Sampling Bias* - this is where the sample is not representative of the population e.g. the sample includes only respondents of a certain gender.
- *History* - this is where an unrelated event occurs and influences the outcomes for instance. right before the pre-test, a natural event occurs e.g. bereavement. As a result, pre-test scores are lower due to psychological trauma emanating from the bereavement.
- *Experimenter Effect* - this is where the characteristics or behaviors of the experimenter(s) unintentionally influence the outcomes
- *Hawthorne Effect* - this is the tendency of participants to change their behaviors simply because they know they are being studied e.g. the participants actively avoid anxiety-inducing situations for the period of the study because they are conscious of their participation in the research.
- *Testing Effect* - this is where the administration of a pre- or post-test affects the outcomes e.g. because participants become familiar with the pre-test format and questions, they are less anxious during the post-test and exhibit less anxiety then.
- *Aptitude Treatment* - this is where interactions between characteristics of the group and individual variables together influence the dependent variable.
- *Situation Effect* - this is where factors like the setting, time of day, location, researchers' characteristics, etc. limit generalizability of the findings.

Countering Threats to External Validity of a Research Design

There are several ways of countering some of the threats to external validity of a research design in educational research;

- Replications counter almost all threats by enhancing generalizability to other settings, populations and conditions.
- Field experiments counter testing and situation effects by using natural contexts.

- Random sampling counters selection bias by making sure everyone in a population has an equal chance of being selected for a study sample.
- Re-calibration or reprocessing also counters selection bias using algorithms to correct weighting of factors (e.g., age) within study samples.

CHAPTER 4:

SAMPLE SIZE AND SAMPLING IN EDUCATIONAL RESEARCH

In most research circumstances, it may not be possible to collect data from all persons in the target population, which leaves sampling as the only viable option. This chapter talks about target population in detail, sampling, reasons for sampling, commonly used sampling techniques in educational research, guidelines on how to choose the right sampling method, sampling frame, sample space, determination of sample size in educational research and ethical issues surrounding sampling and research in education.

Target Population

The target population of a study is the entire set of units, subjects or individuals for which the research findings are to be used to make inferences and generalizations. Thus, the target population defines those units for which findings of the investigation are meant to generalize. For this reason, target population of a study must be clearly defined, as the definition has ramifications on whether sampled cases are eligible or ineligible to be potential respondents. The geographic and temporary characteristics of the target population need to be delineated, as well as types of units being included. In some cases, the target population is restricted to exclude population members that are difficult or impossible to interview e.g., children, incapacitated/disabled persons etc.

While describing target population of a study, relevant demographic attributes and exact number of individuals therein should be clearly stated, as this would make sampling procedures easy to plan and execute. Relevant demographics of the target population in educational research can be obtained from relevant government offices e.g. Kenya National Bureau of Statistics, County Director of Education (for school enrolment, dropout rates, number of schools etc.), Sub-County Quality Assurance and Standards office (number of teachers, school infrastructure etc.), Teachers Service Commission (staffing issues), Kenya Institute of Curriculum Development (teaching resources etc.) and Kenya National Examinations Council (academic performance in KCSE and KCPE). It is also important for researchers to quote the source of demographic statistics mentioned in their work for the sake of credibility and authenticity purposes.

If for instance, a researcher wants to identify approximately how many parents read a particular article in their child's school newsletter, the target population would be all parents of children at that school. The target units would then be the individual parents, and the school could provide a list of parents' contact details, which would serve as the sampling frame. Other examples of a target population are; a company's customer base for example all Safaricom M-PESA subscribers, the total number of registered voters in a particular county ward, total number of female students studying Engineering in a particular university, total number of Gor Mahia football club fans in Kenya etc.

After target population of a study has been clearly identified and defined, a sampling frame is then constructed from this target population. Note that if the target population units are not clearly stated, generalizing findings to the entire population can be misleading, rendering the whole research useless.

What is Sampling?

Sampling is the process of selecting a subset of the target population who will be used as respondents to give information that the researcher uses to address the research objectives. Responses given by the sampled individuals are then used to make statistical inferences that estimate characteristics of the whole population under investigation.

Reasons for Sampling in Educational Research

There are several sampling methods or techniques that are widely used by researchers in education, so that they do not need to research the entire population to collect actionable insights. The following are some of the reasons why sampling is done by educational researchers;

- ✓ It saves time
- ✓ It is a cost-effective method and hence forms the basis of any research design.
- ✓ It is practically impossible to reach out all respondents
- ✓ It may be meaningless to destroy many valuable items that are identical

Commonly Used Sampling Methods in Educational Research

Sampling methods or techniques in educational research lie in two broad categories; probability and non-probability sampling. Let's now look at each category in detail.

Probability Sampling Methods

These are sampling techniques where a researcher sets selection criteria, and uses it to choose members of a population randomly. All the members therein have an equal opportunity to be a part of the sample in any probability sampling technique. There are three major uses of probability sampling;

- Reducing sample bias – as it was earlier discussed in chapter three of this book, sampling bias is one of the possible threats to validity of a research design. By using probability sampling methods, the bias in the sample derived from a population is negligible if not non-existent. Probability sampling leads to higher quality data collection as the sample appropriately represents the population.
- Obtaining a diverse population - When the population is vast and diverse, it is essential to have adequate representation so that the data is not skewed towards one demographic attribute.
- Creating an accurate sample - Probability sampling techniques help the researchers plan and create an accurate sample, which consequently helps to obtain well-defined data.

Probability sampling methods consider every member of the target population and forms samples based on a fixed objective process. For example, in a population of 100 members, every member will have a 1% chance of being selected to be a part of a sample. Probability sampling eliminates bias in the population and gives all members a fair chance to be included in the sample. There are four types of probability sampling methods;

i. **Simple Random Sampling** - This is one of the best probability sampling techniques that help in saving time and resources. It is also a very objective and reliable method of obtaining information where every single member of a population is chosen in a purely random manner, and those that are selected find themselves in the sample merely by chance. Each individual has the same probability of being chosen to be a part of a sample.

Simple random sampling can be implemented using various techniques e.g.;

- *Balloting* - involves using numbered, folded papers in an open container like bowl. A blindfolded person is then asked to pick one piece of paper from the container. The paper is then unfolded to reveal the "lucky" participant. It is suitable for populations not exceeding 100 in number.
- *Dice tossing* – involves using a fair die to select respondents. This technique is applicable for selecting one respondent from a population not exceeding 6 members,
- *Coin tossing* – involves using a fair coin. It is used to select one subject from a possible two
- *Raffling* - this is where solitaire cards used to select respondents from a population not exceeding 52 individuals,
- *Using technology* - involves using programmable machines like *lotto,* suitable for medium to large populations.it is a considered the digital equivalent of balloting, as it is much faster.
- *Using random number generators* - involves online software that can be instructed to generate random numbers depending on the algorithm issued to it. An example of an online number generator is available at www.random.org. This technique is appropriate for extremely large populations, running into tens of thousands of people.

The following pictures show examples of tools that can be used to execute simple random sampling:

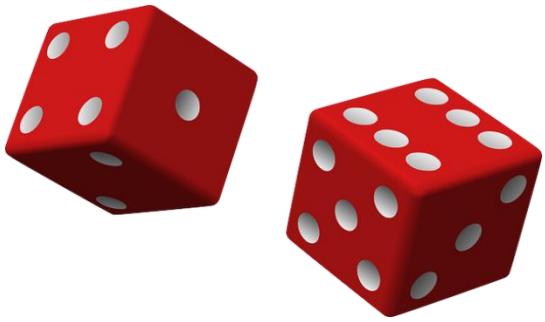


Figure 7: Fair dice – Used for Selecting one from Six Individuals/Groups



Figure 8: Fair coins - Used to Select one from Two Individuals/Groups



Figure 9: Solitaire Cards used to Select Several from about Fifty

ii. **Cluster Sampling** - This is a type of probability sampling of respondents where the researcher divides the target population into geographical sections called *clusters*, from which a representative region is selected. This method is often used to study large populations, particularly those that are geographically dispersed. Researchers use preexisting units such as schools in a zone, municipalities in a city, sub-counties in a county, wards in a constituency, deaneries in a parish etc. as clusters.

This makes it quite simple for researchers to make inferences from the results for example, if in western Kenya, an opinion pollster wishes to determine the number of supporters of a given political party, they can divide it into clusters based on the counties therein i.e., Kakamega, Busia, Bungoma and Vihiga. This way of conducting a survey will be quite effective, as results will be organized into counties and provide insightful data on the popularity of a given political party or presidential candidate.

The simplest form of cluster sampling is single-stage cluster sampling, which involves 4 key steps.

Step 1: Define your population - As earlier mentioned it is important to clearly delineate the characteristics of the population you are interested in collecting data from.

Step 2: Divide your sample into clusters - This is the most important part of this sampling technique. The quality of your clusters and how well they will represent the target population determines the validity of your results. Ideally, you would like your clusters to be as diverse as possible, in order to ensure every characteristic of the entire population is represented in each cluster. Each cluster should therefore be a mini-representation of the entire population. Typical clusters do not often represent the population's characteristics perfectly. This explains why cluster sampling has less statistical robustness as compared to simple random sampling, more so because clusters are usually naturally occurring groups of individuals, such as schools, cities or households, all of which are often more homogeneous than the population as a whole. You should be aware of this when executing your study, as it might affect its validity.

Step 3: Randomly select clusters to use as your sample - If each cluster is itself a mini representation of the larger population, randomly selecting and sampling from the clusters allows you to imitate simple random sampling, which in turn supports the validity of your results.

Conversely, if the clusters are not representative, then random sampling will allow you to gather data on a diverse array of clusters, which should still provide you with an overview of the population as a whole.

Step 4: Collect data from the sample - You then conduct your study and collect data from every unit in the selected clusters.

iii. **Systematic Sampling -** This is a probability sampling method in which sample members of a population are selected at regular intervals. It requires random selection of a starting point for the sample and sample size that can be repeated at regular intervals. This interval is called the sampling interval, and is calculated by dividing the population size by the desired sample size. This type of sampling has a predefined range, and hence this sampling technique is the least time-consuming. For example, a researcher intends to collect a systematic sample of 200 people in a population of 2000. He or she numbers each person in the population from 1 to 2000 and then chooses every 10^{th} individual to be a part of the sample (Total population/ Sample Size = 2000/200 = 10).

These could be accessed from one vantage location during peak hours e.g. at the exit of a parking lot, waiting bay of a hospital, gate of a university just before lecture hours begin etc. Despite the sample population for systematic sampling being selected in advance, this method of sampling is still thought of as being random as long as the periodic interval is determined beforehand and the starting point is random.

There are six steps involved in selecting a systematic random sample;

Step 1: Develop a defined structural audience to start working on the sampling aspect
Step 2: As a researcher, figure out the desired size of the sample
Step 3: Assign a number to every member of the sample
Step 4: Define the standard distance between the elements (interval).
Step 5: Select the members who fit the criteria i.e. 1 in every N individuals.

Step 6: Randomly pick the starting member of the sample and add the interval to the random number to keep adding members in the sample, who will form the sample.

iv. **Stratified Sampling** - Stratified random sampling is a method in which the researcher divides the population into smaller subgroups called *strata*. This grouping is based on shared characteristics of the members and then randomly selecting among these groups to form the final sample. While sampling, these groups can be organized then a sample drawn from each group separately. For example, a researcher looking to analyze the characteristics of students belonging to different types of schools will create strata according to the school types in the research location e.g. mixed schools, boys' schools and girls' schools or national, extra-county, county and sub-county schools. By doing this, the researcher concludes the characteristics of students belonging to different school types. The researchers can analyze which school types to target and which ones to eliminate to create a road-map that would bear desirable results.

The process of classifying the population into groups before sampling is called *stratification*. The strata must be mutually exclusive and all members of the population must only belong to one stratum. When stratifying, researchers tend to use *proportionate sampling,* where they maintain the correct proportions to represent the population as a whole. For example, if the larger population contains 70% day schools and 30% boarding schools, the final sample should reflect these percentages e.g. 7 day schools and 3 boarding schools. *Disproportionate sampling* is typically used only when studying an under-represented group. Stratified random sampling can be executed in four steps;

Step 1: Define your population of interest and choose the characteristic(s) that you will use to divide your groups.

Step 2: Divide your sample into strata depending on the relevant characteristic(s). Each stratum must be mutually exclusive, but together, they must represent the entire population.

Step 3: Define the sample size for each stratum and decide whether your sample will be proportionate or disproportionate. The size of the sample in each stratum should ideally be in proportion to the members of that group within the target population or sampling frame.

Step 4: Draw a random sample from each stratum and combine them to form your final sample.

Non-Probability Sampling Methods

In non-probability sampling, the researcher does not select respondents for the study at random, which means some have a greater chance to be selected than others while others have no chance at all. These sampling methods are not fixed or predefined selection processes, which makes it difficult for all members of a population to have equal opportunities to be included in the final list of selected respondents. Non-probability sampling methods generally involve collection of feedback based on a researcher or statistician's sample selection preferences and capabilities hence are subjective and non-fixed processes. In most situations, the output of a survey conducted with a non-probable sample leads to skewed results, which may not represent the desired target population.

However, there are situations such as the preliminary stages of research or financial constraints which render non-probability sampling methods more practical than the probability methods. The non-probability methods are preferred when there are budget and time constraints in conducting the research but nevertheless, some preliminary data must be collected. Since the surveys and other qualitative research designs are not rigid, researchers using this design find it necessary to pick respondents in the easiest, simplest and cheapest way then have them fill the survey questionnaire.

There are four types of non-probability sampling techniques;

i. **Convenience Sampling** - This method is dependent on the ease of access to subjects such as surveying customers at a mall or passers-by on a busy street. It is usually termed as convenience sampling, because of the researcher's ease of carrying it out and getting in touch with the subjects. Researchers have nearly no authority to select the sample elements, and it's purely done based on proximity and not representation ability.

This non-probability sampling method is used when there is time or cost limitations in collecting feedback. In situations where there are resource limitations such as the initial stages of research, convenience sampling is used. For example, product manufacturers usually conduct convenience sampling at a mall to distribute leaflets of a newly launched product. They do that by standing at the mall's entrance and give out pamphlets, leaflets or brochures randomly.

ii. **Purposive Sampling** - Also called *judgmental sampling*, purposive sampling is where samples are formed by the discretion of the researcher. Researchers purely consider the purpose of the study, along with the understanding of the target population. For instance, when researchers want to understand the level of pre-service teachers' preparation in technological and content knowledge. The researcher will first select pre-service teachers in the target university randomly, then purposively select the lecturers of the selected pre-service teachers to get data that will supplement what the pre-service teachers would have said.

iii. **Snowball Sampling** - Snowball sampling is a sampling method that researchers apply when the subjects are difficult to trace. For example, it will be extremely challenging to survey shelter-less people, illegal immigrants, HIV positive people, bang smokers or ex-convicts. In such cases, using the snowball theory, researchers can track a few categories to interview and derive results.

    Researchers also implement this sampling method in situations where the topic is highly sensitive and not openly discussed, for example, surveys to gather information about prostitution. Not many prostitutes will readily respond to the questions. Still, researchers can contact people they might know or volunteers associated with the cause to get in touch with the victims and collect information. The first respondent(s) can thereafter link you up with their friends, making the sample to keep growing.

iv. **Quota Sampling** - This is a type of non-probability sampling technique where researchers assign quotas to a group of people in order to create subgroups of individuals that represent characteristics of the target population as a whole. Examples of these characteristics include; gender, age, tribe, education level and job group. Once the

subgroups are formed, the researchers will use their own discretion to select the subjects from each segment to produce the final sample. It is important for researchers to maintain the correct proportions to represent the population. For example, if the target population comprises of 65% low-income earners, 30% middle-income earners and 5% high-income earners, the final sample should reflect these percentages.

Quota sampling can be executed in 5 steps;

Step 1: Divide sample into subgroups depending on any relevant characteristics e.g. you could divide a university population by job group.

Step 2: Evaluate the proportions of the subgroups to determine the specific elements that will be chosen from each quota. For example, professors might be ⅕ of the population, senor lectures ¼ etc.

Step 3: Select a sample size. For example, if you are sampling 800 dons, your quota sample might be 80 or any other figure.

Step 4: Choose your participants, adhering to the subgroups characteristics. For example, 20% of your sample should be lecturers.

Step 5: Continue with the selection process until your quotas are filled.

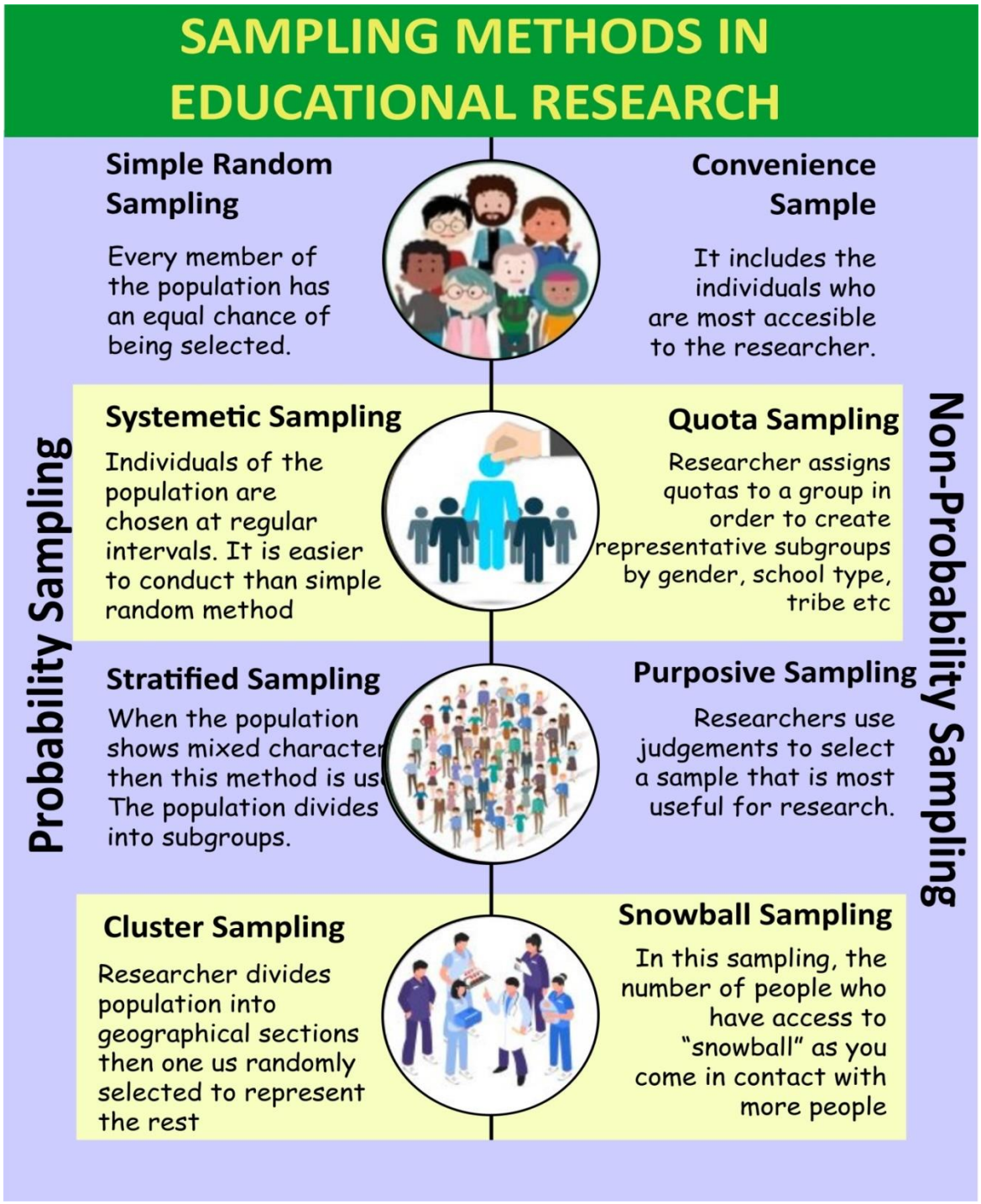


Figure 10: Summary of Sampling Methods

Choosing the Right Sampling Method

For any research in education, it is essential to choose a sampling method that is the most appropriate for meeting objectives of your study. The appropriateness of your sampling method relies on various factors. The following are some useful steps that researchers can follow to guide them in deciding the best sampling method(s) or procedures;

i. Write down the research objectives. They must be SMART and informed by the purpose of the study
ii. Define the target population clearly. Specify all demographic attributes e.g., "all form three female students in mixed public secondary schools in Likuyani sub-county who take computer studies".
iii. Identify the sampling techniques that might effectively achieve the research objectives.
iv. Test each of these methods and examine whether they help in achieving your objectives.
v. Select a method that is practical and works best for the research if there are several options.
vi. Select your respondents using clear and reliable criteria.

The Research Sampling Frame

A sampling frame is the list of all the items or subjects in your target population. It is the complete list of everyone or everything you want to study. The difference between a population and a sampling frame is that the population is general and the frame is specific. For example, the population could be "People who live in Coastal Kenya." The sampling frame would include names all residents of Mombasa, Kilifi, Taita-Taveta, Tana River and Lamu counties. A sampling frame can be a list of just about anything.

Other examples include;

- List of all public universities in Kenya
- List of all extra-county secondary schools in West Pokot county
- List of all supermarkets in Nairobi CBD
- List of all first year students undertaking bachelor of Education Science in Murang'a University of Technology

Qualities of a Good Sampling Frame for Educational Research

For clear-cut sampling procedures, one cannot to just use any list they come across. Careful consideration must be made to ensure the sampling frame has accurate facts and figures, which would enable one get sufficient respondents. A good sampling frame should therefore have the following five qualities;

i. Must include all individuals in the target population
ii. Must exclude all individuals not matching all the specified attributes of the target population
iii. Must include accurate information that can be used to identify the selected individuals e.g. registration numbers of students and their names, since some might have identical names. Avoid using acronyms, since some could be confusing e.g. UOE for University of Eldoret. Don't forget University of Embu has the same acronym!
iv. A unique identifier for each member. This could be a serial number (e.g. from 1 to 1000). Check to make sure there are no duplicates in the frame
v. Logical organization to the list. For example, put them in alphabetical or chronological order.
vi. Up to date information. This may need to be periodically updated due to natural attrition.

It should be noted that in some cases, it might be impossible or very difficult to get a sampling frame. For example, getting a list of all prostitutes in a isn't likely, for the obvious reason that most prostitutes won't want to be identified, none are registered, some operate on part-time basis while some reform along the way and might fear victimization. This is where creativity and use of techniques like snowball sampling must be used to make up for the lack of sampling frame.

Sample Space

While a sampling frame is a list of things that a researcher draws their sample from, a sample space is a list of all possible outcomes for a probability sampling exercise. For example, you might have a sampling frame of names of people in a certain town for a survey you're going to be conducting on family size. The sample space is all possible outcomes from your survey; 4 person, 7 people, 10 people… 20 or more.

In order to answer a research question that seeks to determine how many hours of sleep a random sample of college students have per night, one way of ensuring a diverse sample is to ask each

selected student to record the number of hours of sleep they had the previous night. In this case, if we let *h* denote the number of hours slept, we would write the sample space as:

$$S = \{h : h \geq 0 \text{ hours}\}$$

Another example of defining sample space is, if we were interested in learning about students who passed a university common unit, say UCU 100, when trying to answer a certain research question, we might ask all current non-fresh students "what grade did you obtain in UCU 100?" In that case, we would write our sample space as:

$$S = \{A, B, C, D, E\}$$

Determination of Sample Size in Educational Research

In educational research, there is no standard sample size. This is because the number of respondents needed depends on the research design, nature of research objectives and the resources available to execute the study. Either way, a large sample size has the advantage of yielding very precise findings but has the disadvantage of high variation in the responses obtained.

Selecting the Required Sample in Qualitative Research

Everyone wants to know the magical number for sample size when doing qualitative research. Of course, when you're doing quantitative research, numbers matter a lot, especially when aiming at making inferences about the target population. In qualitative research however, the sample size doesn't have to be representative. The crucial thing is getting the right individuals to give sufficient data for answering the research questions, however few the individuals may be! For this to happen however, several factors must be considered. Here are some important points that qualitative researchers should consider when thinking about sample size;

- **Quality over Quantity:** Whether your study is targeting secondary school students from a given research location, or teenagers from a given locality who use Facebook, you are most likely looking for insights into your research problem. To this end, the first concern when conducting qualitative research should be to ensure the right individuals are recruited into the sample. The right respondents for the study are those who are most knowledgeable about the topic under investigation, regardless of their number.

- **Research Design:** The type of qualitative study is another important factor to consider when choosing sample size. There are various methods that can be used to gather insightful data, but not all may be applicable to your study and its research purpose. In-depth interviews, focus groups and ethnographies are the most common methods used in qualitative research, each method being unique in the information it can provide and the setting it can be used in. Moreover, the types of questions being studied have an equally important role to play in deciding sample size. Thus before choosing a sample size, make sure all the parameters, aims, and controls are clearly outlined.
- **Principle of saturation:** A sample size should be large enough to sufficiently describe the phenomenon of interest, and address the research question at hand. But at the same time, a large sample size risks having repetitive data. The goal of qualitative research should thus be the attainment of *saturation*. Saturation occurs when adding more participants to the study does not result in obtaining any additional information. Any qualitative study should endeavour to have a sample size large enough to uncover a variety of opinions, but to limit the sample size at the point of saturation.

The big question is "*is there a magical number for sample size in qualitative research?*"
Considering all the three factors mentioned above, it is recommended is that for in-depth interviews, a sample size of between 10 (for homogenous populations less than 100 people) to 25 (for homogenous populations greater than 100 people) respondents is adequate, provided the recruitment process is rigorous and informed by the three factors earlier mentioned, especially saturation.

For focus group discussions, it is recommended a total of 4 focus groups (for studies with populations less than 100) and 8 focus groups (for populations exceeding 100) provided the groups are independent, equal sized and heterogeneous.

Selecting the Required Sample in Quantitative Research

Credibility of findings in quantitative research heavily relies on whether or not the sample size used was representative of the target population. Accurate and scientific determination of the required sample size is one of the most important factors in inferential statistics, which are part and parcel of quantitative research. If the sample size is too small, it may not yield valid results,

may not adequately represent the realities of the population being studied or may lead to statistical errors. On the other hand, while larger sample sizes yield smaller margins of error and are more representative, a sample size that is too large may significantly increase the cost and time taken to conduct the research. To strike a balance between the two factors, the following considerations must be put in place in order to determine the required sample size;

- **Confidence Interval (Margin of Error):** Confidence intervals measure the degree of uncertainty or certainty in a sampling method and how much uncertainty there is with any particular statistic. In simple terms, the confidence interval tells you how confident you can be that the results from a study reflect what you would expect to find if it were possible to survey the entire population being studied. The confidence interval is usually a plus or minus (±) figure. For example, if your confidence interval is 6 and 60% percent of your sample picks an answer, you can be confident that if you had asked the entire population, between 54% (60-6) and 66% (60+6) would have picked that answer.
- **Confidence Level:** The confidence level refers to the percentage of probability, or certainty that the confidence interval would contain the true population parameter when you draw a random sample many times. It is expressed as a percentage and represents how often the percentage of the population who would pick an answer lies within the confidence interval. For example, a 99% confidence level means that should you repeat an experiment or survey over and over again, 99 percent of the time, your results will match the results you get from a population. The larger your sample size, the more confident you can be that their answers truly reflect the population i.e. the larger your sample size for a given confidence level, the smaller your confidence interval. In Educational research, 95% is the default confidence level.
- **Standard Deviation:** In calculating the sample size, the standard deviation is useful in estimating how much the responses you receive will vary from each other and from the mean number and the standard deviation of a sample can be used to approximate the standard deviation of a population. The higher the distribution or variability, the greater the standard deviation and the greater the magnitude of the deviation. For example, once you have already sent out your survey, how much variance do you expect in your responses? That variation in responses is the standard deviation.

- **Population Size:** The other important consideration to make when determining your sample size is the size of the entire population you want to study. A population is the entire group that you want to draw conclusions about. It is from the population that a sample is selected, using probability or non-probability samples. The population size may be known (such as the total number of professors in a given university), or unknown (such as the number of students in a college who smoke bhang), but there's a need for a close estimate, especially when dealing with a relatively small or easy-to-measure groups of people.

  For quantitative research in education, the following formula by Krejcie and Morgan (1970) is recommended, as it encompasses all the parameters earlier mentioned, hence meeting the all-important issue of being representative:

$$\mathbf{S} = \frac{\boldsymbol{x^2 NP(1-P)}}{\boldsymbol{d^2(N-1)+x^2P(1-P)}}$$

  Where: S = required sample size
  X = z-value, 1.96 at 0.05 alpha
  N = population size
  P = population proportion (assumed to be 0.5)
  d = degree of accuracy expressed as a proportion (0.05)

To calculate the required number of respondents, we therefore need to; determine the population size (if known), determine the confidence interval, determine the confidence level., determine the standard deviation (a standard deviation of 0.5 is a safe choice where the figure is unknown), convert the confidence level into a Z-Score (for educational research, the z-score for the 95% confidence level by default 1.96). Putting these figures into the sample size formula will yield the required sample size. Fortunately for those who are not good in mathematics, there is a conversion table (see appendix 4) that used this formula to derive sample size for a finite population.

Ethical Issues in Sampling and Research in Education

Ethical issues are simply the rules and regulations a researcher must adhere to in order to protect the rights of all individuals sampled to participate their study. Ethical issues can be specified as one of the most important parts of postgraduate research. Dissertations may even be cancelled and degree withdrawn if this part is missing!

The following ten points represent the most important principles related to ethical considerations in dissertations in order to prevent such disaster from happening to you;

i. Research participants should not be subjected to harm in any ways whatsoever.
ii. Respect for the dignity of research participants should be prioritized.
iii. Full consent should be obtained from the participants prior to the study. If the respondents are minors (below 18), written permission should be obtained from their parents or guardians. The permission should be in form of a signed consent form.
iv. Protection of the privacy of research participants has to be ensured.
v. Adequate level of confidentiality of the research data should be ensured.
vi. Anonymity of individuals and institutions they're affiliated to while participating in the research has to be ensured.
vii. Any deception, falsification or exaggeration about the objectives or findings of the research must be avoided.
viii. Affiliations in any forms, sources of funding, as well as any possible conflicts of interests have to be declared.
ix. Any type of communication in relation to the research should be done with honesty and transparency.
x. Any type of misleading information, as well as representation of primary data findings in a biased way must be avoided.

A well written ethical considerations section in your dissertation needs to expand discussions of each of the following points to at least one paragraph;

✧ Voluntary participation of respondents in the research is important. Moreover, participants have rights to withdraw from the study at any stage if they wish to do so.

- ✧ Respondents should participate on the basis of informed consent. The principle of informed consent involves researchers providing sufficient information and assurances about taking part to allow individuals to understand the implications of participation and to reach a fully informed, considered and freely given decision about whether or not to do so, without the exercise of any pressure or coercion.
- ✧ The use of offensive, discriminatory or other unacceptable language needs to be avoided in the research instruments.
- ✧ Privacy and anonymity or respondents is of a paramount importance.
- ✧ Acknowledgement of works of other authors used in any part of the dissertation with the use of APA referencing format.
- ✧ Maintenance of the highest level of objectivity in discussions and analyses throughout the research. This means you need to avoid irrelevant topics during interviews or discussions.
- ✧ Adherence to chapter 6 of the constitution of Kenya

Most universities have their own additional or complementary Codes of Ethical Practice e.g. seeking a research permit and letter of introduction, checking plagiarism using *turnitin* software and attaching the report in your research report etc. It is critically important for you to thoroughly adhere to this code in every aspect of your research and declare your adherence to ethical considerations as part of your thesis or dissertation. You may want to append scan copies of your research permit, informed consent form and plagiarism check report.

CHAPTER 5:

DATA COLLECTION IN EDUCATIONAL REEARCH

The underlying need for data collection is to capture credible evidence that seeks to sufficiently address all the hypotheses or questions that have been posed at the onset of the research. Through accurate data collection, researchers in education can deduce quality information that is a prerequisite for making informed policy decisions. To improve the research, it is important that data is collected, so that one can draw inferences and make informed decisions. This chapter gives the definition of data collection, highlights primary and secondary data, examines the importance of data collection then discusses the tools and instruments for data collection.

Definition of Data Collection

Data collection is the systematic process of gathering specific information to offer concrete answers to relevant questions/hypotheses. It focuses on finding out all there is to a particular research topic. Data is collected for the sake of further manipulation, so as to make sense of it. Hypothesis testing eliminates assumptions, while making a proposition from the basis of reason. The core forms in which data can be collected are primary and secondary data. While the former is collected by a researcher through first-hand sources, the latter is collected by an individual other than the user.

Primary Data

Primary data collection is the gathering of raw data directly from the source, though first-hand experience. Primary data collection is a direct approach that is tailored to specific research objectives. It can be a long process but does provide important first-hand information in many academic cases. Primary data is therefore original and unfalsifiable, as it is like raw material. Popular examples of primary data sources are; interviews, self-administered questionnaires, field observations, laboratory experiments, life histories, case studies, eyewitness accounts, personal narratives and memoirs. It suffices to say that the source of primary data is the population sample from which data is collected. The sample is selected by some of the different types of sampling methods and techniques discussed in the previous chapter of this book.

Advantages of Primary Data

i. **Is specific** - Performing own research allows you to address and resolve issues specific to your study variables. The collected information is the exact information that the researcher wants to know and he or she reports it in a way that benefits the specific situation in an educational institution or office.
ii. **Better accuracy** - Primary data is much more accurate because it is directly collected from a given population.
iii. **High level of control** - The researcher can control easily the research design and method. In addition, they have higher level of control over how the information is gathered.
iv. **Current** - The primary sources of data yield latest and up-to-date information as you collect it directly from the field in real-time. Usually, secondary data may not be so up-to-date and recent.
v. **You are the owner** - Information collected by the researcher is their own and is typically not shared with others. Thus, the information can remain hidden from other current and potential competitors.

Disadvantages of Primary Data

i. **More expensive** - It could be very expensive to obtain primary data collection methods because the marketer or the research team has to start from the beginning. It means they have to follow the whole study procedure, organizing materials, process and etc.
ii. **Time consuming** - It takes a lot of time to conduct research from the beginning to the end in comparison to the time needed to collect secondary data.
iii. **Has many limitations** - Primary data is limited to the specific time, place or number of participants and etc. To compare, secondary data can come from a variety of sources to give more details.
iv. **Not always possible** - some researches can sometimes be just too large to be performed by an individual.

Secondary Data

Secondary data collection on the other hand, is referred to as the gathering of second-hand data collected by an individual who is not the original user, from already existing sources. Examples of secondary data are figures from published books, journals, websites or online portals. Your choice between Primary data collection and secondary data collection depends on the nature, scope, and area of your research as well as its objectives.

Advantages of Secondary Data

i. **Cost effectiveness** - because someone else has already collected the data, the researcher does not need to invest any money, time, or effort into the data collection stages of his or her study. While sometimes secondary data must be purchased by a researcher looking to use it to inform a study they're working on, these costs are almost always lower than what the expenses would be if the researcher were to get the same data set from scratch.
ii. **Time saving** - the data from a secondary data set is typically already cleaned and stored in an electronic format, so the researcher can spend his or her time rolling up their sleeves and analyzing the data instead of spending time having to prepare the data for analysis.
iii. **Easily available** - leveraging the findings from studies that the government has conducted provides researchers with access to a volume of data that would have simply been impossible for the researcher to amass themselves. Longitudinal data at this scale is extremely powerful. The government could have been collecting data on a single population for long, extended periods of time e.g. school enrollment for the last 30 years. This information can easily be found at the Kenya Bureau of statistics or Ministry of planning, as they conduct census every 10 years and make the data available for any interested member of the public. Instead of investing that time, energy and resources, it is wise to use the government's publicly available data

Disadvantages of Secondary Data Analysis

i. **Relevance issues** - the biggest disadvantage of working with secondary data is that the it might not answer the researcher's specific research question to the degree that the research would have hoped. If a researcher sets out to perform a study with a very particular

question in mind, a secondary data set might not contain the precisely specific information that would allow the researcher to answer their research questions.

ii. **Validity issues** - when a researcher has a specific question or objective in mind, it can sometimes be difficult to identify secondary data that is valid for use, as the data might not have been collected during the time frame the researcher was hoping for, or in correct the desired geographical region etc.

iii. **Credibility issues** - no matter what a researcher does to clean a secondary data set, they will never be able to know exactly how the data was collected, and how well that process was executed. Without being the one who is actually developing instruments and distributing them to the appropriate populations, it's impossible to know the extent to which the researchers that collected the data went to ensure validity or quality, or if they experienced issues such as low response rates or respondents misunderstanding what a question was truly asking.

iv. **Lack of control** - since the researcher conducting the study did not collect the data he or she will be using, he or she ultimately has no control over what their secondary data set contains.

Importance of Data Collection

There are a bunch of underlying reasons for collecting data, especially for a researcher. Walking you through them, here are a few reasons;

(i) Integrity of the research - A key reason for collecting data, be it through quantitative or qualitative methods is to ensure that the integrity of the research question is indeed maintained.

(ii) Reduce the likelihood of errors - The correct use of appropriate data collection of methods reduces the likelihood of errors consistent with the results.

(iii) Decision making - To minimize the risk of errors in decision-making, it is important that accurate data is collected so that the researcher doesn't make uninformed decisions.

(iv) To support the need for a new idea, change, and/or innovation - To prove the need for a change in the norm or the introduction of new information that will be widely accepted, it is important to collect data as evidence to support these claims, like the old adage goes, "numbers don't lie".

Methods and Instruments for Data Collection

Data collection instruments refer to the devices or gadgets used to capture data, such as a paper questionnaire or computer-assisted interviewing software, checklists, interviews guides, cameras, sound recorders are all tools used to collect data. It is important to decide the tools for data collection because research is carried out in different ways and for different purposes. The objective behind data collection is to capture quality evidence that allows analysis to lead to the formulation of convincing and credible answers to the posed questions. The rationale behind data collection is to capture quality evidence for the research conclusions. Let's look at the most commonly used data collection methods and instruments for educational research.

INTERVIEWS

An interview is an oral conversation between two individuals with the sole purpose of collecting relevant information to satisfy a research purpose. Interviews entail collecting data by asking a series of oral questions face to face, on telephone or via video conferencing. The data can be collected by listening to individuals, recording/filming their responses, or a combination of two or more of these methods. Interviews are used when the researcher wants to understand experiences, views or perspectives of respondents in more elaborate detail.

There are three main types of interviews, with each having a slight variation from the other. They include; structured interview, semi-structured interview and unstructured/in-depth interview.

- **Structured Interviews** - Simply put, it is a verbally administered questionnaire. In terms of depth, it is surface level and is usually completed within a short period of time. For speed and efficiency, it is highly recommendable, but it lacks depth. In structured interviews, the questions as well as their order is already scheduled. Your additional intervention consists of giving more explanation to clarify your question (if needed), and to ask your respondent to provide more explanation if the answer they provide is vague (probing)
- **Semi-Structured Interviews** - In this method, there are several predetermined key questions, which cover the scope of the areas to be explored in the research. It allows

a little more leeway for the researcher to explore the subject matter by modifying the wording and order of questions.

- **Unstructured Interviews** - It is an in-depth interview that allows the researcher to collect a wide range of information with a purpose. It is less formal and the least structured, in which the wording and questions are not predetermined. This type of interview is most appropriate for collecting complex information, with a higher proportion of opinion-based information.

Advantages of Interviews

- ✓ They yield detailed information
- ✓ There is freedom of flexibility, as it allows for control over the order and flow of questions
- ✓ Data collected is accurate as it is from the horse's mouth
- ✓ It is done one-on-one, which leads to high response rates

Disadvantages of Interviews

- ➢ May be time-consuming
- ➢ Are relatively expensive to conduct
- ➢ Cannot be used for a large number of respondents
- ➢ Data analysis may be difficult, as it is majorly qualitative data
- ➢ Interviewing can be tiresome if participants are many
- ➢ Risk of bias is high due to possible fatigue
- ➢ May require skillful expertise to conduct

Data Collection Instruments and Tools for Interviews

For collecting data through interviews, here are a few tools you can use to collect data;

- **Audio Recorder -** An audio recorder is used for recording sound on disc, tape, or film. Audio information can meet the needs of a wide range of people, as well as provide alternatives to print data collection tools. It is mostly used by journalists when reporting news.

- **Digital Camera -** A digital camera can be used for transmitting images or responses to a monitor screen when need arises
- **Camcorder -** A camcorder is used for collecting data through interviews. It provides a combination of both an audio recorder and a video camera. Modern smartphones have inbuilt camcorders. The data provided is qualitative in nature and allows respondents to answer questions asked exhaustively. If you need to collect sensitive information during an interview, a camcorder might not work for you as you might need to maintain your subject's privacy.
- **Smartphone -** This has inbuilt devices that are mentioned above albeit with varying names e.g., sound recorder, voice changer etc
- **Computer-Assisted Interviewing System –** due to advancement in technology in the world today, there are many software that can help a researcher conduct interviews e.g. *iMocha*, *interviewstream* etc.

Interviews are executed using an instrument called an *interview schedule*. An interview schedule or *interview guide* is basically a document containing a set of structured questions that have been prepared, to serve as a guide for interviewers, researchers and investigators in collecting information or data about a specific topic or issue. The schedule will be used by the interviewer, who will fill in the answers received during the actual interview. There are two major types of interview schedules or guides that are widely used by interviewers in educational research;

In-depth interview schedule - used for open-ended interviews, which are aimed at obtaining in-depth information, often on serious topics or sensitive issues. The questions are open-ended, with prompts provided for the interviewer to ask for clarification or further information if necessary. The interviewee is given more room or leeway to talk about all the topics that will crop up during the interview, so he/she is free to use their own words and let ideas flow out naturally. The interview schedule contains indications of the interviewee's awareness of the purpose of the interview and how long it should take. The questions must be crafted in a manner that will elicit answers relevant to the topic or issue under investigation.

All questions in the interview schedule should be relevant, or have an impact on the purpose or objective of the interview. Remove any irrelevant questions, or those whose answers won't be

beneficial to the study. It takes the one-step-at-a-time approach, with each question meant to tackle only one issue, instead of addressing several issues all at once. This has a tendency to confuse not only the interviewee, but the interviewer too, and result in the latter losing control of the direction of the interview. Instead of using questions answerable with a simple Yes or No, the questions should be open-ended, which can be used as a starting or reference point for more questions. This way, the interviewer can go deeper in getting relevant information. The questions should be neutral. Leading questions should be avoided, as they have the potential to dictate the answer an interviewee gives.

Structured interview schedule is often compared with the format used in survey forms or questionnaires because of their similarities. The difference lies in the usage in the sense that obviously, the interview schedule is used by the interviewer during a face-to-face interaction with the respondent, while the questionnaire is simply filled out by the respondent. This interview schedule contains questions that will be asked, and it is also in the same place where the interviewer will record answers to those questions.

Essentially, preparing an interview schedule for a structured interview is the same as preparing a questionnaire. It's just that the questionnaire will be used solely by the interviewer, and the respondent (or interviewee) will not get to lay their eyes on the contents. For more flexibility however, some interviewers combine the features of these two types when they prepare their interview schedule. It would really be up to the interviewer and what they deem most effective in achieving the research objectives.

As such, there is no single standard template for an interview schedule. Generally, the format will depend on type and purpose of the interview being conducted, as well as the target respondents. However, a good interview schedule must have the following three key sections;

- **Opening** - Some researchers call this stage the "warm-up", where the goal is to create an atmosphere that will accommodate the open and free flow of ideas between the interviewer and interviewee. At the start of the interview, the interviewer should welcome the interviewee and make an effort to put him or her at ease. If the respondent is relaxed, the interview is likely to go smoothly. The interviewer will then proceed to inform the interviewee of the objectives of the interview, the topics or points that will be discussed in

the course of the interview, the estimated duration of the interview and any other information relevant to the interview. The researcher should allow the interviewee to raise any concerns or reservations they could be having before starting the interview begins. The interview must not proceed if all concerns of the interviewee have not been adequately addressed.

- **Body -** This contains the meat of the interview schedule i.e. the topics and the questions to be asked. Again, the content will depend on the topic and the type of interview. The main thing that you should never overlook is that the questions should fulfill the objective of the interview. Instead of an interview outline, which includes only a list of topics and subtopics, a typical interview schedule also contains major questions, as well as follow-up questions designed to probe or clarify the answers to the previously asked major questions. When preparing the body of the interview schedule, it is important to leave more than enough space where the interviewer may write down the responses or answers of the interviewee.
- **Closing** – at this stage, the interview is about to be wrapped up. The reason that it is included in the interview schedule is to ensure that the interview will not end abruptly, which may sound awkward or rude to the interviewee. The closing will cover a summary of all main points discussed during the interview, followed by a brief discussion on the next steps that will be taken after the interview.

FOCUS GROUP DISCUSSIONS

This is a structured discussion, with the purpose of stimulating conversation or debate around a specific topic. In typical terms, a Focus Group Discussion (FGD) is just an interview conducted to a group of respondents at the same time, where all the participants contribute towards the topic under discussion. Focus group discussion is led by a facilitator, who poses questions and the participants give their thoughts and opinions in orderly manner. Focus Group Discussion gives us the possibility to cross check one individual's opinion with other opinions gathered. A well organized and facilitated FGD is more than just a question-and-answer session. In a group situation, members tend to be more open and the dynamics within the group and interaction can enrich the quality and quantity of information needed. When planning a Focus Group Discussion, some practical issues to consider include;

- Size - ideal focus group should have 6 to 8 participants, 1 Facilitator and 1 Note-taker, a total of 10 persons maximum.
- Preparation – involves identifying the purpose of the discussion, assembly of participants and organizing the questions to be discussed
- Running the discussion – the FGD facilitator needs to make an interesting opening remark and manage the discussion well, by giving equal time to all respondents. The facilitator has the liberty to interject incase a respondent digresses from the main topic or is repeating what has already been noted already. A good closing remark should also be given at the end of the discussion, thanking all the participants of the discussion.
- Following up - plan some follow up activities after the discussion e.g. ferrying respondents to their respective homes, drafting letters of appreciation etc.

A Focus Group Discussion is implemented using an instrument called a Focus Group Discussion Schedule. Questions therein are open-ended, with the aim of stimulating an informal discussion with participants, in order to understand their perceptions, beliefs, fears, questions and information needs with regards to the topic under investigation. An FGD will usually take around one hour.

A facilitator and a note taker are needed to conduct a successful FGD. If possible, the facilitator and note taker should be of the same gender as the participants and speak their language or local dialect. The facilitator must be someone speaking the local language, with good communication skills and able to make respondents feel at ease. In cases where the facilitator does not speak the local language, they can work with a translator.

The facilitator needs to be able to probe further, basing on the responses received or may rephrase questions if respondents do not understand them. They must also be able to keep participants on topic and ensure that everyone has a chance to express themselves. An effective facilitator should be very familiar with questions in the FGD schedule. He/she should also be a careful listener and demonstrate empathy with participants. They should avoid leading questions, remain neutral and not react to participants' answers in order not to bias the session. If some of the questions are sensitive, ask them in a respectful manner and in line with the local culture.

The note taker must be able to speak the local language to record the group discussion points effectively. As well as recording what is said during the group discussion, the note taker should also record the behaviour of the participants e.g., remarkable attitudes, spontaneous reactions, interactions among the participants, etc. Moreover, the note taker should maintain confidentiality whilst recording the discussion by using letters or numbers to identify participants instead of actual names. The note taker may also decide to write only brief notes during the discussion but immediately after the interview, s/he should write the notes in detail so that all important information is recorded.

An effective note taker should; have good listening and writing skills, be familiar with the list of questions and topic under investigation, take notes in a comprehensive way but not literally, observe and remain impartial, ask, with the consent of the facilitator, a participant to repeat their answer if they do not hear it the first time.

Advantages of Focus Group Discussions

i. Collects multiple and wide ranging ideas in one session
ii. Allows in-depth discussion
iii. Group interactions can generate meaningful insights into the issue under investigation
iv. Can be conducted within a short time frame
v. Are relatively cheaper in comparison to interviews

Disadvantages of Focus Group Discussions

i. Findings may be difficult to conclude, due to lack of consensus in responses
ii. Requires a skilled facilitator and a quick note-taker
iii. Limited number of questions can be asked in a session
iv. Group setting, composition or meeting venue can influence opinions
v. Data can be difficult to analyze and interpret
vi. Not appropriate for all topics or populations

GROUP INTERVIEWS

In the event the researcher does not raise the minimum of 6 respondents required for a FGD in educational research, then group interviews can be used. This is just an oral interview consisting of 2 to 5 respondents. All other logistics of FGD apply to group interviews.

QUESTIONNAIRES

This is a self-reporting instrument consisting of a series of written questions, where a respondent is required to read by themselves and give a written response. Researchers in education use questionnaires to obtain information about the thoughts, feelings, perspectives, opinions, beliefs, values, personality and behavioral intentions of a group of respondents. For clarity, it is important to note that a questionnaire isn't a survey, rather it forms a part of it. A survey is a process of data gathering involving a variety of data collection methods, including but not limited to a questionnaire. Questionnaires can be administered in-person, via mail or online (via Google forms). On the downside, mail and online questionnaires stand the risk of impersonation and low return rates.

On a questionnaire, there are three kinds of questions commonly used in educational research; fixed-response, likert-type and open-ended. Each of the questions is tailored to the nature and scope of the research. The most common fixed questionnaire in education is the likert-type. Therein, the respondent is presented with a number of statements on an odd number of options to choose from (mostly 3-point or 5-point).

The following are examples likert-type scales;

3 point likert-type scale:	[Above Average, Average, Below Average]
5 point likert-type scale:	[Strongly Agree, Agree, Undecided, Disagree, Strongly Disagree]
7 point likert-type scale:	[Very Strongly Agree, Strongly Agree, Agree, Undecided, Disagree, Strongly Disagree, Very Strongly Disagree]

Tips for Constructing a Good Questionnaire

To yield the required information when using questionnaires as your primary data collection instrument, the following points are noteworthy;

- ✧ Give simple, clear instructions on how to fill it
- ✧ Arrange items in logical sequence
- ✧ Give room for neutral voices. An odd number likert scale is recommended i.e., 3, 5 or 7 responses to chose from. The middle response is the neutral voice (undecided/not sure).
- ✧ Keep it short to increase completion rate i.e., 10-20 items for each variable should suffice.
- ✧ Mix positive and negatively worded statements in shuffled order, to avoid selection bias
- ✧ Use language that is simple and commensurate to respondents' academic level
- ✧ Avoid unnecessary jargon or colloquial language
- ✧ Avoid loaded or compound statements e.g., chemistry lessons are boring and irrelevant. A respondent could find the lesson boring but relevant hence might be confused which response to pick. Separate them to avoid such dilemma.
- ✧ Avoid seeking irrelevant demographic information
- ✧ Avoid formatting and grammatical errors. These makes the whole work less scholarly
- ✧ Hypothetical or leading questions should be avoided e.g. isn't it quite obvious that….? A respondent is coerced by such statements to select an affirmative response.
- ✧ Give ethically sound incentives in the case of mail questionnaires e.g. postage paid envelope.

Advantages of Using Questionnaires in Educational Research

- ✧ Can be administered to many respondents over a short period of time
- ✧ Can be used to compare and contrast previous research to measure change
- ✧ Are easy to analyze
- ✧ They offer actionable data
- ✧ Respondent identity is protected
- ✧ They can cover many areas of a topic

- ✧ Can be administered in more than one form e.g., online via Google forms, by mail of physically
- ✧ Data can be converted to other forms e.g., from ordinal to interval. This gives room for elaborate statistical operations
- ✧ Come can be adapted or re-used

Disadvantages of Questionnaires

- ✧ Answers may be dishonest
- ✧ Respondents may lose interest midway leading low completion rates
- ✧ Not all questions can be analyzed easily e.g., open ended questionnaires.
- ✧ Are relative expensive if administered in hard copy to many respondents

OBSERVATION

Observation is a data collection method that involves careful observation of procedures or behaviour of respondents. It is used when the researcher wants to investigate how well the respondents execute a procedure or activity. Observation data can be collected in a laboratory or in a natural setting e.g. students in a classroom, shoppers in a supermarket, farmers on their farms etc. While collecting data this way, the observer should be as un-obstructive as possible. It is one of the most reliable ways of data collection if all indicators of the variable under investigation are clearly specified. Observation is executed using an instrument called an *observation checklist*.

A good observation checklist should be accompanied by an observation guide i.e., a document giving clear instructions of what to observe and how many points to award for each observed trait. In Educational settings, a researcher may use an observation checklist to assess students as they perform a skill, interact with peers in different learning situations, micro-teaching and teaching practice. Observation checklists for research purposes are highly structured, and may be scheduled over different periods of time in different learning contexts.

Items in the observation checklists allow the researcher to record information quickly and objectively, about how sampled individuals perform in relation to specific outcomes from the variable being investigated. Observation checklists, written in a yes/no format can be used to assist in observing student performance relative to specific criteria. They may be directed toward

observations of an individual or group. These tools can also include spaces for brief comments, which provide additional information not captured in the checklist.

An example:

An educational researcher has asked his research assistant (observer) to look at the issue of time management by pre-service teachers, so the observer's checklist includes these questions;

- Does the teacher follow the timings on the lesson plan? [Yes] [No]
- Does the teacher tell learners how long they have for an activity? [Yes] [No]
- Does the teacher tell learners when time is nearly up? [Yes] [No]

In such a case, it is important that learners to be aware of the function of the observer's presence in the classroom and be pre-informed about the observation. If not, the observer may intimidate the class and the data gathered may not be accurate due to the Hawthorne effect, which will be discussed in chapter six of this book.

Advantages of Observation Checklists

i. Allows you to get first-hand information
ii. Can reveal unanticipated information and sources of error
iii. Flexible in the course of data collection

Disadvantages of Observation Checklists

i. Are time consuming to use
ii. It may be difficult to observe multiple processes simultaneously
iii. Recording may be subjective if no clear instructions are given
iv. It may be difficult to score or interpret the observations if there are no clear indicators. This emphasizes the need for constructing an elaborate conceptual framework of the study at the initial stages of the research
v. Observers' presence may intimidate those being observed hence unexpected change of behaviour

DOCUMENT ANALYSIS GUIDES

Document analysis is a form of quantitative or qualitative data collection exercise in which documents are interpreted by the researcher to give voice and meaning around a research topic. Document analysis is executed using an instrument called a *document analysis guide (DAG)*. It is used when data or literature are already available, hence no need to collect it directly from respondents. Analyzing documents incorporates coding content into themes similar to how focus group or interview transcripts are analyzed. A rubric can also be used to grade or score the different items in a document.

There are three primary types of documents that are analyzed;

- **Public Records** – these are official, ongoing records of an organization's activities. Examples include student transcripts, mission statements, annual reports, policy manuals, student handbooks, strategic plans, occurrence books and course outlines.
- **Personal Documents** – these entail first-person accounts of an individual's actions, experiences, and beliefs. Examples include diaries, calendars, e-mails, blogs, Facebook posts, incident reports, student portals, unpublished thesis, fee statements, bank statements, whatsapp chats etc.
- **Physical Evidence** - Physical objects found within the study setting (often called artifacts). Examples include flyers, posters, meeting agendas, handbooks, training materials, minutes of a meeting, event programmes, CCTV footage etc.

Document analysis is important because it allows researchers to make sense of their resources of information and select appropriate pieces for their studies. Documents may contain qualitative data, which refers to non-numerical information that researchers organize into categories. Organizing the data allows researchers to combine document analysis with other sources of research data.

Researchers use documentary material because the information it contains can serve many purposes, including;

- Providing context within research
- Introducing questions researchers may ask

- Suggesting situations researchers should observe
- Providing additional insights researchers may find valuable
- Giving researchers a way to track progress
- Serving as a tool for confirming evidence found in other research

Researchers often use data collected during document analysis in a process called triangulation, which occurs when researchers compare two or more sources of data in the same category to determine their similarities and differences. This process has an important role in establishing credibility. Researchers may also use document analysis as a part of mixed-method studies, which combine quantitative and qualitative data to create a synergistic relationship that enables more efficient use of data.

Researchers approach the process of document analysis differently, depending on the purpose of their study, types of documents they're analyzing and the ways they prefer to conduct research. The following seven basic steps can guide you when conducting document analysis;

1. **List your resources** - Before beginning your document analysis, decide on the types of documents you intend to use. Select your texts based on factors such as: -

    - ✓ Medium, such as newspaper, magazine or website
    - ✓ Genre, such as politics, marketing or social issues
    - ✓ Inclusion criteria, such as only certain newspapers or websites
    - ✓ Parameters, such as date range and location

2. **Decide how to organize the information** - Organizing information in your document analysis involves defining two elements; units of meaning and set of categories. A unit of meaning refers to the elements you're looking for in your texts, such as words, phrases or images. Sets of categories are characteristics of the content. Once you define the elements of your information, you can use a coding system with rules you establish to organize it by theme, which makes analyzing your results easier.
3. **Make copies of notes** - Making copies of all your sources allows you to highlight and annotate the information without damaging the original documents. It also allows for portability, so if you conduct research in multiple environments, you can easily take them

from one place to another. When you're finished with your current study, you can file the copies for later research.

4. **Ensure authenticity** - only authoritative sources of information benefit your research, so it's important to ensure their trustworthiness before using their data. You can ensure the reliability of a document by validating factors such as; credibility, dependability, conformability, transferability and authenticity. Authentic sources typically have a listed author and copyright information included in the documentation. e.g., official school letter heads, official rubber stamps and authorized signatures.
5. **Check for biases** - even documents that seem objective can contain biased information, which can consequently skew their accuracy. Types of bias you can inspect for during document analysis include;
    - Confirmation bias - Using others' opinions to confirm beliefs
    - Culture bias - Making assumptions based on how someone perceives their own culture or another culture
    - Question-order bias - Using answers to questions to influence respondents to answer other questions the same way
    - Halo effect - Seeing a person in a certain way because of a single attribute
    - Wording bias - Elaborating on the answer to a question to make it fit a research topic

6. **Ask questions** - To conduct document analysis effectively, ask questions that help you determine the background of the document and how resources can help you with your research. Researchers often ask questions about who, what, when, where and why when assessing information to gather important details. Examples of these types of questions include;
    - Who wrote, researched or created this document?
    - What is this document about?
    - When was this document written?
    - Where did this document come from?

7. **Evaluate the document** - Once you have completed these steps, you have a framework for approaching your document analysis. You can identify patterns in the data, extract

useful information and determine how it applies to your research question. Finally, you can interpret your findings and use them to build on existing knowledge.

Advantages of Document Analysis Guides

i. Capture accurate information
ii. Are highly reliable
iii. Can be done with aid of technology
iv. Can be read and reviewed multiple times and remain unchanged by the researcher's influence or research process

Disadvantages of Document Analysis/Guides

i. Some may require expertise to fill since language used in some documents may be technical or coded
ii. Some require a lot of time to construct and use
iii. Cannot be used to collect highly sensitive information e.g., bank account statement
iv. Data is limited to what is available
v. Data may be incomplete

STANDARDIZED TESTS

Tests are commonly used in quantitative research to measure several variables including but not limited to; aptitude, academic achievement, personality and intelligence. The most common type of test is the standardized test, which is developed by psychometricians and usually includes psychometric information on reliability, validity and reference group norms, difficulty index, discrimination power etc. There are many types of tests are available for use e.g., standardized IQ tests, personality tests, aptitude tests etc. However, experimental researchers sometimes need to construct their own tests to measure very specific constructs that are operationalized in a manner unique to their study. The following are examples of tests;

i. **Achievement tests** – These are teacher-made tests in which scores are often used to determine the level of instruction for which a student is prepared. High achievement scores generally indicate that a level of mastery of grade-level material has been reached,

and that the student is prepared for advanced instruction. Conversely, low achievement scores can indicate the need for further remediation or repeating a year of study or unit.

ii. **Aptitude tests** - is a way for employers to assess a candidate's abilities through a variety of different testing formats. Aptitude tests will test your ability to; perform tasks, react to situations at work, personality traits and cognitive ability, amongst other variables.

iii. **Intelligence tests** - An intelligence quotient (IQ) test is an assessment that measures a range of cognitive abilities and provides a score that is intended to serve as a measure of an individual's intellectual abilities and potential. IQ is basically a type of standard score that indicates how far above or how far below, his/her peer group an individual stands in mental ability.

IQ tests are among the most administered psychological tests. Today, many IQ tests are standardized, and scores are derived by comparing individual performance against the norm for people in that age group. The most widely used intelligence tests include the *Stanford-Binet* Intelligence Scale and the *Wechsler* scales. Modern intelligence tests often focus on abilities such as mathematical skills, memory, spatial perception and language abilities.

The capacity to see relationships, solve problems, and remember information are important components of intelligence, so these are often the skills on which IQ tests focus. IQ tests can be used for a wide range of purposes, including;

- Educational measurement and evaluation
- Assessment and diagnosis of intellectual disability of a leaner
- Cognitive research about memory, speed, and attention of a leaner

Advantages of Standardized Tests

i. Can be used on a large number of respondents over a short time
ii. Some are easy to construct
iii. Data is easy to analyze and interpret
iv. Validity and reliability can be assessed easily

Disadvantages of Standardized Tests

i. May be time consuming to administer and score
ii. Require experts to construct and score
iii. Are expensive
iv. Some may require permission in order to use
v. Some become obsolete with time

CHAPTER 6:

VALIDITY AND RELIABILITY OF EDUCATIONAL RESEARCH INSTRUMENTS

This chapter talks about; definition of validity, types of validity, assessment of validity of research instruments, reliability of research instruments, methods of assessing reliability, factors that affect reliability and lastly, the trustworthiness of qualitative research instruments.

What is Validity?

The conclusions you draw from your research (whether from analyzing survey, focus groups, experimental design, or other research methods) are only useful if they're valid. How "true" are these results? How well do they represent the thing you're actually trying to study? All these questions are talking about validity.

Validity of a research instrument is the measure of how well it does its intended job. Unfortunately, some researchers sometimes come up with their own operational definitions of what they consider as valid in their study. In educational research, testing for validity and reliability of a data collection tool is mandatory. However, some rogue researchers have wrongfully insinuated that instrument validity does not apply to their research. Nothing could be further from the truth! Validity is always important - only that it is harder to determine in qualitative research.

To disregard validity is tantamount to putting the trustworthiness of your work in question, and to call into question others' confidence in your research findings. Even when qualitative measures are used in research, they need to be looked at using measures of reliability and validity in order to sustain the trustworthiness of the results.

Validity and reliability make all the difference between "good" and "bad" research reports. Quality research depends on a commitment to testing and increasing the validity, as well as the reliability of your research instruments and eventual results. Any research worth its weight is concerned with whether what is being measured is what was intended. It therefore evaluates the ways in which observations are influenced by the circumstances in which they are made. The basis of your conclusion plays an important role in addressing the broader substantive issues of any given study. For this reason, let us look at various types of internal validity that must be assessed in educational

research. Note that external validity is not considered here, as it is more of a research design issue, rather than having something to do with instruments.

Types of Validity of Research Instruments

Internal validity is the concept of how much confidence you have in the result of your research. This is because the lesser the possibility of confounding variables in research, the greater the internal validity and the more confident a researcher can be of the research process. What this means is that internal validity is the degree to which you can establish the cause-and-effect association between your treatment condition and your research result. It is how best you can avoid traps in methodology that can alter your result as a researcher.

Internal validity also shows that knowing the standard data helps a researcher to take out insignificant results in the research. With an accurate selection of sample groups and proper data collection, the result of the relationship between the data variables will be very valid. This reiterates the need for researchers to do everything possible to ensure their instruments have acceptable levels of validity.

In educational research, there are four main types of internal validity that are associated with research instruments therein;

i. **Face validity -** this form of validity is concerned with how valid your results seem based on what they look like on face value. This is the least scientific form of validity, as it is not quantifiable using statistical methods. Face validity is not validity in a technical sense of the term, as it is only concerned with what we see or perceive. Here we look at how valid a measure or research instrument appears on the surface and make subjective judgments based on that.

For example:

> *Imagine you give a survey that appears to be valid to the respondent and the questions are selected because they look valid to the researcher. The researcher asks a group of random people, untrained observers, if the questions appear valid to them.*

In research, it's never enough to rely on face judgments alone. More quantifiable methods of validity are necessary in order to draw acceptable conclusions. There are many instruments of measurement to consider, so face validity is useful only in cases where you need to distinguish one instrument from another. Face validity should therefore never be trusted on its own merits.

ii. **Content validity -** this is concerned with whether the data captured by the research instrument covers all the indicators of the variable under investigation. This is a less subjective measure, but unlike face validity, we ask whether the content of a measure covers the full domain of the content. If a researcher wanted to measure attitude, they would first have to decide all indicators of attitude e.g., feeling of dislike, disapproval of a person etc. leaving out some of them lowers validity of the instrument.

    Content validity is partially considered to be a subjective form of measurement because it still relies on people's perception for measuring variables or indicators that would otherwise be difficult to measure. Where content validity distinguishes itself (and becomes useful) is through its use of experts in the field or individuals belonging to a target population. This type of validity can be made more objective through the use of rigorous statistical tests. For example, you could have a content validity assessment to inform you how items used in an achievement test represent the course outline, how clear they are, and the extent to which they address the research objective**.**

iii. **Construct validity -** A construct represents a collection of behaviors that are associated in a meaningful way to create an image or an idea invented for a research purpose. Construct validity is the degree to which your research instrument measures the construct (as compared to things outside the construct).

iv. **Criterion-Related Validity -** Criterion-related validity (also called *instrumental validity*) is a measure of the quality of your research instrument. The accuracy of a measure is demonstrated by comparing it with a measure that is already known to be valid. In other words, if your attitude questionnaire has a high correlation with standard attitude questionnaires or those previously used ones, it has high criterion-related validity. If a

teacher sets a test whose outcomes are similar to KCSE trends, then the test has high criterion-related validity.

Assessment of Validity of Research Instruments

To help researchers decide whether or not their data collection instruments are valid, it is recommended computing an average of the applicable types of validity. Experts that are well-versed with the research topic should be used to assess the relevant types of validity on a scale of, say 1 to 10. In educational research, an average coefficient of 0.65 (or 6.5/10) and above is an indicator that the instrument is generally good enough hence acceptable to measure what it purports to measure. Anything less than that and the instrument should be reviewed as per the expert's recommendations.

Content validity is the most important type of validity that must be assessed in all instruments used for collecting data in educational research. To validate the content of a research instrument, experts scrutinizing the instrument should critically examine and compare items therein against the respective variables on the conceptual framework. They may also use course outlines, course descriptions, syllabuses and schemes of work in the case of achievement tests. Rating of the items should then be done, basing on how well they represent the indicators of the variables to be measured. The researcher should therefore identify a person who is well versed with the research topic and use him or her as the rater of content validity.

The researcher should supply the expert (rater) with some background information e.g. what is the instrument being used for, who is the respondent? where and when will the data be collected etc. In the case of postgraduate research, the first university supervisor automatically becomes one of the raters but two other experts who have nothing to do with the research should also be used as raters to get their independent view. An average of the three scores can then be used to make the ultimate decision on whether the instrument is valid or not. The author of this book recommends a threshold of 6.5 (out of 10) and above for any research in education as the greenlight to use the instrument with minor corrections as recommended by the expert reviewers. For any average validity score values less than 6.5 but greater or exceeding 0.5, the instrument needs an overhaul. For values less than 0.5, the instrument should be discarded, as it does not measure what it is intended to measure to acceptable scholastic standards.

Reliability of Educational Research Instruments

Apart from validity, another measure that is used to determine whether or not research instruments are suitable for data collection is reliability. An instrument that is both valid and reliable is perfect for use in the actual study. Reliability refers to the ability of a research instrument in elicit consistent responses when administered to the same set of respondents under similar research conditions. Quantitative research instruments are those that collect continuous data i.e., interval or ratio scales of measurement. Such kind of data allows for several complex mathematical operations and therefore reliability of quantitative instruments must be determined for credible findings and conclusions. There are several methods of assessing reliability of a quantitative research instrument, discussed in the next section.

Methods of Assessing Reliability of Research Instruments

The following methods are commonly used to assess the reliability of research instruments. Researchers should carefully select the most appropriate method.

i. **Equivalent Forms Method** - also called *parallel forms* reliability. It measures the correlation between two equivalent versions of an instrument. You use it when you have two different assessment tools or sets of questions designed to measure the same variable. Respondents' scores in one version of an instrument are correlated against those of the other version. A correlation coefficient value of +0.65 or higher indicates a reliable instrument.

ii. **Test-Retest Method** - When researchers measure a construct that they assume to be consistent across time, then the scores they obtain should also be consistent across time. Test-retest reliability is the extent to which this is actually the case. For example, intelligence is one variable that that doesn't change significantly over time. A person who is highly intelligent today will still be highly intelligent next week, next month and even next year. This means that any good measure of intelligence should produce roughly the same scores for this individual next week as it does currently.

    Clearly, an instrument that produces inconsistent scores over time cannot be a very good measure of a variable that is supposed to be consistent. Assessing test-retest reliability requires using the instrument on a group of people at one time, using it again on the same

group of people at a later time, and then computing a correlation between the two sets of scores. This is typically done by computing Pearson's correlation coefficient (r) or Spearman's Rank Coefficient depending on the type of data collected by the instrument. A value of +0.65 or higher indicates a reliable instrument.

iii. **Internal Consistency Method -** A third kind of reliability is internal consistency, which is the consistency of people's responses across the items on a multiple-item measure. In general, all the items on such measures are supposed to reflect the same underlying construct, so people's scores on those items should be correlated with each other. Like test-retest reliability, internal consistency can only be assessed by collecting and analyzing quantitative data. There are several methods of calculating the internal consistency of a research instrument. These include; split-half method, Kuder-Richardson method and the Cronbach's Alpha method. Each of these is discussed separately in detail next, as separate forms of reliability assessment.

iv. **Split-Half Method**: Split-half reliability on its own is a statistical method used to measure the internal consistency of scores of a test item. It is a form of internal consistency reliability and had been commonly used before the coefficient α was invented. Split-half reliability is a convenient alternative to other forms of reliability, including test–retest reliability and parallel forms reliability because it requires only 1 administration of the test. As can be inferred from its name, the method involves splitting a test into halves and correlating the respondents' scores on the two halves of the instrument. The resulting correlation is then adjusted for test length using the Spearman-Brown prophecy formula or Pearson's Product Moment Correlation Coefficient if the variables were continuous. In summary, this method involves splitting the items on a test into two sets, such as the first and second halves of the items or the even- and odd-numbered items. A score is then computed for each set of items, and the relationship between the two sets of scores examined. A value of +0.65 or higher indicates a reliable instrument.

v. **Kuder-Richardson Method**

This is a measure of Internal Consistency of a test or quantitative research instrument. It entails estimating the reliability of test scores from a single administration of a single form of a test by means of formulas developed by Kuder and Richardson, hence the name. Like

the spilt-half method, these formulas provide a measure of internal consistency but does not require splitting the test into halves for scoring purposes.

Kuder-Richardson formula 20

The Kuder-Richardson formula 20, abbreviated KR-20, is used to determine the reliability of a test with items of varying difficulty. For example, some items might be very easy, others trickier. It should be used if there is only one correct answer for each question. It is analogous to Cronbach's α, except Cronbach's α is used for non-dichotomous (continuous) measures. KR-20 should not be used for questions with possibility for partial credit or for scales like the likert scale questionnaires. If all questions in your binary test are equally challenging, use the KR-21, which will be discussed shortly. If on the other hand you have a test with more than two answers or possibilities of partial credit, use Cronbach's Alpha instead.

The Kuder-Richardson formula 20 is based on the proportion of persons getting each item right and the standard deviation of the total scores. The result of the analysis using this formula is equal to all possible split-half coefficients for the group tested. However, it is rarely used among educational researchers, due to the fact that the computation is rather cumbersome, especially if the test has many items, unless information is already available concerning the proportion of each item. Since Cronbach's α was published in 1951, there has been no known advantage to KR-20 over Cronbach. KR-20 is therefore seen as a derivative of the Cronbach formula, with the advantage to Cronbach that it can handle both dichotomous and continuous variables.

Computing the Kuder and Richardson Formula 20

As earlier mentioned, the Kuder and Richardson Formula 20 test checks the internal consistency of measurements with dichotomous choices. It is equivalent to performing the split-half methodology on all combinations of questions and is applicable when each question is either right or wrong. To compute KR-20, every correct question should be scored 1 and every incorrect question scored 0. The test statistic is then computed using the formula;

$$\text{KR-20} = \frac{n}{n-1}\left(1 - \frac{\sum pp\prime}{s^2}\right)$$

Where;

n = number of items in the test or instrument

s^2 = variance of the total scores of all the people taking the test

p = number of people in the sample who answered item p correctly

p' = number of people in the sample who answered item p incorrectly

(hint: the product of p and p' is computed for each item in the test and these products summed up for all the n items in the test to give $\sum pp'$)

Note that values of KR-20 range from 0 to 1, where a high value (at least 0.7) indicates the test is reliable, while too high a value (0.9 or greater) indicates a homogeneous test. High values of KR-20 therefore indicate that the research instrument is likely to correlate with alternate forms of the test or instrument. It should also be noted that KR-20 is usually affected by item difficulty, variance of the scores and number of items in the test or instrument.

Example:

A questionnaire with 11 items was administered to 12 students by a certain educational researcher. The results were as listed in the green portion of the table below:

Respondent	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Student 1	1	1	1	1	1	1	1	1	1	1	1
Student 2	1	1	1	1	1	1	1	1	0	1	0
Student 3	1	0	1	1	1	1	1	1	1	0	0
Student 4	1	1	1	0	1	1	0	1	1	0	0
Student 5	1	1	1	1	1	0	0	0	1	0	0
Student 6	0	1	1	0	1	1	1	1	0	0	0
Student 7	1	1	1	1	0	0	1	0	0	0	0
Student 8	1	1	1	1	1	0	0	0	0	0	0
Student 9	0	1	0	1	1	0	0	0	0	1	0
Student 10	1	0	0	1	0	1	0	0	0	0	0
Student 11	1	1	1	0	0	0	0	0	0	0	0
Student 12	1	0	0	1	0	0	0	0	0	0	0

The reliability of this questionnaire can be calculated using KR-20 as shown below:

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Total
Student 1	1	1	1	1	1	1	1	1	1	1	1	11
Student 2	1	1	1	1	1	1	1	1	0	1	0	9
Student 3	1	0	1	1	1	1	1	1	1	0	0	8
Student 4	1	1	1	0	1	1	0	1	1	0	0	7
Student 5	1	1	1	1	1	0	0	0	1	0	0	6
Student 6	0	1	1	0	1	1	1	1	0	0	0	6
Student 7	1	1	1	1	0	0	1	0	0	0	0	5
Student 8	1	1	1	1	1	0	0	0	0	0	0	5
Student 9	0	1	0	1	1	0	0	0	0	1	0	4
Student 10	1	0	0	1	0	1	0	0	0	0	0	3
Student 11	1	1	1	0	0	0	0	0	0	0	0	3
Student 12	1	0	0	1	0	0	0	0	0	0	0	2
Total	10	9	9	9	8	6	5	5	4	3	1	69
P	0.83	0.75	0.75	0.75	0.67	0.5	0.42	0.42	0.33	0.25	0.08	
P’	0.17	0.25	0.25	0.25	0.33	0.5	0.58	0.58	0.67	0.75	0.92	
PP’	0.14	0.19	0.19	0.19	0.22	0.25	0.24	0.24	0.22	0.19	0.07	**2.15**

From the table, N = 11, and $\sum$PP’ = 2.15

The variance (calculated separately) = 6.52

Substituting these values of N, $\sum$PP’ and variance (s^2) in the KR-20 formula gives;

$$\text{KR-20} = \frac{11}{11-1}\left(1 - \frac{2.15}{6.52}\right)$$

$$= \frac{11}{10}(1 - 0.3298)$$

$$= 1.1 \times 0.6702$$

$$\mathbf{= 0.7372}$$

The value of KR-20 obtained is 0.7372 as shown, which indicates that the test has high reliability.

Kuder-Richardson Formula 21

The Kuder-Richardson formula 21, abbreviated KR-21, is a less accurate but simpler formula to compute as compared to KR-20 as previously described. The KR-21 formula is applied to the results of any test of same item difficulty (i.e. the mean score of each question is approximately equal to the mean score of all the questions in the test). It is also used where test items have been scored on the basis of the number of correct answers. A simplified version of the formula is as follows:

$$r = \frac{n}{n-1}\left[1 - \frac{m(n-m)}{ns^2}\right]$$

Where;

n = the number of items in the test

m = the mean of the test scores

s^2 = variance of the test scores.

Example:

A researcher in his pilot study used an achievement test containing 60 items and the mean of the respondents' scores was 45 and a variance of 36 was obtained. The estimate reliability of this test can be calculated as follows;

$$r = \frac{60}{60-1}\left[1 - \frac{45(60-45)}{60(36)}\right] = 0.699$$

The result of this formula KR-21 approximates that of KR-20 and therefore has a relatively smaller reliability estimate in most cases. This method of reliability estimates whether the items in the test are homogenous i.e. it seeks to determine whether each test item measures the same quality or characteristics as every other therein. If this is established, then the reliability estimate will be similar to that provided by the split-half method. On the other hand, if the test lacks homogeneity, an estimate smaller than split-half reliability will be obtained.

The Kuder-Richardson formula and the Split-half methods are widely used in determining reliability in educational research because they are robust and simple to apply. Nevertheless, the following limitations restrict their value;

- They are not appropriate for speed tests in which test retest or equivalent form methods are better estimates
- Just like the parallel forms method, they do not indicate the constancy of a test taker's response day in day out. Only the test-retest procedures indicate the extent to which test results are generalizable over different periods of time.
- They are most adequate only for teacher-made achievement tests, as these are usually power tests.

vi. **Cronbach's Alpha Method** – this is arguably the most common measure of internal consistency used by educational researchers. Conceptually, α is the mean of all possible split-half correlations for a set of items in a test. For example, there are 252 ways to split a set of 10 items into two sets of five. Cronbach's α would be the mean of the 252 split-half correlations. Note that this is not how α is actually computed, but it is a correct way of interpreting the meaning of this statistic. Again, a value of +0.65 or greater is generally taken to indicate good internal consistency.

Cronbach's alpha is used to assess the reliability, or internal consistency, of a set of scale or test items. As previously mentioned, the reliability of any given measurement refers to the extent to which it is a consistent measure of a concept, and Cronbach's alpha is one way of measuring the strength of that consistency. Cronbach's alpha is computed by correlating the score for each scale item with the total score for each observation (usually individual survey respondents or test takers), and then comparing that to the variance for all individual item scores. The Cronbach's alpha statistic is computed using the formula;

$$\alpha = \frac{n \times \bar{c}}{\bar{v} + \bar{c}\,(n-1)}$$

where: n refers to the number of test items

$\bar{c}$ refers to the average of all covariances between items

$\bar{v}$ refers to the average variance of each item

Cronbach's alpha is thus a function of the number of items in a test, the average covariance between pairs of items, and the variance of the total score. The resulting α coefficient of reliability ranges between 0 and 1 in providing this overall assessment of an instrument's reliability. If all of the test items are entirely independent from one another (i.e. are not correlated or share no covariance), then α = 0, and if all items in the instrument have high covariances, then α will be close to 1 as the number of items in the scale approaches infinity. In other words, the higher the α coefficient, the more the items have shared covariance and probably measure the same underlying concept.

Although the thresholds α coefficients of any instrument are entirely arbitrary and depend on one's theoretical knowledge of the scale or instrument in question, It is recommended that a minimum α coefficient + 0.65 for all instruments used for any educational research passes the reliability test. Consequently, any instrument whose α coefficients is less than 0.5 is unacceptable in educational research, especially for instruments purporting to be measure only one dependent variable. A coefficient of between 0.5 and 0.65 may still be used but after major corrections as per expert recommendations.

This type of reliability is best calculated using SPSS as manual calculation is tedious and pointless in the 21st century. The following is an example:

Imagine a university lecturer wants to conduct an investigation on his students to evaluate how satisfied they are with timeliness of his examination processing. He develops the following four items in his close-ended questionnaire;

Item 1 - My concerns were addressed in a reasonable amount of time.

Item 2 - I am satisfied with the timeliness of the exam processing.

Item 3 - The time I waited for my examination results was reasonable.

Item 4 - I am satisfied with the CAT and examination marks I received.

Assuming these questions all use a 5-point Likert scale, ranging from 1-Very Dissatisfied to 5 - Very Satisfied, the lecturer asks 60 of his B.Ed Arts students to fill the questionnaire during the pilot study phase before distributing the same to the rest of the university students he teaches.

For this analysis, the computer software SPSS calculates the overall Cronbach's alpha. Then it recalculates the statistic after omitting each item because that process can provide valuable information about specific items. An overall Cronbach's alpha of at least 0.65 is acceptable in educational research.

Under Omitted Item Statistics, SPSS recalculates Cronbach's alpha after removing an item. If omitting an item substantially increases Cronbach's alpha, the researcher should consider removing that question from the instrument because it is suspect.

vii. **Inter-Rater Reliability Method** - Many behavioural measures in educational research involve significant judgment on the part of an observer or a rater. Inter-rater reliability is the extent to which different observers are in agreement in their judgments. For example, if you were interested in measuring university students' social skills, you could make video recordings of them as they interacted with another student whom they are meeting for the first time. Then you could have two or more observers watch the videos and rate each student's level of social skills.

    To the extent that each participant does in fact have some level of social skills that can be detected by an attentive observer, different observers' ratings should be highly correlated with each other. Inter-rater reliability is often assessed using Cronbach's α when the judgments are quantitative or an analogous statistic called Cohen's κ (the Greek letter kappa) when they are categorical.

    At this point, I think the debate as to whether qualitative instruments can be assessed for reliability is now settled. A threshold of 0.4 and above is recommended for qualitative instruments like interview schedules while for quantitative instruments, 0.65 is the recommended threshold.

Factors that Affect Reliability of an Instrument

The various factors that influence a research instrument's reliability may be extrinsic or intrinsic. Extrinsic factors are factors that lie outside the test or instrument, but tend to make the test reliable or unreliable. They include;

i. **Group reliability** - when a group of respondents being investigated are homogeneous in ability, their reliability coefficient is likely to be low. However, where the respondents vary widely in their ability (are heterogeneous) the reliability of test score is likely to be higher, meaning reliability coefficient of the test is high.
ii. **Guessing -** guessing by the respondents may raise the total score in an instrument, which makes its reliability co-efficient extremely high, leading to error variance.
iii. **Environmental conditions** – the testing environment need to be a conducive one e.g. sitting arrangement, noise levels, ventilation, lighting etc. A poor testing environment causes destruction of the mental processes while respondents are filling an instrument. Panicking may also interfere with respondents' cognition process. These conditions cause momentary fluctuations in the mindset of the respondents, thereby raising or lowering the scores which affect reliability coefficient of the test.

Intrinsic factors, are also known as internal factors include aspects of an instrument in itself, which may lead to inconsistent responses. These include;

i. **Length of the test** - An instrument with many items is likely to have high reliability coefficient compared to a test with few items. More items increase the potential variability thereby improving the instruments' reliability. A questionnaire for instance should have at least 10 items while an achievement test should have atleast 15 questions if they are short answer type.
ii. **Range of total scores** - When the standard deviation of the total score is high, reliability coefficient is also high, when the standard deviation of the total score is low, then the coefficient is likely to be low. For multiple choice questions therefore, the items should be a minimum of four while for likert scales, 5 points are recommended for educational research.

iii. **Homogeneity of test items** - When test items measure the same indicators or traits from one item to another then the reliability coefficient will be low. Items in the instrument should therefore seek information on different indicators of the same variable.

iv. **Difficulty index of test items** - When items in an achievement test are too easy (answer too obvious) or too difficult (questions too tricky to understand), the instrument may not give a clear picture of the individual being investigated. Items in a good instrument should not be such that they are unanswered or are answered the same way by all randomly sampled respondents, as this lowers the reliability coefficient of the instrument.

v. **Discrimination power** - this mostly applies to achievement tests, whereby when the test is made by discriminating items, the item total test correlation is likely to be high, thus raising the reliability coefficient. Where test items do not discriminate between the superior and inferior learners, then the total correlation results to low reliability coefficient.

vi. **Scoring reliability** - scorer reliability means how closely two or more raters or scorers agree in scoring or awarding same number of marks to an item. If they do not agree, reliability coefficient of the test is likely to be lowered.

Trustworthiness of Qualitative Research Instruments

It is not very easy to determine reliability of qualitative instruments, since the type of data collected by such instruments (categorical) does not allow for any form of mathematical calculations that are part and parcel of reliability analysis, unless the tips suggested in the previous section are implemented. The best alternative for qualitative research instruments is to assess their trustworthiness. Trustworthiness is about establishing the following qualitative measures;

- **Credibility** - how confident the qualitative researcher is in the truth of the research study's findings. Researchers can use *triangulation* to show that the research study's findings are credible. This is discussed in chapter 7 of this book
- **Transferability** - how the qualitative researcher demonstrates that the research study's findings are applicable to similar situations, populations and phenomena. Researchers can use detailed descriptions to show that the research study's findings can be applicable to other contexts, circumstances and situations.
- **Confirmability** - the degree of neutrality in the research study's findings. This occurs when the findings are based on participants' responses and not any potential bias or

personal motivations of the researcher. Researchers can provide an audit trail, which highlights every step of data analysis that was made in order to provide a rationale for the decisions made. Transcripts from interviews might come in handy in this regard.

- **Dependability** - the extent to which the study could be replicated by other researchers and yield similar findings. If another person wants to repeat the study, they should have enough information about the methodology used in order to obtain similar findings. Researchers can use inquiry audit (an outside person to review and examine the process) in order to establish dependability.

CHAPTER 7:

DATA ANALYSIS IN EDUCATIONAL RESEARCH

This chapter presents; the definition of data analysis, reasons for data analysis, steps in data analysis process, types of data, types of data analysis and the various ambiguities in classifying data variables,

Definition of Data Analysis

Data analysis is the process of cleaning, editing and processing raw data in order to extract actionable, relevant and sensible information that helps the researcher make informed decisions about their research questions or hypotheses. This is a very important step in research as it helps reduce the risks poor decision-making by providing useful insights and statistical evidence for backing up any claims.

Reasons for Data Analysis

Data analysis is the most crucial part of any research, as it summarizes collected data to generate evidence the research conclusions. Proper data analysis leads to accurate conclusions. The following are the main reasons why data is analyzed;

i. To establish trends and patterns in the research variables
ii. To determine causes of the observed changes in respondents' characteristics
iii. To predict future relationships between research variables
iv. To obtain evidence for accepting or rejecting the research hypothesis
v. To make informed decisions about the population

Preparing Data for Analysis

Once data has been collected using various instruments, the next is to make it ready for rigorous analysis. Several preliminary steps must be carried out on the raw data to facilitate analysis;

i. **Validation** - Data validation is done to understand if the collected data sample is per the preset standards. This involves physical inspection to confirm that what was collected in the field is what is before the researcher's eyes.

ii. **Screening** – This must be done to make sure each respondent was selected in compliance with the inclusion criteria, which ensures ethical standards are maintained. Thorough screening confirms if the respondents answered all the questions or if the interviewer asked all the questions.

iii. **Editing** - More often, an extensive research data sample comes loaded with errors. Respondents sometimes fill in some fields incorrectly or sometimes skip them accidentally. Data editing is a process where the researchers confirm that the provided data is free of such errors and outliers.

iv. **Coding** - this is the most critical phase of data preparation and entails grouping and assigning values to the responses given in the research instruments. If a survey is completed with a 1000 sample size, the researcher will create an age bracket to distinguish the respondents based on their age. Thus, it becomes easier to analyze small data buckets rather than massive data piles. Coding data enables entry into the computer for further analysis using statistical software.

Types of Data

As we are going to see shortly, there are several methods of analyzing data, which all depend on the type of data collected by the research instruments or the scale of measurement on which the data or variable lies. We will therefore discuss the types of variables before embarking on specific methods of data analysis. All research instruments capture information in form of a variable. In fact, most postgraduate students who have their research topics rejected by their supervisors for being "un-researchable" is due to their lack of clear measurable variables. A variable therefore is not only something that we can measure, but also something that we can manipulate or control. Research data variables can broadly be classified as either discrete or continuous.

Discrete variables are also called categorical (can be put into classes) or qualitative variables (since they emanate from qualitative research). Categorical variables can be measured on either nominal or ordinal scale. Continuous variables on the other hand are variables that can take any value including decimals. They are also known as quantitative variables since they emanate from quantitative research. Continuous variables can further be measured on either interval or ratio scale.

Let us now look at the four scales of measurement in detail before embarking on the methods used to analyze data;

i. **Nominal data -** is discrete data that results from measuring variables that have two or more categories, but which do not have an intrinsic order. Nominal data has the property of identification and nothing more e.g. serial number or name. The number used in a nominal scale does not represent any quantity. Examples of nominal variables include gender (male/female), nationality (citizen/non-citizen), tribe, race (black/white/half-caste), school type (national/extra-county, county/sub-county) etc.

   Nominal variables which have only two subsets or levels are called *dichotomous variables*. For example, if we were looking at gender, we would most probably categorize somebody as either male or female. This is an example of a nominal dichotomous variable. Nominal scale is a measurement scale where data that can only be counted and put into categories, with no particular order of the categories. It has the property of identification and nothing more e.g. index number or name of a student. The number used in a nominal scale does not represent any quantity.

ii. **Ordinal data –** is discrete data that results from measuring variables that have two or more subsets just like nominal variables but the subsets can be ordered or ranked. So if you asked someone if they liked the measures put in place by the Kenyan government to manage the Covid-19 pandemic and they could answer either “Not very much”, “They are OK” or “Yes, a lot”, then you have an ordinal variable.

   This is because you have 3 categories of responses, which you can rank from the most positive (Yes, a lot), to the middle response (They are OK), to the least positive (Not very much). However, whilst we can rank the levels, we cannot place a value to them e.g. we cannot say that “They are OK” is twice as positive as “Not very much”. Ordinal scale is basically a form of quantitative measurement that indicates a numerical order such e.g. $2 < 3$ or $5 > 4$, i.e. the order and a succession of the numbers may be from top to bottom, greater to least, highest to lowest etc on some property. However, it lacks the elements of arithmetic i.e. additions or subtractions cannot be bone as they are meaningless.

iii. **Interval data –** this is continuous data obtained by measuring variables for which their central characteristic is that they can be measured along a continuum and they have a numerical value (for example temperature measured in degrees Celsius or Fahrenheit).
So the difference between 20°C and 30°C is the same as 30°C to 40°C. Interval scale, which is sometimes called equal internal scale, is a measurement that has equal units of measurement and an arbitrary zero e.g. 0^{o}C does not mean that there is no temperature.

iv. **Ratio data -** this typically interval data, but with the added condition that a value of 0 (zero) on the measurement variable indicates that there is none of that variable. So, temperature measured in degrees Celsius or Fahrenheit is not a ratio variable because 0°C does not mean there is no temperature. However, temperature measured in Kelvin is a ratio variable as 0 Kelvin (often called absolute zero) indicates that there is no temperature whatsoever.

Other examples of ratio variables include height, mass, distance and many more. The name ratio" reflects the fact that you can use the ratio of measurements. So, for example, a distance of ten metres is twice the distance of 5 metres. Ratio scale is therefore a measurement scale that has equal units of measurement and an absolute zero point i.e. the zero point is real and indicates total absence of the property measured e.g. if you have zero shillings or there is zero weight means there is nothing at all. Or if Makori weighs 100 kg and Andayi weighs 50 kg, it means Makori is twice as heavy as Andayi. An object of zero mass means it is totally weightless.

Ambiguities in Classifying a Type of Variable

In some education and social science cases, the measurement scale for data may be ordinal, but the variable is treated as continuous. For example, a Likert scale that contains five values - strongly agree, agree, undecided, disagree, and strongly disagree - is ordinal. However, where a Likert scale contains seven or more value - strongly agree, moderately agree, agree, neutral, disagree, moderately disagree, and strongly disagree - the underlying scale is sometimes treated as continuous (although where you should do this is a cause of great dispute).

It is worth noting however that how we categorize variables is somewhat of a researcher's choice. Whilst we categorized gender as a dichotomous variable (you are either male or female), social scientists may disagree with this, arguing that gender is a more complex variable involving more

than two distinctions, but also including measurement levels like gender-queer, intersex and transgender. At the same time, some researchers would argue that a Likert scale, even with seven values, should never be treated as a continuous variable.

Types of Data Analyses

Data analysis methods can be classified into two broad categories; qualitative and quantitative data analysis. The type of analysis a researcher uses depends on the type of data or variables as earlier discussed. Categorical variables call for qualitative analysis while continuous variables call for quantitative data analysis. Mixed methods researches call for a combination of both qualitative and quantitative data analysis methods. We will therefore discuss the types of variables before embarking on specific methods of data analysis.

Qualitative Data Analysis

Qualitative data is typically generated from; interview transcripts, open-ended questionnaires, focus group discussion notes, texts/documents, audio/video clips, observational notes and even photographs. Analyzing such data is not easy because while tools like MS Excel and SPSS visualize quantitative data with ease, there are no such mainstream tools for analyzing qualitative data. Majority of qualitative data is therefore analyzed manually.

Qualitative data analysis is a process of gathering, structuring and interpreting qualitative data to understand what it represents. In educational research, qualitative data may include reasons why students choose a particular course at university, *WhatsApp* status comments etc.

Steps of Qualitative Data Analysis

Qualitative data analysis is done in 5 major steps;

(i) Gathering and collecting your qualitative data
(ii) Organizing and connecting into your qualitative data
(iii) Coding your qualitative data
(iv) Analyzing the qualitative data for insights
(v) Reporting on the insights derived from your analysis

QUALITATIVE DATA ANALYSIS METHODS

There are many different types of Qualitative Data Analysis (QDA), all of which serve different purposes and have unique strengths and weaknesses. In educational research, there are 6 commonly used QDA methods, cumulatively called “the big 6”. These include;

(i) Qualitative Content Analysis
(ii) Narrative Analysis
(iii) Discourse Analysis
(iv) Thematic Analysis
(v) Grounded Theory (GT)
(vi) Interpretive Phenomenological Analysis (IPA)

Content Analysis

This is possibly the most common and straightforward QDA method. At the simplest level, content analysis is used to evaluate patterns within a piece of content (for example, words, phrases or images) or across multiple pieces of content or sources of communication e.g. a collection of newspaper articles or political speeches.

With content analysis, you could, for instance, identify the frequency with which an idea is shared or spoken about – like the number of times a prominent politician is mentioned in a newspaper. Or you could identify patterns of deeper underlying interpretations – for instance, by identifying phrases or words in an interview conducted on students that indicate they have a negative attitude towards Mathematics.

Because content analysis can be used in such a wide variety of ways, it’s important to go into your analysis with a very specific question and objective, or you’ll get lost in the process. With content analysis, you’ll group large amounts of text into codes, summarize these into categories, and possibly even tabulate the data to calculate the frequency of certain concepts or variables. Because of this, content analysis provides a small splash of quantitative thinking within a qualitative method.

Naturally, while content analysis is widely useful, it’s not without its drawbacks. One of the main issues with content analysis is that it can be very time consuming, as it requires lots of reading and

re-reading of the texts. Also, because of its multidimensional focus on both qualitative and quantitative aspects, it is sometimes accused of losing important nuances in communication.

Content analysis also tends to concentrate on a very specific timeline and doesn't take into account what happened before or after that timeline. This isn't necessarily a bad thing though – just something to be aware of. So, keep these factors in mind if you're considering content analysis. Every analysis method has its drawbacks, so don't be put off by these – just be aware of them! Content analysis is used to evaluate patterns within a piece of content or across multiple pieces of content or sources of communication

Narrative Analysis

As the name suggests, narrative analysis is all about listening to people telling stories and analyzing what that means. Since stories serve a functional purpose of helping us make sense of the world, we can gain insights into the ways that people deal with and make sense of reality by analyzing their stories and the ways they're told.

You could, for example, use narrative analysis to explore whether how something is being said is important. For instance, the narrative of a prisoner trying to justify their crime could provide insight into their view of the world and the justice system. Similarly, analyzing the ways entrepreneurs talk about the struggles in their careers or cancer patients telling stories of hope could provide powerful insights into their mindsets and perspectives. In other words, narrative analysis is about paying attention to the stories that people tell - and more importantly, the way they tell them. Of course, the narrative approach has its weaknesses, just like all analysis methods.

Sample sizes are generally quite small due to the time-consuming process of capturing narratives. Because of this, along with the multitude of social and lifestyle factors which can influence a subject, narrative analysis can be quite difficult to reproduce in subsequent research. This means that it's difficult to test the findings of some of this research. Similarly, researcher bias can have a strong influence on the results here, so you need to be particularly careful about the potential biases you can bring into your analysis when using this method. Nevertheless, narrative analysis is still a very useful qualitative method – just keep these limitations in mind and be careful not to draw

broad conclusions. Narrative analysis is all about listening to people telling stories to gain insights into the ways that people deal with and make sense of reality.

Discourse Analysis

Discourse is simply a fancy word for written or spoken language or debate. So, discourse analysis is all about analyzing language within its social context. In other words, analyzing language such as a conversation, a speech etc. within the culture and society it takes place in. For example, you could analyze how a school messenger speaks to the principal, or parents speak to their teenage children about sex. To truly understand such conversations or speeches, the culture and history of those involved in the communication is important. For example, a janitor might speak more casually with a CEO in a company that emphasizes equality among workers.

Similarly, a politician might speak more about terrorism if there was a recent terrorist incident in the country. So, as you can see, by using discourse analysis, you can identify how culture, history or power dynamics (to name a few) have an effect on the way concepts are spoken about. So, if your research aims and objectives involve understanding culture or power dynamics, discourse analysis can be a powerful method. Because there are many social influences in how we speak to each other, the potential use of discourse analysis is vast. Of course, this also means it's important to have a very specific research question (or questions) in mind when analyzing your data and looking for patterns and themes, or you might land up going down a winding rabbit hole.

Discourse analysis can also be very time consuming as you need to sample the data to the point of saturation, in other words, until no new information and insights emerge. But this is, of course, part of what makes discourse analysis such a powerful technique. So, keep these factors in mind when considering this QDA method. Discourse analysis is all about analyzing language such as a conversation or a speech, within the culture and society it takes place in.

Thematic Analysis

Thematic analysis looks at patterns of meaning in a data set – for example, a set of interviews or focus group transcripts. But what exactly does that… mean? Well, a thematic analysis takes bodies of data (which are often quite large) and groups them according to similarities – in other words, themes. These themes help us make sense of the content and derive meaning from it.

Let’s take a look at an example:

With thematic analysis, you could analyze 100 reviews of a university student centre to find out what university students think about the food served therein. By reviewing the data, you would then identify the themes that crop up repeatedly within the data – for example, “fresh ingredients” or “friendly waiters”. So, as you can see, thematic analysis can be pretty useful for finding out about respondents’ experiences, views, and opinions. Therefore, if your research aims and objectives involve understanding people’s experience or view of something, thematic analysis can be a great choice.

Since thematic analysis is a bit of an exploratory process, it’s not unusual for your research questions to develop, or even change as you progress through the analysis. While this is somewhat natural in exploratory research, it can also be seen as a disadvantage as it means that data needs to be re-reviewed each time a research question is adjusted. In other words, thematic analysis can be quite time-consuming albeit for a good reason. So, keep this in mind if you choose to use thematic analysis for your educational research and budget extra time for unexpected adjustments. Thematic analysis therefore takes masses of qualitative data and groups them according to similarities (themes), which help us make sense of the content.

Grounded Theory (GT)

Grounded Theory is powerful qualitative analysis method where the intention is to create a new theory (or theories) using the data at hand, through a series of “tests” and “revisions.” For example, you could try to develop a theory about what factors influence students to read, watch a viral video. The important thing with grounded theory is that you go into the analysis with an open mind and let the data speak for itself, rather than dragging existing hypotheses or theories into your analysis. In other words, your analysis must develop from the ground up (hence the name).

In Grounded Theory, you start with a general overarching question about a given population – for example, undergraduate students. Then you begin to analyze a small sample – for example, ten undergraduate students in a university. Ideally, this sample should be reasonably representative of the broader population. You’d then interview these students to identify what factors lead them to watch the “*murife* don’t run!” video that was viral on Kenyan social media a while back.

After analyzing the interview data, a general hypothesis or pattern could emerge. For example, you might notice that undergraduate students are more likely to crack the "*murife don't run*" joke any time they witness their colleague facing a difficult question in a group discussion. From here, you'll look for another small sample – for example, five more graduate students in a different department – and see whether this pattern or this hypothesis holds true for them. If not, you'll look for commonalities and adapt your theory accordingly. As this process continues, the theory develops.

What's important with grounded theory is that the theory develops from the data – not from some preconceived idea. You need to let the data speak for itself. So, what are the drawbacks of grounded theory? Well, some argue that there's a tricky circularity to Grounded Theory. For it to work, in principle, you should know as little as possible regarding the research question and population, so that you reduce the bias in your interpretation. However, in many circumstances, it's also thought to be unwise to approach a research question without knowledge of the current literature. In other words, it's a bit of a "chicken or the egg" situation.

Regardless, grounded theory remains a popular (and powerful) option. Naturally, it's a very useful method when you're researching a topic that is completely new or has very little existing research about it, as it allows you to start from scratch and work your way from the ground up. Grounded theory is used to create a new theory (or theories) by using the data at hand, as opposed to existing theories and frameworks.

Interpretive Phenomenological Analysis (IPA)

IPA is designed to help you understand the personal experiences of a subject (for example, a person or group of people) concerning a major life event, an experience or a situation. This event or experience is the "phenomenon" or phenomena that make up the "P" in IPA. These phenomena may range from relatively common events – such as motherhood, or being involved in a car accident – to those which are extremely rare – for example, someone's personal experience in a refugee camp. So, IPA is a great choice if your research involves analyzing people's personal experiences of something that happened to them.

It’s important to remember that IPA is subject-centred. In other words, it’s focused on the experiences. This means that, while you’ll likely use a coding system to identify commonalities, it’s important not to lose the depth of experience or meaning by trying to reduce everything to codes. Also, keep in mind that since your sample size will generally be very small with IPA, you often won’t be able to draw broad conclusions about the generalizability of your findings. But that’s okay as long as it aligns with your research aims and objectives.

Another thing to be aware of with IPA is personal bias. While researcher bias can creep into all forms of research, self-awareness is critically important with IPA, as it can have a major impact on the results. For example, a researcher who was a victim of a crime himself could insert his own feelings of frustration and anger into the way he interprets the experience of someone who was kidnapped. So, if you’re going to undertake IPA, you need to be very self-aware or you could muddy the analysis. IPA can therefore help you understand the personal experiences of a person or group concerning a major life event, an experience or a situation.

Choosing the Right Qualitative Analysis Method

Selecting the right qualitative analysis method largely depends on your research aims, objectives and questions. In other words, the best tool for the job depends on what you’re trying to build. For example: Perhaps your research aims to analyze the use of words and what they reveal about the intention of the storyteller and the cultural context of the time. Perhaps your research aims to develop an understanding of the unique personal experiences of people that have experienced a certain event. Or perhaps your research aims to develop insight regarding the influence of a certain culture on its members.

As you can see, all these research aims are distinctly different, and therefore different analysis methods would be suitable for each one. Also, remember that each method has its own strengths, weaknesses and general limitations. No single analysis method is perfect. Therefore, it often makes sense to adopt more than one method (this is called triangulation), but this is, of course, quite time-consuming. As we’ve seen, these approaches all make use of coding and theme-generating techniques, but the intent and approach of each analysis method differ quite substantially. So, it’s really important to come into your research with a clear intention before you start thinking about which analysis method (or methods) to use. Start by reviewing your research aims, objectives and

research questions to assess what exactly you're trying to find out, and then select a method that fits. Never pick a method just because you like it or have experience using it. Your analysis method (or methods) must align with your broader research aims and objectives.

In Summary;

- Content analysis blends a little bit of quantitative into a primarily qualitative analysis.
- Narrative analysis is about analyzing how stories are told.
- Discourse analysis is about analyzing conversations and interactions.
- Thematic analysis is about identifying themes and patterns.
- Grounded theory starts from scratch with a specific question and uses the data alone to build a theory in response to that question.
- IPA is about understanding people's unique experiences of a phenomenon.
- No single analysis method is perfect, so it can often make sense to adopt more than one method. This is called triangulation, and is discussed in the next section.

Triangulation in Educational Research

This is the use of more than one research method when carrying out research so that different types of data will complement each other. Triangulation can take various forms;

i. *Investigator triangulation* - involves different researchers or interviewers. This form checks for observer bias, which is a potential threat to validity.
ii. *Data triangulation* - involves collecting both primary and secondary data from different people, at different times, at different places.
iii. *Methodological triangulation* – involves using different techniques within the same methods e.g. using both open and closed questionnaires or asking the same questions in different ways. It may also involve using different instruments to collect same data e.g. questionnaires as well as interviews.

QUANTITATIVE DATA ANALYSIS METHODS

Quantitative data analysis is the process of manipulating data in form of numbers or any data that can be quantified by converting it into numbers without losing any meaning. Such data is obtained from quantitative or continuous variables i.e. variables on interval or ratio scales of measurement.

Categorical variables like gender, ethnicity, or native language are examples of variables that can be "converted" into numbers without losing meaning – for example, English could equal 1, French 2, etc. This contrasts qualitative data analysis, where the focus is on words, phrases and expressions that can't be reduced to numbers.

Quantitative analysis is generally used for three purposes.

(i) To measure differences between groups e.g., the popularity of different subjects among students
(ii) To assess relationships between variables e.g., the relationship between weather conditions and voter turnout during elections
(iii) To test hypotheses in a scientifically rigorous way e.g., a hypothesis about the impact of an *AstraZeneca* vaccine on hospitalization rates among covid-19 patients.
(iv) To determine the association between the use a certain teaching method and students' academic performance.

As we mentioned, quantitative analysis is powered by statistical analysis methods. There are two main "branches" of statistical methods that are used – descriptive statistics and inferential statistics. In your research, you might only use descriptive statistics, or you might use a mix of both, depending on what you're trying to figure out. In other words, depending on your research questions, aims and objectives. I'll explain how to choose your methods later in the book.

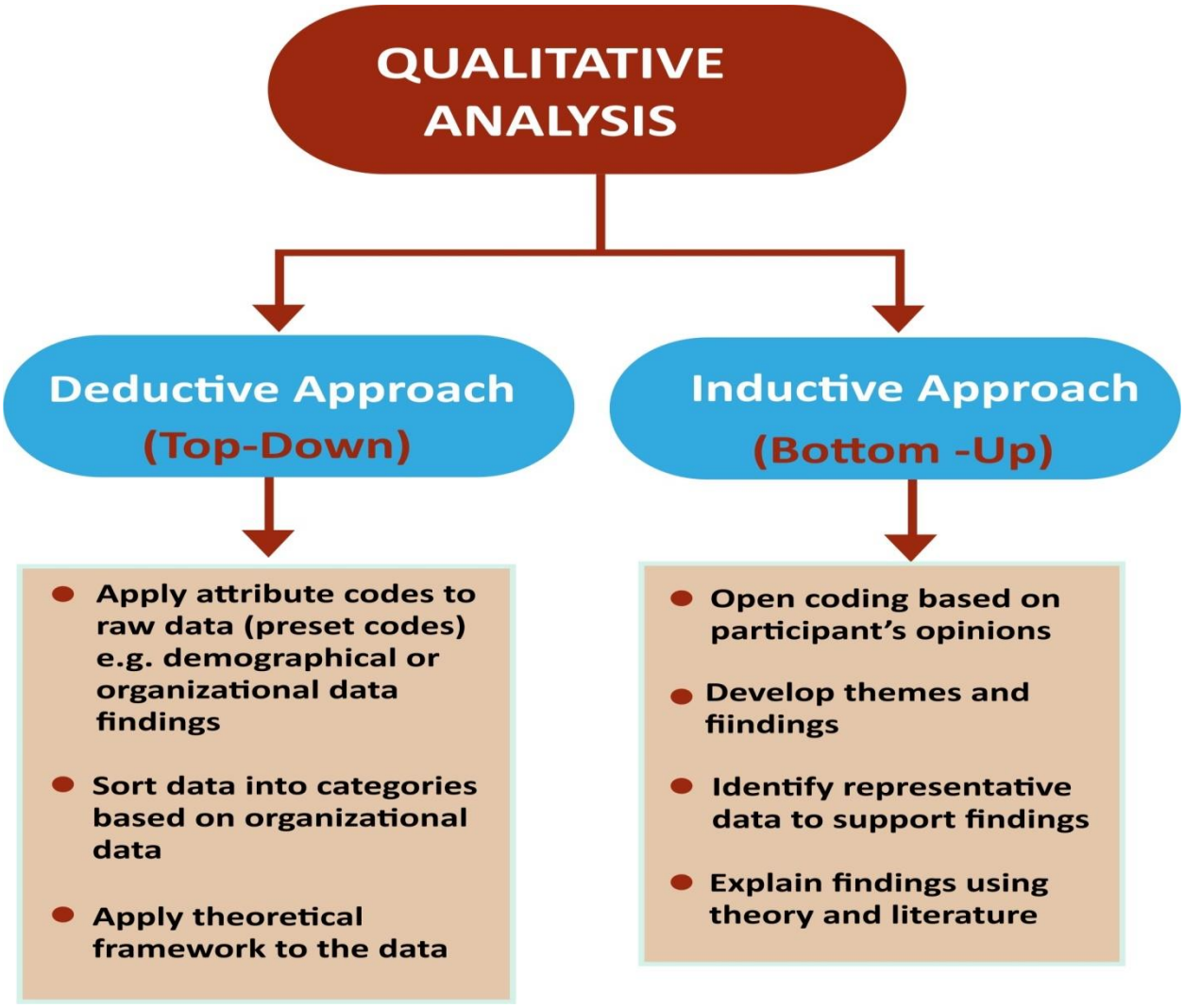


Figure 11: Approaches to Qualitative Analysis

DESCRIPTIVE STATISTICS IN EDUCATIONAL RESEARCH

Descriptive statistics serve a simple but critically important role in your research, which is to describe your data set, hence the name. In other words, they help you understand the details of your sample. Unlike inferential statistics, descriptive statistics do not aim at making inferences or predictions about the entire population. They're purely interested in the details of your specific sample. When you're writing up your analysis, descriptive statistics are the first set of stats you'll cover, before moving on to inferential statistics. But, that said, depending on your research objectives and research questions, they may be the only type of statistics you use. Descriptive analysis is also called a 'univariate analysis' since it is commonly used to analyze a single variable.

Importance of Descriptive Statistics

(i) They help you get both a macro and micro-level view of your data. In other words, they help you understand both the big picture and the finer details.
(ii) They help you spot potential errors in the data – for example, if an average is way higher than you'd expect, or responses to a question are highly varied, this can act as a warning sign that you need to double-check the data.
(iii) They help inform which inferential statistical techniques you can use, as those techniques depend on the descriptive measures like skewness, symmetry and normality of the data.

Let us look at some of the descriptive statistical measures or methods that are commonly used to analyze data in in educational research.

THE MODE

It is the value in a set of data which appears the highest number of times. Where the mode does not exist, it is usually estimated e.g. No mode exists in a distribution where values have the same frequency e.g.

2 3 4 5 7 8

Where one score has higher frequency than others in a distribution, the score is the mode e.g.

2 2 3 3 3 5 7 8

Mode = 3

THE MEDIAN

It is the point in a distribution that divides a set of data right in the middle e.g., below are the marks scored by 11 students in an achievement test.

16 11 12 10 13 17 12 14 12 14 15

Arrange the numbers from the lowest to the highest or vice versa

10 11 12 12 12 13 14 14 15 16 17

Get the median position by adding 1 to N and dividing the sum by 2 thus;

$M_d = \frac{N+1}{2} = \frac{11+1}{2} = 6$

Starting with the lowest value, count up to the sixth value, which is now the median.

10 11 12 12 12 **13** 14 14 15 16 17

Median = 13

If there is an even number of values (scores), the median is the average of the middle two.

When to use the Median in Educational Research

- ✓ When the data set has extreme values (in a skewed data).
- ✓ When the values at the end of a distribution are not known.
- ✓ When the exact midpoint of the distribution is wanted i.e. the 50% point

Advantages of the Median

i. The concept is easy to understand and interpret.
ii. It can be determined for any data set.
iii. It is not easily affected by extreme values in a data set.

Disadvantages of the Median

i. The data must first be arranged in an array (ascending or descending order).

ii. It lacks the useful mathematical properties i.e. it cannot be used for further computation.

THE MEAN

It is commonly referred to as the "average" and is obtained by taking the sum of all the values in a set of data divided by the number of values i.e.

$$\bar{x} = \frac{\sum x}{N}$$

For example, the mean of 12 8 25 26 10 is calculated as follows;

$$\bar{x} = \frac{12 + 8 + 25 + 26 + 10}{5} = 16.2$$

When to Use the Mean in Educational Research

- ✓ When the values (scores) are distributed symmetrically around a central point i.e. when the distribution is not badly skewed.
- ✓ When the measure of central tendency having the greatest stability is wanted.
- ✓ When other statistics (e.g. SD, coefficient of correlation etc.) are to be computed later.

Advantages of Using the Mean

i. It uses all values in the distribution hence its more stable.
ii. It is used to draw inferences (conclusions)

Disadvantages of the Mean

i. It is unduly affected by extreme values.
ii. It is difficult to compute compared to the mode and median.

THE RANGE

It refers to the difference between the highest and lowest values in a set of data.

Range = Highest value - Lowest value

Example**:** Find the range of the following data.

13 17 14 16 10 9 11

Solution: Range = 17 - 9 = 8

When to use the Range

- ✓ When the data are too scant or too scattered to justifying the computation of a more precise measure of variability.
- ✓ When knowledge of extreme scores or a total spread is all that is needed.

Advantages of the Range

i. It is easy to calculate
ii. It is easy to interpret

Disadvantages of the Range

i. It only takes two values into account and is therefore affected by extreme scores.
ii. It is unreliable when N is small or when there are large gaps in the frequency distribution.

THE STANDARD DEVIATION

The standard deviation (SD) is the most stable index of variability. It is represented by the symbol s or σ (sigma). The SD of a set of data is the square root of the variance.

$$= \sqrt{\frac{\sum \left(X - \bar{X}\right)^2}{N}}$$

Example**:** Calculate the standard deviation for the data below.

5 2 7 4 8

Solution:

X	$X - \bar{X}$	$\left(X - \bar{X}^2\right)$
2	-3.6	12.96
4	-1.6	2.56
5	-0.6	0.36
7	1.4	1.96
8	-2.4	5.76
		23.6

$$\text{S. D} = \sqrt{\frac{\sum\left(X - \bar{X}\right)^2}{N}}$$

$$= \sqrt{\frac{23.6}{5}}$$

$$= \sqrt{4.72}$$

$$= 2.17$$

Interpretation of S.D values is that the bigger the value, the larger the spread while the smaller the value, the less the spread.

When to use the SD

- ✓ When the statistic having the greatest stability is sought.
- ✓ When extreme deviation should exercise a proportionally greater effect upon the variability.
- ✓ When coefficient of correlation and other statistics are to be computed later.

SKEWNESS

Skewness refers to whether the distribution is symmetrical with respect to its dispersion from the mean. If on one side of the mean has extreme scores but the other does not, the distribution is said to be skewed.

(a) *Normal distribution* - If the dispersion of scores on either side of the mean are roughly symmetrical, the distribution is said to be normal. It is bell shaped.

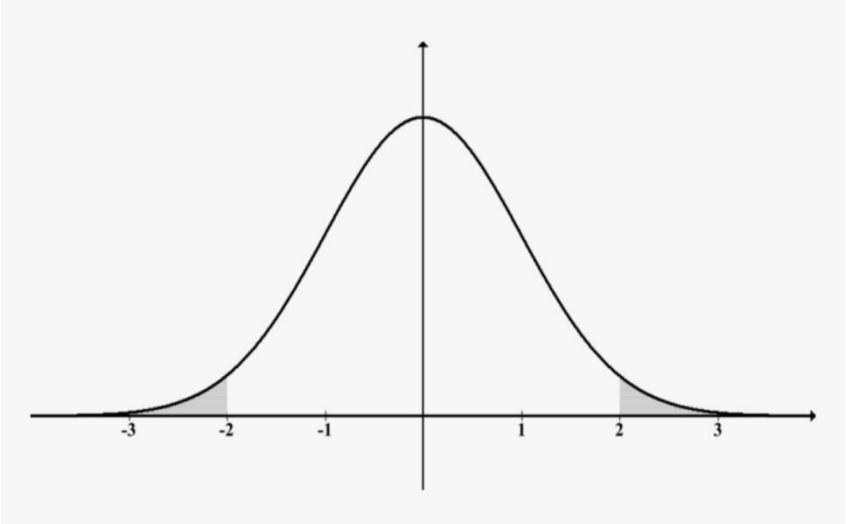


Figure 12: Shape of a Normal Distribution

(b) *Positively skewed distribution* - It is where the majority of scores are at the lower end of the distribution with a tail of scores at the upper end of the distribution.

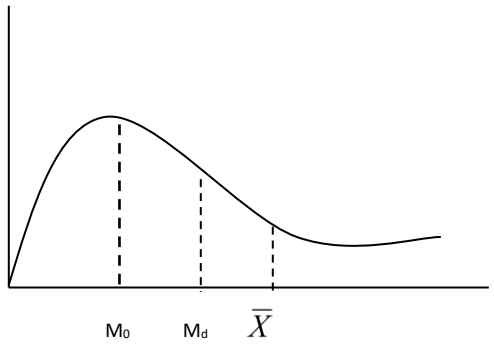


Figure 13: Shape of a Positively Skewed Distribution

In a positively skewed distribution, the Mean > Median > Mode. In a class test, it would mean that majority of the students scored below the class mean implying that;

- the test items may have been above the ability level of the students
- majority of the students are of below average ability
- the concept being tested may not have been well understood by the students

(c) *Negatively skewed distribution* - It refers to a distribution where the scores are clustered at the upper end of the scale with a tail of scores at the lower end of the distribution.

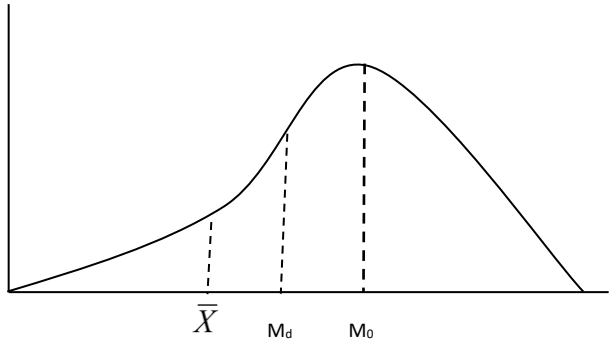


Figure 14: Shape of a Negatively Skewed Distribution

In a negatively skewed distribution, the Mean < Median < Mode. In a class test, it would mean that majority of the students scored above the class mean implying that;

- majority of the students may be of above average ability
- the test items may have been easy
- the concept being tested may have been well understood by the students

KURTOSIS

Kurtosis is a statistic that measures the extent to which a distribution contains outliers. It is a statistical measure used to describe the degree to which scores cluster in the tails or the peak of a frequency distribution. The peak is the tallest part of the distribution, and the tails are the ends of the distribution.

There are four kinds of kurtosis: - leptokurtic, platykurtic, Mesokurtic and bimodal distributions. Statisticians define these types relative to the normal distribution. Higher kurtosis values indicate that the distribution has more outliers falling relatively far from the mean. Distributions with smaller values have a lower tendency for producing extreme values. When you're assessing a sample, outliers have the greatest impact on this statistic.

Kurtosis is a unit-less measure of a distribution's shape. Consequently, analysts use the value for a normal distribution as the baseline for comparing other distributions. For instance, statisticians describe leptokurtic distributions as having higher kurtosis than the normal distribution. These

distributions have “heavy tails,” indicating that they have relatively long tails that contain more outliers. It is characterized by the piling up if scores in the centre of the distribution. Conversely, platykurtic distributions have “light tails” that are shorter and include fewer extreme values. It is characterized by scores spread across forming a “platform-like” distribution. The four types are graphs in the next two figures for comparison at one glance:

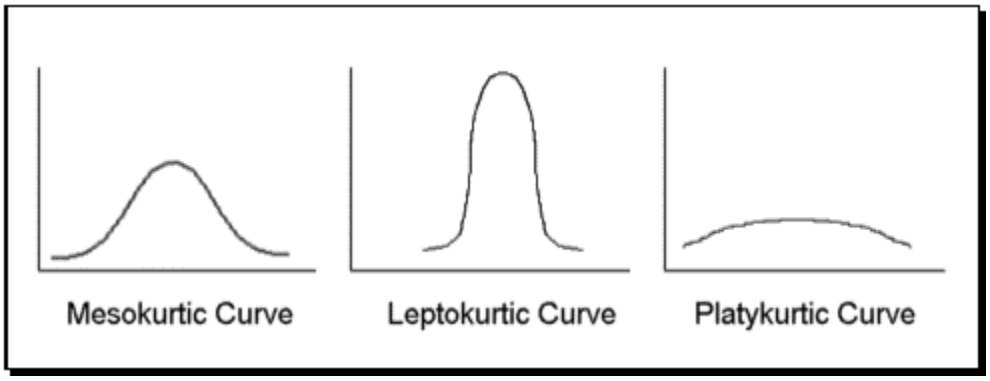


Figure 15: Shape of Mesokurtic, Leptokurtic and Platykurtic Distributions

The fourth and least common type of kurtosis is the Bimodal distribution. This is where a variable has a high concentration of frequencies around two separate values or where frequency distributions of two different populations are represented in single graph e.g. average adult height of males and females.

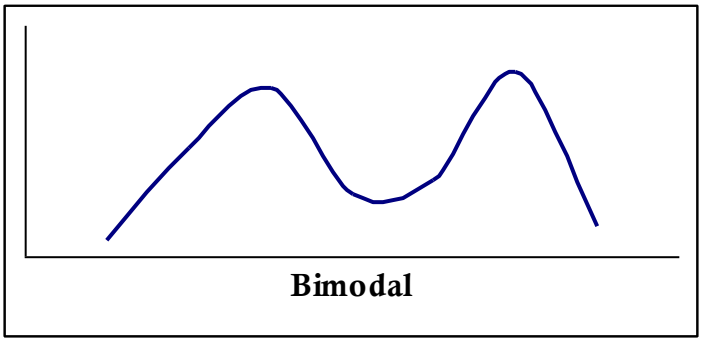


Figure 16: Shape of a Bimodal Distribution

INFERENTIAL STATISTICS AND HYPOTHESIS TESTING

As earlier mentioned, while descriptive statistics are all about the details of your specific data set, inferential statistics aim to make inferences about the population. In other words, you'll use inferential statistics to make predictions and generalizations about what you'd expect to find in the full population. What kind of predictions, you ask? Well, there are two common types of predictions that researchers try to make using inferential statistics; firstly, predictions about differences between groups - for example, differences between students' attitude towards a given subject grouped by their gender and secondly, relationships between variables - for example, the relationship between academic performance and the number of hours per week that a student studies privately.

Inferential statistics therefore allow you to connect the dots and make predictions about what you expect to see in the target population, based on what you observe in your sample data. For this reason, inferential statistics are used for hypothesis testing i.e. to test hypotheses that predict changes, relationships or differences. The process of testing a hypothesis follows the following general steps, regardless of the inferential test used:

Table 6: Steps of Hypothesis Testing

STEP	ACTION	DESCRIPTIONS
1	State the null (H_O) and alternate (H_A) hypotheses	H_O: There's no significant difference/association etc. H_A: There's a significant difference/association/relationship etc. **Note**: Use 2-tailed test if hypothesis doesn't suggest direction, Use 1-tailed test if it does
2	State level of significance	$\alpha = 0.05$ for educational research

3	Identify and compute the test statistic	Use a parametric test or its non-parametric equivalent if assumptions of the former aren't met e.g. t-test/U-test etc.
4	Formulate the decision rule	Find the critical value from the relevant normal distribution tables or the associated p-value if using SPSS
5	Make the decision	***Reject*** H_O if calculated test statistic ≥ critical value or if $p \leq \alpha$ for SPSS users ***Do not reject*** H_O *and accept* H_A if calculated test statistic < critical value or if $p > \alpha$ if using SPSS Note: only one of the 2 decisions is possible in hypothesis testing

Parametric Versus Non-Parametric Inferential Tests

There are a number of inferential statistical analysis methods that can be used to testing a null hypothesis, some of which are parametric while others are non-parametric. The key difference between the two broad types of tests is that parametric tests rely on statistical distributions in data whereas non-parametric tests do not depend on any distribution. Furthermore, parametric tests make assumptions about parameters of the distribution from which the sample is drawn, while non-parametric tests do not make any assumptions. Note that parametric tests are the most robust, only that their respective assumptions must be met, failure to which may lead to *type I* or *type II* error.

Type I error is rejecting a null hypothesis that is true in the population (false positive) while type II error is accepting a null hypothesis that is false in the population. Note that we also have *type III* error but it is irrelevant in the current discussion. However, it means correctly rejecting a false

null hypothesis but for the wrong reason. Non-parametric tests on the other hand are less robust but work equally well when the assumptions of their parametric equivalents are not met.

Non-parametric tests are valid for both non-normally distributed data and normally distributed data, so why not use them all the time? You may ask. It would seem prudent to use non-parametric tests in all cases, which would save one the bother of testing for normality. However, parametric tests are preferred for the following reasons;

i. In educational research, we are rarely interested in a significance test alone. We would also like to say something about the population from which the samples are drawn from. This is best achieved with estimates of parameters and confidence intervals.
ii. It is difficult to do flexible modelling with non-parametric tests, for example allowing or controlling for confounding factors using multiple regression.
iii. Parametric tests usually have more statistical power than their non-parametric equivalents. In other words, one is more likely to detect significant differences when they truly exist.

Assumptions of each test are very important when one is deciding on an appropriate test to use. Let us now discuss some of the most common inferential tests in educational research and their respective assumptions where applicable;

T-TEST

This is a parametric test that compares the means of two groups of random samples, to determine whether or not they're significantly different. This type of testing is very useful for understanding the effect of treatment on the experimental group. For example, you might want to compare the mean achievement test scores between two groups of students - one that has been taught face-to face and one that has been taught online, to assess whether online teaching (the treatment) is effective or not.

There are three sub-types of t-tests: - one sample t-test, paired samples t-test and independent samples t-test. Their non-parametric equivalents are; Kolmogorov-Smirnov test, Wilcoxon Signed Rank Test and Mann-Whitney U-test respectively. The latter trio are used when the assumptions of their parametric equivalents are not met by the data to be analyzed. When choosing a t-test, you will need to consider two things - whether the groups being compared come from a single

population or two different populations, and whether you want to test the difference in a specific direction.

One-Sample, Two-Sample or Paired T-Test?

If the groups come from a single population (e.g., measuring before and after an experimental treatment), perform a paired t test. This is a within-subjects design. If the groups come from two different populations (e.g. two different species, or people from two separate cities), perform a two-sample t test, also referred to as independent t test. This is a between-subjects design. If there is one group being compared against a standard value (e.g., comparing the acidity of a liquid to a neutral pH of 7), perform a one-sample t test.

One-Tailed or Two-Tailed T-Test?

If you only care whether the two populations are different from one another, perform a two-tailed t test. If you want to know whether one population mean is greater than or less than the other, perform a one-tailed t test. Due to limitations of the scope of this book, let us only look at in independent samples t-test in detail. When we take two samples from the same population, then the mean of the two samples may be identical. But when samples are taken from two different populations, then the mean of the sample may differ. In this case, it is used to draw conclusions about the means of two populations, and used to tell whether or not they are similar.

Assumptions in Independent Samples T-Test

The assumptions of t-test for independent samples are the same as the general assumptions for any other parametric test. These are;

- ✓ **Independence** - the dependent variable score for the two groups should be independent of the score for any other unit. Independence of observations: the data were collected using statistically valid sampling methods, and there are no hidden relationships among observations. The assumption of independence of scores is the most important assumption. When that assumption is violated, the resulting statistical tests can be misleading. The independence assumption is satisfied by the research design, which should feature random selection of subjects and random assignment to treatment groups. Randomization tends to distribute effects of extraneous variables evenly across groups.

- ✓ **Equality of variance** - in the population, the variance of dependent variable scores in each treatment group should be equal. (Equality of variance is also known as homogeneity of variance or homoscedasticity). Ensure variation within each group being compared is similar for every group by conducting the Levene's test of homogeneity of variances. This test for homogeneity of variance provides an F-statistic and a significance (p) value. We are primarily concerned with the significance value. If it is greater than 0.05 (i.e. $p > .05$), our group variances can be treated as equal. However, if $p < 0.05$, we have unequal variances and we have violated the assumption of homogeneity of variances.

If the Levene's test for equality of variances is statistically significant, which indicates that the group variances are unequal in the population, you can correct for this violation by not using the pooled estimate for the error term for the t-statistic, but instead using an adjustment to the degrees of freedom using the Welch-Satterthwaite method. In all reality, you will probably never have heard of these adjustments because SPSS statistics hides this information and simply labels the two options as "Equal variances assumed" and "Equal variances not assumed" without explicitly stating the underlying tests used. However, you can see the evidence of these tests as shown in the table, which displays differences in the t-statistic and the degrees of freedom when homogeneity of variance is not assumed.

Table 7: Sample T-Test Results

	Equal Variances assumed	Equal Variances not assumed	
Levene's Test for Equality of Variances	F	28.636	
	Sig.	.000	
t-test for Equality of Means	T	2.816	2.245
	DF	16	7.005
	Sig. (2-tailed)	.025	0.068

From the result of Levene's Test for Equality of Variances in the table, we can reject the null hypothesis that there is no difference in the variances between the groups and accept the alternative hypothesis that there is a statistically significant difference in the variances between groups. The

effect of not being able to assume equal variances is evident in the final column of the table where we see a reduction in the value of the t-statistic and a large reduction in the Degrees of Freedom (DF).

This has the effect of increasing the p-value above the critical significance level of 0.05. In this case, we therefore do not accept the alternative hypothesis and accept that there are no statistically significant differences between means. This would not have been our conclusion had we not tested for homogeneity of variances!

- ✓ **Normality** - in the population, dependent variable scores should be normally distributed within treatment groups. The values of the dependent variable follow a normal distribution. Violations of normality can be a problem when sample size is small and therefore, it is important to be on the lookout for any indication of non-normality. There are many different ways to check for normality. Given a small sample size however, the best option for testing normality is to look at the following descriptive statistics;

    - ✧ *Central tendency* - The mean and the median are summary measures used to describe central tendency. With a normal distribution, the mean = the median.
    - ✧ *Skewness* - is a measure of the asymmetry of a probability distribution. If observations are equally distributed around the mean, the skewness value is zero; otherwise, the skewness value is positive or negative. As a rule of thumb, skewness between -2 and +2 is consistent with a normal distribution.
    - ✧ *Kurtosis* - is a measure of whether observations cluster around the mean of the distribution or in the tails of the distribution. The normal distribution has a kurtosis value of zero. As a rule of thumb, kurtosis between -2 and +2 is consistent with a normal distribution.

    Note that for large samples, the *Shapiro-Wilk test* is the best option for statistically determining the normality of data. If your data does not fit all these assumptions, you can try a non-parametric alternative to the t test. However, the t-test is described as a robust test with respect to the assumption of normality. This means that some deviation away from normality does not have a large influence on Type I error rates.

The exception to this is if the ratio of the smallest to largest group size is greater than 1.5 (largest compared to smallest). If you find that either one or both of your group's data is not approximately normally distributed and groups sizes differ greatly, you have two options; (i) transform your data (via SPSS) so that the data becomes normally distributed, or (ii) run the Mann-Whitney U-test which is a non-parametric test that does not require the assumption of normality.

Procedure for Performing Independent Samples T-Test

Step 1: Set up the hypothesis. The null hypothesis assumes that the means of the two groups are not significantly different.

Step 2: State the alternative hypothesis: It assumes that the means of the two groups are significantly different.

Step 3: Calculate the value of the independent sample t-test by using the formula:

$$t = \frac{mean1 - mean2}{\sqrt{(\frac{var1}{n1} + \frac{var2}{n2})}}$$

Where: Var1 and Var2 = variance of each group, N_1 and N_2 = Number of respondents in each sample group. Note that the "t" must always be in lowercase.

Step 5: Make a decision. In hypothesis testing for the independent sample t-test, statistical decisions are made about whether or not the two population means are identical. Compare the calculated value of the independent sample t-test with the critical value from the t-test distribution table (see appendix 1) at the required degrees of freedom.

Degrees of Freedom = $N_1 + N_2 - 2$

If the calculated value of the independent sample t-test is greater than the critical value of the 0.05 significance level, you will reject the null hypothesis and say that the means of the two groups are different. If the calculated value of the

independent sample t-test is less than the table value, then you will say that the means of the two groups are the same.

Note that for all t-tests, a larger t-value shows that the difference between group means is greater than the pooled standard error, indicating a more significant difference between the groups.

Scholarly Reporting Results of T-Test

When reporting the result of an independent t-test, you need to include the t-statistic value, the degrees of freedom (DF) and the significance value of the test (p-value). The format of the test result is: t (DF) = t-statistic, p = significance value. Descriptive statistics may also be included to help the reader get a clearer picture.

Example:

The difference in academic motivation between the experimental group (M = 71.46; SD = 8.206) and the control group (M = 75.54; SD = 8.569) was significant (t (30) = 33.7190; p = .007). For higher level studies like doctoral dissertations, the researcher should provide enough information for readers to fully understand the results when you have run an independent t-test, you should include the result of normality tests, Levene's Equality of Variances test, the two group means and standard deviations, the actual t-test result and the direction of the difference (if any). In addition, you might also wish to include the difference between the groups along with a 95% confidence interval.

CHI-SQUARE

This is a non-parametric inferential test that is used to show association between two categorical variables. The chi-squared (x^2) statistic is a single number that tells you how much difference exists between your observed counts and the counts you would expect if there were no relationship at all in the population. There are a few variations of the chi-square statistic. Which one you use depends upon how you collected the data and which hypothesis is being tested. However, all of the variations use the same idea, which is that you are comparing your expected values with the values you actually collect.

One of the most common forms can be used for contingency tables. *The Chi-Square test of independence* is the most common in educational research data analysis and is used to determine if there is a significant relationship between two nominal (categorical) variables. The frequency of each category for one nominal variable is compared across the categories of the second nominal variable. The data can be displayed in a contingency table where each row represents a category for one variable and each column represents a category for the other variable.

First you have to calculate the expected value of the two nominal variables using this formula:

Expected value = (Row total × column total ÷ sum total

After calculating the expected value, we will apply the following formula to calculate the value of the chi-square test of independence statistic (x^2) using the formula:

$$x^2 = \sum \frac{(O-E)^2}{E}$$

Where; x^2 = chi-square test of independence

O = observed value of two nominal variables

E = expected value of two nominal variables

Degree of Freedom is calculated by using the formula:

DF = (r-1) (c-1)

Where; DF = Degree of Freedom, r = number of rows, c = number of columns

We then compare the calculated x2 with the critical value in the chi-square distribution (see appendix 1) and make the decision accordingly.

For example:

A researcher wants to examine the relationship between marital status of teachers in Murang'a municipality (single, married or divorced) and their preferred choice of supermarket for shopping in Murang'a town (between Magunas Teketeke and Mathais). The chi-square test of independence can be used to examine this relationship.

The null hypothesis for this example is;

H_O: "There is no statistically significant association between teachers' marital status and their preferred supermarket for shopping in Murang'a town".

The alternative hypothesis is;

HA: "There is a statistically significant association between marital status and preferred supermarket".

First we have to calculate the expected value of the two nominal variables. The table below is the contingency table from sample data that was collected by the researcher.

Table 8: Sample Contingency Table

OBSERVED frequencies (numbers actually seen in the 2 supermarkets)			
	Mathais	Magunas	Total
Single	40	30	70
Married	89	67	156
Divorced	3	7	10
Total	132	104	236

We can now proceed and calculate the expected values of the two nominal variables thus:

Table 9: Sample Expected Frequencies Table

EXPECTED frequencies (Row total × column total ÷ sum total			
	Mathais	Magunas	Total
Single	39	31	70
Married	87	69	156
Divorced	6	4	10
Total	132	104	236

The calculated values of the x^2 statistic are as follows;

Table 10: Sample Chi-statistics Table

$(O-E)^2/E$			
	Mathais	Magunas	Total
Single	0.0256	0.0323	0.0579
Married	0.0460	0.0580	0.1040
Divorced	1.5000	2.2500	3.7500
Total	1.5716	2.3403	**3.9119**

$$x^2 = \sum \frac{(O-E)^2}{E} = 3.9119$$

$$DF = (r-1)\,(c-1) = (3-1)\,(2-1) = 2$$

NB: the values of r and c must not include the totals'.

The null hypothesis for chi-square assumes that there is no association between the two variables thus:

H_O: Teachers' preferred supermarket is independent of their marital status.

The alternative hypothesis for chi-square assumes that two variables are dependent of each other thus:

H_O: Teachers' preferred supermarket depends on their marital status

Hypothesis testing for the chi-square test of independence as it is for other tests like t-test and ANOVA, where a test statistic is computed and compared to a critical value. The critical value for the chi-square statistic (x^2 calculated) at **0.05** level of significance and **2** degrees of freedom from the x^2 table (see appendix 1) is **5.99** as the table shows.

Since the calculated chi-square test statistic 3.909 is less than the critical value 5.99, we do not reject the null hypothesis. This implies that the sampled teachers go to either supermarket regardless of their marital status. In other words, there's no group of teachers (single, married, divorced) that prefer a particular supermarket in Murang'a town for their shopping.

Reporting Chi-Square Analysis Results

In educational research, inferential statistics are resorted in APA format, whose requirements for citing statistical test results are quite precise, so you need to pay attention to the basic format, and also to the placing of brackets, punctuation, italics, and the like. This is the basic format for reporting a chi-square test result (where the bold text means you substitute in the appropriate value from your study).

*[x^2 (**degrees of freedom**, N = **sample size**) = **chi-square statistic value**, p = **p-value**].*

Example:

Imagine we conducted a study that looked at whether there is a link between gender and the academic achievement in Kiswahili. We might report the results like this:

> A chi-square test of independence was performed to examine the relation between students' gender and their academic achievement in Kiswahili The relation between these variables was significant, *[χ^2 (8, N = 84) = 19.7, p = .012]*. Female students were more likely than men to perform well in Kiswahili.

CORRELATION

This type of analysis assesses the direction and strength of association between two variables. In other words, if one variable increases, does the other variable also increase, decrease or stay the same? For example, if the average lecture attendance goes up, do semester grades increase too? We'd expect some sort of linear relationship between these two variables intuitively, but correlation analysis allows us to measure it relationship scientifically.

The most common correlation analysis tool is the Pearson's Product-Moment Correlation Coefficient. This correlation is a parametric test that is used to measure the linear relationship between two continuous variables. However, in a non-linear relationship, this correlation coefficient may not always be a suitable measure of dependence. The possible range of values for

the correlation coefficient is -1.0 to +1.0. In other words, the values cannot exceed 1.0 or be less than -1.0. A correlation of -1.0 indicates a perfect negative correlation, while a correlation of +1.0 indicates a perfect positive correlation.

If the correlation coefficient value is greater than zero, it is a positive relationship i.e. the variables are directly proportional to each other. Conversely, if the value is less than zero, it is a negative relationship i.e. the variables are directly proportional to each other. A value of zero indicates that there is no relationship whatsoever between the two variables.

The most common correlation is the Pearson's Product Moment Correlation Coefficient (PPMCC). The PPMCC is calculated using the formula;

$$r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

r_{xy} is based on an interval scale and the two variables must be similar. The points on the scatter diagram should be uniformly distributed. It provides a linear relationship. Interpretation of r_{xy} values is as follows:

- + 1.0 is described as perfect, direct relationship
- + 0.5 is described as moderate, direct relationship
- 0.0 indicates no relationship
- - 0.5 is described as moderate, inverse relationship
- - 1.0 is described as perfect, inverse relationship

Assumptions of PPMCC

Given how simple Karl Pearson's Coefficient of Correlation is, the assumptions behind it are often forgotten. It is important to ensure that the assumptions hold true for your data, else the Pearson's Coefficient may be inappropriate. The assumptions and requirements for computing Karl Pearson's Coefficient of Correlation are as follows;

i. **Normality** means that the data sets to be correlated should approximate the normal distribution. In such normally distributed data, most data points tend to hover close to the mean.

ii. **Homoscedascity** comes from the Greek prefix hom, along with the Greek word skedastikos, which means 'able to disperse'. Homoscedascity means 'equal variances'. It means that the size of the error term is the same for all values of the independent variable. If the error term, or the variance, is smaller for a particular range of values of independent variable and larger for another range of values, then there is a violation of homoscedascity. It is quite easy to check for homoscedascity visually, by looking at a scatter plot. If the points lie equally on both sides of the line of best fit, then the data is homoscedastic.

iii. **Linearity** simply means that the data follows a linear relationship. Again, this can be examined by looking at a scatter plot. If the data points have a straight line (and not a curve) relationship, then the data satisfies the linearity assumption.

iv. **Continuous variables** are those that can take any value within an interval. Ratio variables are also continuous variables. To compute Karl Pearson's Coefficient of Correlation, both data sets must contain continuous variables. If even one of the data sets is ordinal, then Spearman's Coefficient of Rank Correlation would be a more appropriate measure.

v. **Paired observations** mean that every data point must be in pairs. That is, for every observation of the independent variable, there must be a corresponding observation of the dependent variable. We cannot compute correlation coefficient if one data set has 12 observations and the other has 10 observations.

vi. **No spurious outliers** present in the data. While statistically there's no harm if the data contains outliers, they can significantly skew the correlation coefficient and make it inaccurate. When does a data point become an outlier? In general, a data point that's beyond +3.29 or -3.29 standard deviations away, it is considered to be an outlier. Outliers are easy to spot visually from the scatter plot.

To verify most of these assumptions, a scatter plot is invaluable. It is therefore suggested that a scatter plot be created first before computing the correlation coefficient.

Worked out example:

The following scores were obtained by 5 students in their two semester examinations. Use PPMCC to determine whether there was any association between the two tests.

Student No.	1	2	3	4	5
SEM 1	56	78	90	43	33
SEM 2	62	85	80	40	28

Solution:

$$\bar{x} = \frac{300}{5} = 60$$

$$\bar{y} = \frac{295}{5} = 59$$

Candidate No.	x	y	$(x - \bar{x})$	$(x - \bar{x})^2$	$(y - \bar{y})$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	56	62	-4	16	3	9	-12
2	78	85	18	324	26	676	468
3	90	80	30	900	21	441	630
4	43	40	-17	289	-19	361	323
5	33	28	-27	729	-31	961	837
$\sum$	300	295		2258		2448	2246

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} = \frac{2246}{\sqrt{2258 \times 2448}} = 0.953$$

The interpretation is that there is a strong positive relationship between the 1st semester and second semester examination scores. This means that a student who scored highly in the first semester examinations also scored highly in the second semester examinations. This can also be interpreted to mean that the tests are internally consistent/reliable or that the independent variable (Exam I) has the potential for predicting the dependent variable (Exam II).

Scholarly Reporting of Correlations

To report the results of a correlation, include the following: the degrees of freedom in parentheses, the r value (the correlation coefficient), the p value.

Example:

> It was found that a strong positive correlation exists between average hours of study and academic performance, *[r (357) = .42, p < .001]*.

Although we have a significant relationship or associations between the two variables, it does not mean that one variable causes the other. Researchers need to remember that there may be a third variable (covariate) present that affects both variables accounted for in a correlation coefficient. Therefore, the existence of a significant correlation coefficient for a pair of variables by itself does not imply causation between the two variables.

LINEAR REGRESSION ANALYSIS

Linear regression is the next step up after performing a correlation. It is used when we want to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable (or sometimes, the outcome variable) while the variable we are using to predict the other variable's value is called the independent variable (or sometimes, the predictor variable). For example, you could use linear regression to understand whether exam performance can be predicted based on revision time. If you have two or more independent variables, rather than just one, you need to use multiple regressions, otherwise you use simple linear regression for one independent variable against one dependent variable.

Assumptions of Linear Regression

When you choose to analyze your data using linear regression, part of the process involves checking to make sure that the data you want to analyze can actually be analyzed using linear regression. You need to do this because it is only appropriate to use linear regression if your data "passes" seven assumptions that are required for linear regression to give you a valid result. In practice, checking for these seven assumptions just adds a little bit more time to your analysis, requiring you to click a few more buttons in SPSS statistics when performing your analysis, as

well as think a little bit more about your data, but it is not such a daunting task. The following are the assumptions of linear regression;

i. Your dependent variable should be measured at the continuous level (i.e. it is either an interval or ratio variable). Examples of continuous variables include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth.
ii. Your independent variable should also be measured at the continuous level (i.e. it is either an interval or ratio variable).
iii. There needs to be a linear relationship between the two variables. Whilst there are a number of ways to check whether a linear relationship exists between your two variables, I suggest creating a scatter-plot using SPSS statistics, where you can plot the dependent variable against your independent variable and then visually inspect the scatter-plot to check for linearity.
iv. There should be no significant outliers. An outlier is an observed data point that has a dependent variable value that is very different to the value predicted by the regression equation. As such, an outlier will be a point on a scatter-plot that is vertically far away from the regression line indicating that it has a large residual. The problem with outliers is that they can have a negative effect on the regression analysis (e.g. reduce the fit of the regression equation) that is used to predict the value of the dependent (outcome) variable based on the independent (predictor) variable.

    This will change the output that SPSS statistics produces and reduce the predictive accuracy of your model. Fortunately, when using SPSS statistics to run a linear regression on your data, you can easily include criteria to help you detect possible outliers, called "case-wise diagnostics".
v. Independence of observations, which you can easily check using the Durbin-Watson statistic, which is a simple test to run using SPSS statistics.
vi. Your data needs to show homoscedasticity, which is where the variances along the line of best fit remain similar as you move along the line.
vii. Finally, you need to check that the residuals (errors) of the regression line are approximately normally distributed. Two common methods to check this assumption include using either a histogram (with a superimposed normal curve) or a normal P-P Plot.

Interpreting and Reporting the Output of Multiple Regression Analysis

The reason why many researchers avoid regression analysis is the many tables that are produced once the SPSS analysis is complete. SPSS statistics will generate quite a few tables of output for a multiple regression analysis. However, there's no cause for alarm, as only the three main tables are required to understand results from the multiple regression procedure, assuming that no assumptions have been violated. In this book, I focus only on the three main tables you need to understand your multiple regression results, assuming that your data has already met the seven assumptions required for multiple regressions to give you a valid result:

Determining how Well the Regression Model Fits

The first table of interest is the Model Summary table. This table provides the R, R^2, adjusted R^2, and the standard error of the estimate, which can be used to determine how well a regression model fits the data:

Table 11: Sample Model Summary Table for Regression Analysis

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.846[a]	.716	.705	7.956
a. Predictors: (Constant), Height, Age, Weight				

The "R" column represents the value of R, the multiple correlation coefficient. R can be considered to be one measure of the quality of the prediction of the dependent variable, in this case, academic performance. A value of 0.846, in this example, indicates a good level of prediction. The "R Square" column represents the R^2 value (also called the *coefficient of determination*), which is the proportion of variance in the dependent variable that can be explained by the independent variables (technically, it is the proportion of variation accounted for by the regression model above and beyond the mean model). You can see from our value of 0.716 that our independent variables explain 71.6% of the variability of our dependent variable, academic performance. However, you also need to be able accurately report your data.

Statistical Significance

The second important output in regression analysis is the ANOVA table. The F-ratio in the ANOVA table, tests whether the overall regression model is a good fit for the data.

Table 12: Sample ANOVA Table for Regression Analysis

ANOVA[A]						
Model		Sum of Squares	DF	Mean Square	F	Sig.
1	Regression	8787.581	3	2929.194	46.278	.002[b]
	Residual	21963.501	347	63.295		
	Total	30751.083	350			
A. Dependent Variable: Academic Performance Score						
B. Predictors: (Constant), Age, Weight, Height						

The ANOVA table above shows that the independent variables statistically significantly predict the dependent variable, F (3, 347) = 46.278, p = .002 (this ANOVA result implies that the regression model is a good fit of the data).

Estimated Model Coefficients

From the previous example, the general form of the equation to predict academic performance from age, weight and height is:

Predicted Acad. performance = 4.16 + (0.415 x age) + (0.186 x weight) + (0.103 x height)

This is obtained from the Coefficients table, as shown in Table 13:

Table 13: Sample Coefficients Table for Regression Analysis

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.160	3.040		-1.368	.172
	Age	.313	.038	.415	8.266	.000
	Weight	.131	.035	.186	3.696	.000
	Height	.124	.055	.103	2.249	.025
a. Dependent Variable: Academic Performance						

Unstandardized coefficients indicate how much the dependent variable varies with an independent variable when all other independent variables are held constant. Consider the effect of age in this example. The unstandardized coefficient, B, for age is equal to 0.313 from the coefficients table. This means that for each one-year increase in age, there is an increase in academic performance score of 0.313 marks.

Statistical Significance of the Independent Variables

You can test for the statistical significance of each of the independent variables. This tests whether the unstandardized (or standardized) coefficients are equal to 0 (zero) in the population. If $p < .05$, you can conclude that the coefficients are statistically significantly different from 0 (zero). The t-value and corresponding p-value are located in the "t" and "Sig." columns respectively.

You can see from the "Sig." column that all independent variable coefficients are statistically significantly different from 0 (zero), as they are all less than 0.05 the set alpha. Although the intercept, B, is tested for statistical significance, this is rarely an important or interesting finding. Putting it all together, you could write up the results as follows:

> *A multiple regression was run to predict academic performance from age, height and weight. These variables statistically significantly predicted academic performance, [$F (3, 347) = 46.278$, $p = .002$, $R^2 = .716$]. All three variables added statistically significantly to the prediction at $\alpha = .05$, $p < .05$.*

ONE-WAY ANOVA

ANOVA, which stands for Analysis of Variance, is a statistical test used to analyze the difference between the means of more than two groups. So it's basically a t-test for more than two groups of respondents. The non-parametric equivalent of one-way ANOVA is Kruskal Wallis-H, which works the same but is robust when assumptions of the latter are not met by the data to be analyzed. A one-way ANOVA uses one independent categorical variable to form the groups, while a two-way ANOVA uses two independent categorical variables. One-way ANOVA is used when you have collected data about one categorical independent variable and one quantitative dependent variable. The independent variable should have at least three levels (i.e. at least three different groups or categories). ANOVA tells you if the dependent variable changes according to the level of the independent variable.

For example:

> *Your independent variable is method of teaching. In your experimental research design, you teach three different student classes with lecture method, group discussions and computer assisted instructions (each group with 1 one of those) to find out if there is a difference in students' achievement according to the method used the teach them a particular topic or subject.*

The null hypothesis (H_O) of ANOVA is that there is no significant difference between the group means. The alternative hypothesis (H_A) is that at least one group differs significantly from the overall mean of the dependent variable. Note that if you only want to compare two groups, a t-test should be used instead.

ANOVA determines whether the groups created by the levels of the independent variable are statistically different by calculating whether the means of the treatment levels are different from the overall mean of the dependent variable. If any of the group means is significantly different from the overall mean, then the null hypothesis is rejected. It uses the F-test for statistical significance. This allows for comparison of multiple means at once, because the error is calculated for the whole set of comparisons rather than for each individual two-way comparison (which would happen with a t test).

The F-test compares the variance in each group mean from the overall group variance. If the variance within groups is smaller than the variance between groups, the F-test will find a higher F value, and therefore a higher likelihood that the difference observed is real and not due to chance.

Assumptions of ANOVA

The assumptions of ANOVA test are the same as the general assumptions for any other parametric test. These are;

- ✓ **Independence** - the dependent variable score for each group should be independent of the score for any other unit. The assumption of independence of scores is the most important assumption. When this assumption is violated, the resulting statistical tests can be misleading. Therefore, if your data fails to meet this assumption because you have a confounding variable that you need to control for statistically, use an ANOVA with blocking variables (ANCOVA). The independence assumption is satisfied by the research design, which should feature random selection of subjects and random assignment to treatment groups. Randomization tends to distribute effects of extraneous variables evenly across groups.

- ✓ **Equality of variance** - in the population, the variance of dependent variable scores in each treatment group should be equal. (Equality of variance is also known as homogeneity of variance or homoscedasticity). Ensure variation within each group being compared is similar for every group. If the variances are different among the groups, then ANOVA probably isn't the right fit for the data. The Kruskal-Wallis H test should be done instead, as does not give an invalid result even if the assumption of variance is not met by your data.

- ✓ **Normality** - in the population, dependent variable scores should be normally distributed within treatment groups. The values of the dependent variable follow a normal distribution. Violations of normality can be a problem when sample size is small and therefore, it is important to be on the lookout for any indication of non-normality. There are many different ways to check for normality. Given a small sample size however, the best option for testing normality is to look at the following descriptive statistics;

- ✧ *Central tendency* - The mean and the median are summary measures used to describe central tendency. With a normal distribution, the mean = the median.
- ✧ *Skewness* - is a measure of the asymmetry of a probability distribution. If observations are equally distributed around the mean, the skewness value is zero; otherwise, the skewness value is positive or negative. As a rule of thumb, skewness between -2 and +2 is consistent with a normal distribution.
- ✧ *Kurtosis* - is a measure of whether observations cluster around the mean of the distribution or in the tails of the distribution. The normal distribution has a kurtosis value of zero. As a rule of thumb, kurtosis between -2 and +2 is consistent with a normal distribution.

Performing One-Way ANOVA

While you can perform One-Way ANOVA by hand, it is difficult to do so with a large sample size, like it might be the case for educational research, which mostly involves at least 40 students per class. Imagine having 10 such groups which you want to compare! This is where SPSS comes in handy. We'll therefore do without an example here. Having confirmed that the critical assumptions are tenable, we can proceed with a one-way analysis of variance. That means taking the following steps:

Step 1: Write statistical hypotheses to be tested by experimental data.

Step 2: Specify a significance level for the hypothesis test.

Step 3: Compute the F-test statistic and the degrees of freedom associated with it

The degrees of freedom are calculated as follows;

$DF_1 = K - 1$ and $DF_2 = N - K$

Where K = the number of comparison groups and

N = the total number of respondents in all the groups.

Note the p-value for the test statistic obtained.

Step 5: Accept or reject the null hypothesis, based on the critical or p-value at the 0.05 significance level.

About Critical and P-Values

In any inferential test or experiment, a p-value is the probability of obtaining a result more extreme than the observed experimental outcome, assuming the null hypothesis is true. With ANOVA, the F-ratio is the observed experimental outcome that we are interested in. So, the P-value would be the probability that an F statistic would be more extreme (i.e. bigger) than the actual F ratio computed from experimental data (critical value). Here's the decision rule for accepting or rejecting the null hypothesis:

- *If the P-value obtained > the significance level (0.05),* ***accept*** H_O.
- *If the P-value is equal to or less than 0.05,* ***reject*** H_O.

OR for manual computations;

- *If the F-ratio < the critical value,* ***accept*** H_O.
- *If the F-ratio is equal to or greater than the critical value,* ***reject*** H_O.

Post-hoc Testing and Reporting of ANOVA Results

ANOVA will tell you if there are differences among the levels of the independent variable, but not which of the differences are significant. To find how the treatment levels differ from one another, you need to perform a Tukey's HSD (Honestly-Significant Difference) post-hoc test, which is within the SPSS program under ANOVA.

Tukey's test runs pairwise comparisons among each of the groups, and uses a conservative error estimate to find the groups which are statistically different from one another. When reporting the results of an ANOVA, include a brief description of the variables you tested, the F-value, degrees of freedom, and p-values for each dependent variable and explain what the results imply.

Scholarly Reporting One-Way ANOVA Results

To report the results of an ANOVA, include the following; the degrees of freedom (between groups, within groups) in parentheses, the F value (also referred to as the F statistic), the p value.

Example: Reporting from fictitious One-way ANOVA results from a study whose independent variable was teaching method and dependent variable being attitude of learners towards Mathematics:

> *A one-way ANOVA demonstrated that the effect of teaching method was significant for learners' attitude towards Mathematics, [F (2, 78) = 4.58, p = .013].*

ANCOVA

ANCOVA stands for Analysis of Covariance. The one-way ANCOVA can be thought of as an extension of one-way ANOVA to incorporate a third variable called a covariate. Like the one-way ANOVA, the one-way ANCOVA is used to determine whether there are any significant differences between two or more independent groups on a dependent variable. However, whereas the ANOVA looks for differences in the group means, the ANCOVA looks for differences in adjusted means (i.e. adjusted for the covariate). As such, compared to the one-way ANOVA, the one-way ANCOVA has the additional benefit of allowing you to "statistically control" for the third variable (sometimes known as a "confounding variable"), which you believe will affect your results. This third variable that could be confounding your results should be included in your one-way ANCOVA analysis.

Note that it is possible to have more than one covariate and although covariates are traditionally measured on a continuous scale, they can also be categorical. However, when the covariates are categorical, the analysis is not often called ANCOVA. In addition, the "one-way" part of one-way ANCOVA refers to the number of independent variables. If you have two independent variables rather than one, you could run a two-way ANCOVA.

Assumptions of ANCOVA

Before I introduce you to these assumptions, do not be surprised if, when analyzing your own data using SPSS Statistics, one or more of these assumptions is violated (i.e., is not met). This is not uncommon when working with real-world data rather than textbook examples, which often only show you how to carry out a one-way ANCOVA when everything goes well! However, don't worry. Even when your data fails certain assumptions, there is often a solution to overcome this. First, let's take a look at these assumptions;

i. The dependent variable and covariate variable should be measured on a continuous scale (i.e. interval or ratio level). Examples of variables that meet this criterion include revision time of learners (measured in hours), intelligence (measured using IQ score), Academic achievement (measured from 0 to 100) etc. As stated earlier, you can have categorical covariates (e.g. categorical variables such as gender, which has two categories, but the analysis is not usually referred to as an ANCOVA in this situation.

ii. Your independent variable should consist of two or more categorical, independent groups. Examples of independent variables that meet this criterion include; gender (male and female), school type (mixed, boys and girls), CBC pathway (STEM, Social Sciences and Humanities) etc.

iii. You should have independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves. For example, there must be different participants in each group with no participant being in more than one group. This is more of a study design issue than something you can test for, but it is an important assumption of a one-way ANCOVA. If your study fails this assumption, you will need to use another statistical test instead of a one-way ANCOVA (e.g., a repeated measures design).

iv. There should be no significant outliers. Outliers are simply data points within your data that do not follow the usual pattern (e.g., in a study of 100 students' IQ scores, where the mean score was 108 with only a small variation between students, one student had a score of 156, which is very unusual, and may even put her in the top 1% of IQ scores nationally). The problem with outliers is that they can have a negative effect on the one-way ANCOVA, reducing the validity of your results.

v. Your residuals should be approximately normally distributed for each category of the independent variable. We talk about the ANCOVA only requiring approximately normal residuals because it is quite robust to violations of normality, meaning that the assumption can be violated to a degree and still provide valid results. You can test for normality using two Shapiro-Wilk tests of normality; one to test the within-group residuals and one to test the overall model fit. Both of these are easily tested for using SPSS Statistics.

vi. There needs to be homogeneity of variances. You can test this assumption in using Levene's test for homogeneity of variances.

vii. The covariate should be linearly related to the dependent variable at each level of the independent variable. You can test this assumption in SPSS Statistics by plotting a grouped scatter plot of the covariate, post-test scores of the dependent variable and independent variable.

viii. There needs to be homoscedasticity. You can test this assumption in SPSS Statistics by plotting a scatterplot of the standardized residuals against the predicted values.

EFFECT SIZE

Effect size is a term used to describe the strength or magnitude of an effect. This effect is usually expressed as a measure of difference or association. Like most statistical tests, effect sizes come in two distinct groups, and effect sizes generally range from 0 to 1.0. The first type of effect size is based on magnitude of difference between groups, and this is known as the d family of effect sizes. The second type of effect size is the measure of association or the variance accounted for by two or more variables, which is known as the r family of effect sizes. For example, a t-test produces the effect size d, while a correlation coefficient produces the effect size r.

Generally, the effect size values such as d or r, are only transformations of the difference between groups or associations between variables, which are then weighted (divided by) the size of the sample and/or its standard error. This implies that the higher the difference or association, and the greater the sample, the bigger the effect size. Cohen (1988) suggests small, medium, and large effects sizes for a T-test would respectively be about 0.2, 0.5, and 0.8, while small, medium, and large effects sizes for the correlation coefficient would respectively be about 0.1, 0.3, and 0.5 units.

When reporting results from the calculation of a test statistic, it is always a good idea to report more than just the p value. It is far better, and more thorough, to include the effect size and the confidence interval along with the p value result of the test. It is important to note that just because a researcher finds a statistically significant result does not mean the result is sizeable, important, or useful in the real world. Statistical significance is not the only measure of a result. A result

should also be practically significant, which means the strength or size of the effect represents a finding that is practically important to others.

Practical significance always involves judgment by other researchers or consumers of research that takes into account factors such as cost and political considerations of interventions tied to the estimated effects e.g. rolling out the CBC curriculum. In educational researches that compare different instructional strategies, effect size is the magnitude of the difference between the groups. The absolute effect size is the difference between the average, or mean, outcomes in two different intervention groups.

In another example, residents' self-assessed confidence in performing a procedure improved an average of 0.4 point on a Likert-type scale ranging from 1 to 5, after simulation training. While the absolute effect size in the first example appears clear, the effect size in the second example is less apparent. Is a 0.4 change a lot or trivial? Accounting for variability in the measured improvement may aid in interpreting the magnitude of the change in the second example. Thus, effect size can refer to the raw difference between group means, or absolute effect size, as well as standardized measures of effect, which are calculated to transform the effect to an easily understood scale. Absolute effect size is useful when the variables under study have intrinsic meaning (e.g. number of hours of private study).

Calculated indices of effect size are useful when the measurements have no intrinsic meaning, such as numbers on a Likert scale; when studies have used different scales so no direct comparison is possible; or when effect size is examined in the context of variability in the population under study. Calculated effect sizes can also quantitatively compare results from different studies and thus are commonly used in meta-analyses.

Why Report Effect Sizes?

The effect size is the main finding of a quantitative study. While a P value can inform the researcher whether an effect exists and whether it is significant, the P-value will not reveal the size of the effect. In reporting and interpreting studies, both the substantive significance (effect size) and statistical significance (P value) are essential results to be reported since not all significant findings

are important. For this reason, effect sizes should be reported, especially if you have made recommendations that you want the government to adopt for policy formulation.

So why isn't the P-Value Enough?

Statistical significance is the probability that the observed difference between two groups is due to chance. If the P value is larger than 0.05, any observed difference is assumed to be explained by sampling variability. With a sufficiently large sample, a statistical test will almost always demonstrate a significant difference, unless there is no effect whatsoever, that is, when the effect size is exactly zero, yet very small differences, even if significant, are often meaningless. Thus, reporting only the significant P value for an analysis is not adequate for readers and policy makers to fully take the results seriously.

For example, if a sample size is 10 000, a significant P value is likely to be found even when the difference in outcomes between groups is negligible and may not justify an expensive or time-consuming intervention over another. The level of significance by itself does not predict effect size. Unlike significance tests, effect size is independent of sample size. Statistical significance, on the other hand, depends upon both sample size and effect size. For this reason, P values are considered to be confounded because of their dependence on sample size. Sometimes a statistically significant result means only that a huge sample size was used.

How to Calculate Effect Size

Depending upon the type of comparisons under study, effect size is estimated with different indices. The indices fall into two main study categories, those looking at effect sizes between groups and those looking at measures of association between variables. For two independent groups, effect size can be measured by the standardized difference between two means, or mean (group 1) minus mean (group 2) / standard deviation.

The denominator standardizes the difference by transforming the absolute difference into standard deviation units. Cohen's d is an example of this type of effect size index. Cohen classified effect sizes as small ($d = 0.2$), medium ($d = 0.5$), and large ($d \geq 0.8$). According to Cohen, "a medium effect of 0.5 is visible to the naked eye of a careful observer. A small effect of 0.2 is noticeably smaller than medium but not so small as to be trivial.

A large effect of 0.8 is the same distance above the medium as small is below it." These designations large, medium, and small do not take into account other variables such as the accuracy of the assessment instrument and the diversity of the study population. However, these categories provide a general guide that should also be informed by context.

Choosing the Right Quantitative Analysis Method

To choose the right statistical methods, you need to think about two important factors:

- *The type of quantitative data* - Different types of data require different statistical methods of data analysis. This is because each inferential statistical method has assumptions which if violated, might lead to misleading findings. For example, some techniques work with categorical data (for example, yes/no type questions, or gender or ethnicity), while others work with continuous numerical data (for example, age, weight or income) – and, of course, some work with multiple data types. If you try to use a statistical method that doesn't support the data type you have, your results will be largely meaningless. So, make sure that you have a clear understanding of what types of data you've collected (or will collect) before choosing the appropriate inferential statistical method of analysis.
- *The research questions or hypotheses* - The nature of your research questions and research hypotheses will heavily influence which statistical methods and techniques you should use. If you're just interested in understanding the attributes of your sample (as opposed to the entire population), then descriptive statistics are probably all you need.

  For example, if you just want to assess the means (averages) and medians (centre points) of variables in a group of people. On the other hand, if you aim to understand differences between groups or relationships between variables and to infer or predict outcomes in the population, then you'll likely need both descriptive statistics and inferential statistics. So, it's really important to get very clear about your objectives and research questions, as well as your hypotheses before you start looking at which statistical techniques to use.
- *Statistical assumptions* - all inferential statistics rely on assumptions, which if violated might lead to type I or type II error in the course of hypothesis testing. It is therefore important to assess all assumptions of a particular parametric test and proceed with it if none has been violated or there is a way of going about it, otherwise, non-parametric tests should be used.

CHAPTER 8:

PRESENTATION AND DISCUSSION OF EDUCATIONAL RESEARCH FINDINGS

After research data has been analyzed, the next step is to make meaning of it. This has serious ramifications, as not all data analyses make sense to laypersons. This chapter talks about the definition of data presentation, methods of data presentation, importance of data presentation, scholarly writing of educational research findings, structure and style of the results section in a thesis, interpretation and discussion of findings in educational research.

Definition of Data Presentation

Data presentation is expression of the data analysis results in a manner that can easily be understood. It is defined as the process of using various graphical formats to visually represent the relationship between two or more data sets so that an informed decision can be made based on them.

Importance of Data Presentation

Data Presentation can be a deal maker or breaker, based on delivery of the content in the context of visual depiction. Data Presentation tools are therefore powerful communication tools that can simplify the data by making it easily understandable and readable at the same time, while attracting and keeping the interest of its readers to effectively showcase large amounts of complex data in a simplified manner.

If the researcher can create an insightful presentation of the data in hand with the same sets of facts and figures, then the results can be interesting to read. In postgraduate research, there are many cases where the user has had very good data but the manner of presentation turns off the readers of their work. To impress the consumers of your research findings, effective presentation of data is needed. The following are the major reasons for an effective presentation;

- ✧ Most readers are interested in the interpretation of data, not the raw data itself. Therefore, after analysis of the data, users should represent the data with a visual aspect for better understanding and knowledge.

- ✧ The researcher should not overwhelm the audience with too many slides of the presentation and inject an ample amount of texts as pictures that will speak for themselves.
- ✧ Providing a brief description would help the user to attain attention in a small amount of time while informing the audience about the context of the presentation.
- ✧ The inclusion of pictures, charts, graphs and tables in the presentation help for better understanding the outcome.
- ✧ An effective presentation would allow accurate decision making.

Methods of Representing Data

Statistical data can be represented in a number of ways, which must be determined according to the data type, method of analysis used and the information to be emphasized. Inappropriately presented data may fail to clearly convey information to the target audience. Even when the same information is being conveyed, different methods of data presentation must be employed, depending on what specific information is going to be emphasized. A method of presentation must be chosen after carefully weighing the advantages and disadvantages of different methods of presentation. For easy comparison of two randomly chosen methods of presentation, let us look at a table and a graph that present the same information.

If one wishes to compare or introduce two values at a certain time point, it is appropriate to use text or the written language. However, a table is the most appropriate when all information requires equal attention, and it allows readers to selectively look at information of their own interest. Graphs allow readers to understand the overall trend in data, and intuitively understand the comparison results between two groups. One thing to always bear in mind regardless of what method is used, however, is the simplicity and ease of data presentation.

The following are some of the common methods of presenting data in educational research.

i. **Pictorial Presentation -** It is the simplest form of data presentation, often used in schools or universities to provide a clearer picture to students, who are better able to capture the concepts effectively through a pictorial presentation of simple data. An example is shown in figure 17, which shows the number of illiterate children in 5 randomly selected small towns in Kenya.

Small Towns	Number of illiterate children
MATUNDA	
SAGANA	
SABATIA	
KOGALO	
MTWAPA	

Figure 17: Example of a Pictogram

It can be quickly deduced from this pictogram that Sabatia town had most illiterate children, followed by Matunda, while Mtwapa has the fewest number of illiterate children. In pictorial data presentation, one picture can be used to represent a given number of children, say 200. Given this, the number of respondents of each gender in the above example can be determined from the data presented in the pictogram.

ii. **Column chart** - It is a simplified version of the pictorial presentation, which involves management of a larger amount of data being shared during the presentations and providing more clarity to the insights from the data. Figure 18 is an example of a column chart, showing a comparison of wildlife population in Kenya between 2017 and 2022 thus:

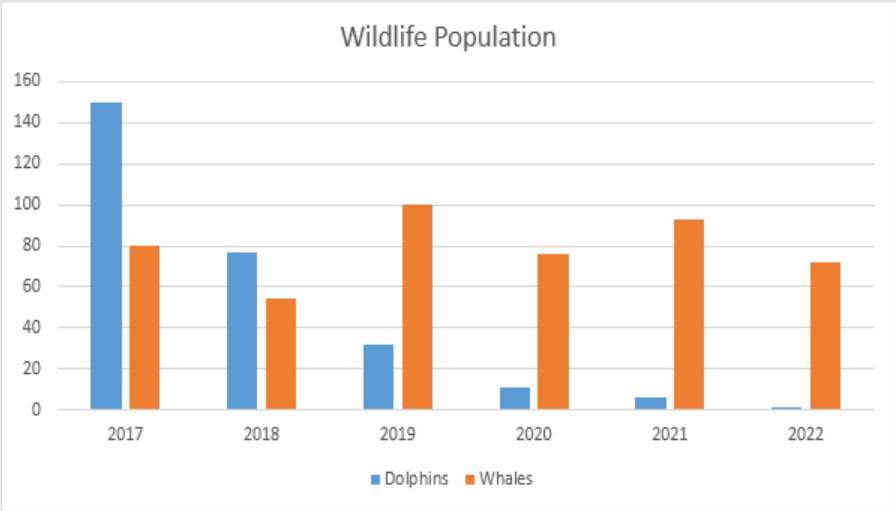


Figure 18: Example of a Column Chart

It can be seen from the column chart that while dolphins have been increasing steadily over the six-year period, that of whales has been lackluster in the same period.

iii. **Pie Charts -** A pie chart, sometimes called a circle chart, is a way of summarizing a set of nominal data or displaying the different values of a given variable (e.g. percentage distribution). This type of chart is a circle divided into a series of segments. Each segment represents a particular category. The area of each segment is the same proportion of a circle as the category is of the total data set. Pie chart usually shows the component parts of a whole. Sometimes you will see a segment of the drawing separated from the rest of the pie in order to emphasize an important piece of information. This is called an exploded pie chart.

The use of the pie charts is quite popular among educational researchers, as the circle provides a visual concept of the whole (100%). Pie charts are also one of the most commonly used charts because they are simple to use. Despite its popularity, pie charts should be used sparingly for two reasons. First, they are best used for displaying statistical information when there are no more than six components only. Otherwise, the resulting picture will be too complex to understand. Secondly, pie charts are not useful when the values of each component are too similar because it is difficult to see the differences between slice sizes.

A pie chart uses percentages to compare information. Percentages are used because they are the easiest way to represent a whole. The whole is equal to 100%. For example, if you spend 7 hours at school and 55 minutes of that time is spent eating lunch, then 13.1% of your school day was spent eating lunch. To present this in a pie chart, you would need to find out how many degrees represent 13.1%. This calculation is done by developing the equation:

Percent ÷ 100 × 360 degrees = the number of degrees

This ratio works because the total percent of the pie chart represents 100% and there are 360 degrees in a circle. Therefore 47.1 degrees of the circle (13.1%) represents the time spent eating lunch. Figure 19 shows an example of a pie chart, indicating the percentages of grade six students who chose each pathway in the CBC curriculum in a certain year.

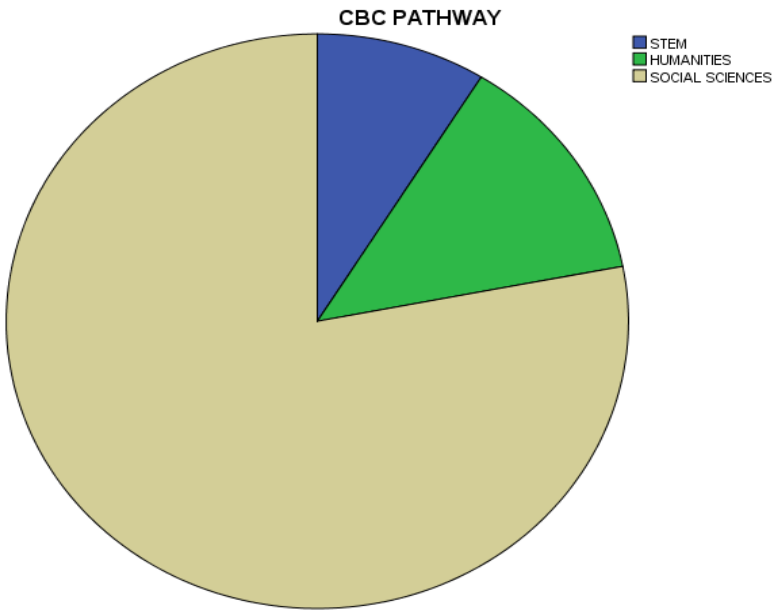


Figure 19: Example of a Pie-Chart

It can be seen that a vast majority of the sample grade six students preferred the social sciences pathway, while the STEM pathway was the least preferred.

iv. **Bar graphs -** A bar graph is a type of graph in which each column (plotted either vertically or horizontally) represents a categorical variable. (A categorical variable is a variable that has two or more categories with no intrinsic ordering to the categories. For example, school type is a categorical variable with three categories; boys', girls' and mixed.

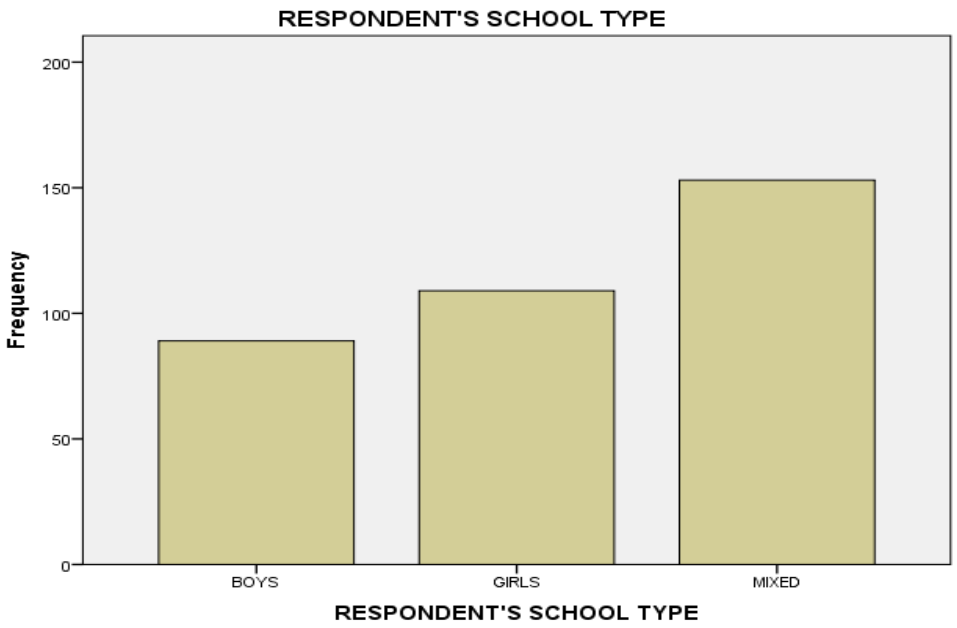


Figure 20: Example of a Bar Graph

A bar graph is used to compare the frequency of a category or characteristic with that of another category or characteristic. The bar height (if vertical) or length (if horizontal) shows the frequency for each category or characteristic. Because each column represents an individual category rather than intervals for a continuous measurement, gaps are included between the bars. Also, the bars can be arranged in any order without affecting the data. Bar charts have a similar appearance as histograms. However, bar charts are used for categorical or qualitative data while histograms are used for quantitative data. Also, in histograms, classes (or bars) touch each other, while in bar charts the bars do not touch each other

v. Histograms

This is a perfect way of showing the spread of numerical data. The main difference between histograms and bar graphs are the gaps between the bars, as histograms have no gaps in between bars.

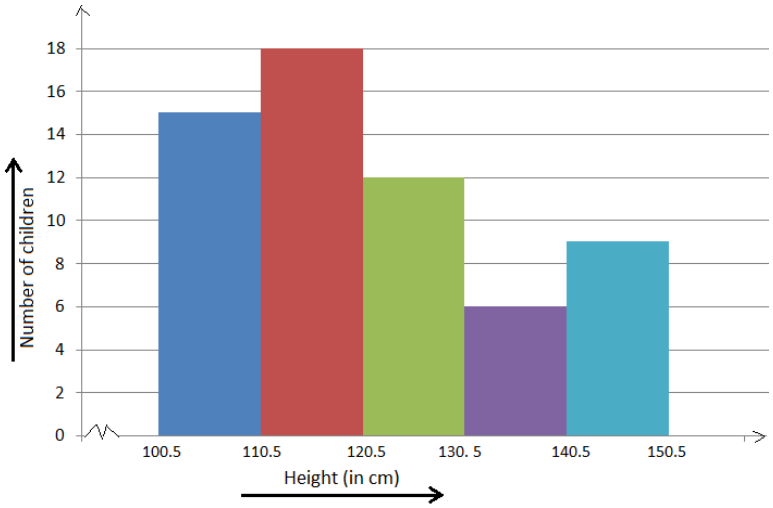


Figure 21: Example of a Histogram

A frequency distribution shows how often each different value in a set of data occurs. A histogram is the most commonly used graph to show frequency distributions. It looks very much like a bar chart, but there are important differences between them. A histogram is used when;

- ✓ The data are numerical
- ✓ You want to see the shape of the data's distribution, especially when determining whether the output of a process is distributed approximately normally
- ✓ Seeing whether a process change has occurred from one time period to another
- ✓ Determining whether the outputs of two or more processes are different
- ✓ You wish to communicate the distribution of data quickly and easily to others

vi. **Maps -** Maps help you with data presentation over an area to display the areas of concern. Map graphs are useful to make an exact depiction of data over a vast case scenario. An example could be a map of Kenya, showing the population of each county.

vii. **Tables -** A tabular presentation of data enables one to view multiple records from a single table. Each record is displayed as a single row in the tabular form and each field is displayed as a column, creating a table within the form itself. Tabular forms are named after the table with which they are associated only in plural. For example, the tabular form associated with the country/region table is called Countries/Regions. In the case of associated tables that have multiple fields in the primary key, the name can be different.

Anyone with basic literacy skills can easily understand the information presented in a table. Tables are the most appropriate for presenting individual information, and can present both quantitative and qualitative information.

Table 14: Example of a Statistical Table in APA Format

Variable	N	Mean	Std. Deviation
Post-Test Achievement Score	351	30.17	9.373
Pre-Test Achievement Score	351	23.41	7.790

The strength of tables is that they can accurately present information that cannot be presented in a graph. A number such as "132.145852" can be accurately expressed in a table. Another strength of tabular presentation is that information with different units can be presented together. For instance, students' gender, age and subject combination can all be presented together in one table, which may not be possible in other ways of data representation. Finally, tables are useful for summarizing and comparing quantitative information of different variables. However, interpretation of information takes longer in tables than in graphs, and therefore tables are not appropriate for studying data trends. Furthermore, since all data are of equal importance in a table, it is not easy to identify and selectively choose the information that requires emphasis.

viii. **Textual presentation -** In qualitative research, text is the main method of presenting data, as it is used to explain results and trends, and provide contextual information. Data are fundamentally presented in paragraphs or sentences. Text can be used to provide interpretation or emphasize certain data. If quantitative information to be conveyed consists of one or two numbers, it is more appropriate to use written language than tables or graphs. For instance, information about the infection rates of Covid-19 in a particular area can be presented with the use of a few numbers, for example:

"The infection rate of coronavirus was 3% in June, 2.5% in July and 1% in August No fatality cases were reported between the three months."

"In Isiolo County, George Wajackoyah obtained 75% of the total votes cast during the 2022 presidential general elections, while his closest rival obtained 10% of the votes cast."

If such information were to be presented in a graph or a table, it would occupy an unnecessarily large space on the page, without enhancing the readers' understanding of the data. If more data were to be presented or other information such as that regarding data trends are to be conveyed, a table or a graph would be more appropriate. By nature, data takes longer to be read when it has been presented as texts and when the main text includes a long list of information, readers and reviewers may have difficulties in understanding the information.

Interpretation of Research Findings

After data has been analyzed and presented, it should be interpreted in line with the research questions or hypotheses. Data interpretation is the process of reviewing data through some predefined processes which will help assign some meaning to the data and arrive at a relevant conclusion. Data interpretation should be the next step of research after data analysis because data, when collected in raw form, may be difficult for the layman to understand, which is why researchers need to break down the information gathered so that others can make sense of it. Interpretation involves taking the result of data analysis, making inferences on the relations observed, and using them as evidence to address the research question or hypothesis.

Interpretation of research data aims at drawing inference from the collected facts and computed statistics or test statistics. The task of interpretation has two major aspects; one is the effort to establish continuity in research through linking the results of given study with those of others, and two is the establishment of some explanatory concepts. In one sense, interpretation is concerned with relationship within the collected data, partially overlapping analysis. Interpretation also extends beyond the data of the study to include the results of other research, theory and hypotheses. Thus, interpretation is the devise through which the factors that seem to explain what has been observed by researchers in the course of the study can be better understood and it also provides a theoretical conception which can serve as a guide for further research.

Process of Data Interpretation

Interpretation of research data is not a very easy job as it requires a great skill and dexterity on the part of researcher. Interpretation is an art that one learns through practice and experience. The researcher may, at times, seek the guidance from experts for accomplishing interpretation.
The technique of interpretation often involves the following steps;

(i) Researchers must give reasonable explanations of the relations which he has found and he must interpret the lines of relationship in terms of the underlying process and must try to find out the thread of uniformity that lies under the surface layer of his diversified research findings. In fact, this is the technique of how generalization should be done and concepts be formulated.

(ii) Extraneous information, if collected during the study, must be considered while interpreting the final results of research study, for it may prove to be a key factor in understanding the problem under consideration.

(iii) It is advisable, before embarking upon final interpretation, to consult someone having insight into the study and who is frank and honest and will not hesitate to point out omissions and errors in logical argumentation. Such a consultation will result in correct interpretation and, thus, will enhance the utility of research results.

(iv) Researchers must accomplish the task of interpretation only after considering all relevant factors affecting the problem to avoid false generalization. He must be in no hurry while interpreting results, for quite often the conclusions, which appear to be all right at the beginning, may not at all accurate.

Importance of Interpretation

The need for interpretation of research data can hardly be over-emphasized. The following are the reasons why data must be interpreted;

(i) It is through interpretation that the researcher can understand the abstract principle that works beneath his findings. Through this he can link up his findings with those of other studies, having the same abstract principle, and thereby can predict about the concrete worlds of events. Fresh inquiries can test these predictions later on. This way of continuity in research can be maintained.

(ii) Interpretation leads to the establishment of explanatory concepts that can serve as a guide for future research studies; it opens new avenues of intellectual adventure and stimulates the quest for more knowledge.

(iii) Researchers can better appreciate only through interpretation why his findings are what they are and can make others to understand the real significance of his research findings.

(iv) The interpretation of the findings of exploratory research study often results into hypotheses for experimental research and as such interpretation is involved in the transition from exploratory to experimental research. Since an exploratory study does not have a hypothesis to start with, the findings of such study have to be interpreted on a post-factum basis in which case the interpretation is technically described as 'post factum' interpretation.

Guidelines for Making Valid Interpretations

The following guidelines are useful in making good interpretations,

- Don't interpret from a single or limited number of instances.
- Don't over-stress both positive and negative points.
- Don't omit evidences contrary to one's opinion.
- Don't overlook important circumstances pertaining to different phenomena studied.
- Don't have any pre-dispositions and pre-conceived notions.
- Don't attribute results to a single factor when other factors are equally important.
- Don't base your judgment on inaccurate instruments of measurement.
- Don't forget to distinguish between material and less-significant issues.
- Don't make any false analogy.
- Don't generalize from insufficient data.
- Try to see the problem in right perspective.
- Make provision for unstudied factors.
- Try to see the differences between cause and effect.
- Recognize the limitations of evidences.
- Base judgments on complete and accurate data.
- Ensure consistency of information and inferences.
- Trust statistical evidences more than verbal evidences.

- Elicit diverse opinions, make appropriate induction-based interpretations
- Make appropriate use of deductive technique of interpretation.

Common Data Interpretation Problems for Novice Researchers

The oft-repeated mantra of those who fear data advancements in the digital age is "big data equals big trouble." While that statement is not accurate, it is safe to say that certain data interpretation problems or "pitfalls" exist and can occur when interpreting data, especially at the speed of thought. Let's identify some of the most common data misinterpretation risks and shed some light on how they can be avoided;

i. **Correlation mistaken for causation -** our first misinterpretation of data refers to the tendency of data analysts to mix the cause of a phenomenon with correlation. It is the assumption that because two actions occurred together, one caused the other. This is not accurate as actions can occur together absent a cause and effect relationship.
    - Digital age example: assuming that increased revenue is the result of increased social media followers… there might be a definitive correlation between the two, especially with today's multi-channel purchasing experiences. But, that does not mean an increase in followers is the direct cause of increased revenue. There could be both a common cause and indirect causality.
    - Remedy: attempt to eliminate the variable you believe to be causing the phenomenon.

ii) Confirmation bias - our second data interpretation problem occurs when you have a theory or hypothesis in mind but are intent on only discovering data patterns that provide support to it while rejecting those that do not.

- Digital age example - your boss asks you to analyze the success of a recent multi-platform social media marketing campaign. While analyzing the potential data variables from the campaign (one that you ran and believe performed well), you see that the share rate for Facebook posts was great, while the share rate for Twitter Tweets was not. Using only the Facebook posts to prove your hypothesis that the campaign was successful would be a perfect manifestation of confirmation bias.
- Remedy - as this pitfall is often based on subjective desires, one remedy would be to analyze data with a team of objective individuals. If this is not possible, another solution is

to resist the urge to make a conclusion before data exploration has been completed. Remember to always try to disprove a hypothesis, not prove it.

iii) Irrelevant data - the third data misinterpretation pitfall is especially important in the digital age. As large data is no longer centrally stored, and as it continues to be analyzed at the speed of thought, it is inevitable that analysts will focus on data that is irrelevant to the problem they are trying to correct. A digital age example: in attempting to gauge the success of an email lead generation campaign, you notice that the number of homepage views directly resulting from the campaign increased, but the number of monthly newsletter subscribers did not. Based on the number of homepage views, you decide the campaign was a success when really it generated zero leads.

iv. **Truncation of axes -** When creating a graph to start interpreting the results of your analysis it is important to keep the axes truthful and avoid generating misleading visualizations. Starting the axes in a value that doesn't portray the actual truth about the data can lead to false conclusions. Digital age example: In the image below we can see a graph from Fox News in which the Y-axes start at 34%, making it seem that the difference between 35% and 39.6% is way higher than it actually is. This could lead to a misinterpretation of the tax rate changes. The remedy is to be careful with the way your data is visualized. Be respectful and realistic with axes to avoid misinterpretation of your data.

v) Small sample size - Another common data analysis and interpretation problem is the use of a small sample size. Logically, the bigger the sample size, the more accurate and reliable are the results. However, this also depends on the size of the effect of the study. For example, the sample size in a survey about the quality of education will not be the same as for one about people doing outdoor sports in a specific area.

- Digital age example: Imagine you ask 30 people a question and 29 answer "yes" resulting in 95% of the total. Now imagine you ask the same question to 1000 and 950 of them answer "yes", which is again 95%. While these percentages might look the same, they certainly do not mean the same thing as a 30 people sample size is not a significant number to establish a truthful conclusion.

- Remedy: Researchers say that in order to determine the correct sample size to get truthful and meaningful results it is necessary to define a margin of error that will represent the maximum amount they want the results to deviate from the statistical mean. Paired to this, they need to define a confidence level that should be between 90 and 99%. With these two values in hand, researchers can calculate an accurate sample size for their studies.

vi) Reliability, subjectivity, and generalizability - When performing qualitative analysis, researchers must consider practical and theoretical limitations when interpreting the data. In some cases, qualitative research can be considered unreliable because of uncontrolled factors that might or might not affect the results. This is paired with the fact that the researcher has a primary role in the interpretation process, meaning he or she decides what is relevant and what is not, and as we know, interpretations can be very subjective.

Generalizability is also an issue that researchers face when dealing with qualitative analysis. As mentioned in the point about small sample size, it is difficult to draw conclusions that are 100% representative because the results might be biased or unrepresentative of a wider population. While these factors are mostly present in qualitative research, they can also affect quantitative analysis. A digital age example: Biased questions in a survey are a great example of reliability and subjectivity issues.

Imagine you are sending a survey to your clients to see how satisfied they are with your customer service with this question: "how amazing was your experience with our customer service team?" Here we can see that this question is clearly influencing the response of the individual by putting the word "amazing" on it. Remedy: A solution to avoid these issues is to keep your research honest and neutral. Keep the wording of the questions as objective as possible. For example: "on a scale of 1-10 how satisfied were you with our customer service team". This is not leading the respondent to any specific answer, meaning the results of your survey will be reliable.

Scholarly Writing of Educational Research Findings

The results section of the research paper or thesis is where you report the findings of your study based upon the information gathered as a result of the methodology [or methodologies] you applied.

The following are useful points when writing the results section;

- The results section should simply state the findings, without bias or interpretation, and arranged in a logical sequence. The results section should always be written in the past tense. A section describing results is particularly necessary if your paper includes data generated from your own research.
- Remember that the results of a study do not prove anything. Research results can only confirm or reject the research problem underpinning your study. However, the act of articulating the results helps you to understand the problem from within, to break it into pieces, and to view the research problem from various perspectives.
- The page length of this section is set by the amount and types of data to be reported. Be concise when using non-textual elements such as figures and tables. In deciding what data to describe in your results section, you must clearly distinguish material that would normally be included in a research paper from any raw data or other material that could be included as an appendix. In general, raw data should not be included in the main text of your paper unless requested to do so by your supervisor.
- Avoid providing data that is not critical to answering the research question. The background information you described in the introduction section should provide the reader with any additional context or explanation needed to understand the results.
- Always re-read the background section of your paper after you have written up your results to ensure that the reader has enough contexts to understand the results [and, later, how you interpreted the results in the discussion section of your paper].

Structure and Writing Style of the Results Section

For most research paper or thesis formats, there are two ways of presenting and organizing the results;

- ✓ Present the results followed by a short explanation of the findings. For example, you may have noticed an unusual correlation between two variables during the analysis of your findings. It is correct to point this out in the results section. However, speculating as to why this correlation exists, and offering a hypothesis about what may be happening, belongs in the discussion section of your paper.

- ✓ Present a section and then discuss it, before presenting the next section then discussing it, and so on. This is more common in longer papers because it helps the reader to better understand each finding. In this model, it can be helpful to provide a brief conclusion in the results section that ties each of the findings together and links to the discussion.

In general, the content of your results section should include the following elements;

- An introductory context for understanding the results by restating the research problem that underpins the purpose of your study. This means you should begin with a paragraph orienting the reader to the chapter. This is important because some readers will jump straight to this location in your thesis, dissertation or paper without reading earlier chapters or sections.
- when using mixed methodology, start with quantitative then qualitative findings
- A clear presentation of your key findings arranged in a logical sequence that generally follows your methodology section. You may use tables or figures as you may deem necessary, with frequent use of precise headings to motivate and help readers follow your presentation with ease.
- Inclusion of non-textual elements, such as, figures, charts, photos, maps, tables, etc. to further illustrate the findings, if appropriate.
- In the text, a systematic description of your results, highlighting for the reader observations that are most relevant to the topic under investigation [remember that not all results that emerge from the methodology that you used to gather the data may be relevant].
- Use of the past tense when referring to your results.
- The page length of your results section is guided by the amount and types of data to be reported. However, focus only on findings that are important and related to addressing the research problem.
- When using non-textual Elements, either place figures, tables, charts, etc. within the text of the result, or include them in the back of the report. Never do both.
- In the text, refer to each non-textual element in numbered order e.g., Table 1, Table 2; Figure 1, Plate 1 etc.
- If you place non-textual elements at the end of the report, make sure they are clearly distinguished from any attached appendix materials, such as raw data.

- Regardless of placement, each non-textual element must be numbered consecutively and complete with caption [caption goes under the figure, table, chart, etc.]
- Each non-textual element must be titled, numbered consecutively, and complete with a clear heading
- In proofreading your results section, be sure that each non-textual element is sufficiently complete so that it could stand on its own, separate from the text.
- Conclude with a recap of the findings using a transitional paragraph such as “chapter four presented quantitative and qualitative findings addressing the four research questions. Chapter five presents conclusion, recommendations and suggestions for further research.

When presenting the research findings, avoid doing the following;

- Discussing or interpreting your results. Save all this for the next section
- Reporting background information
- Ignoring negative results. If some of your results fail to support your hypothesis, do not ignore them. Document them, and then state in your discussion section why you believe a negative result emerged from your study. Note that negative results, and how you handle them, often provides you with the opportunity to write a more engaging discussion section, therefore, don't be afraid to highlight them.
- Including raw data or intermediate calculations. Ask your mentor if you need to include any raw data generated by your study, such as interview transcripts or data files. If raw data is to be included, place it in an appendix or set of appendices that are referred to in the main text of the document.
- Do not use phrases that are vague or non-specific, such as, "appeared to be greater or lesser than..." or "demonstrates promising trends that...."
- Presenting the same data or repeating same information more than once. If you feel the need to highlight something, you will have a chance to do that in the discussion section.
- Confusing figures with tables. Be sure to properly label any non-textual elements in your paper. If you are not sure, look up the term in a dictionary.

Discussion of Findings

The purpose of the discussion is to describe the significance of your findings in light of what was already known about the research problem being investigated, and to explain any new understanding or fresh insights about the problem after you've taken the findings into consideration. The discussion will always connect to the introduction by way of the research questions or hypotheses you posed and the literature you reviewed, but it does not simply repeat or rearrange the introduction; the discussion should always explain how your study has moved the reader's understanding of the research problem forward from where you left them at the end of the introduction.

Importance of a Good Discussion

This section is often considered the most important part of a research paper because it most effectively demonstrates your ability as a researcher to think critically about an issue, to develop creative solutions to problems based on the findings, and to formulate a deeper, more profound understanding of the research problem you are studying.

The discussion section is where you explore the underlying meaning of your research, its possible implications in other areas of study, and the possible improvements that can be made in order to further develop the concerns of your research. This is the section where you need to present the importance of your study and how it may be able to add value and contribute to and/or fill existing knowledge gaps in the research area.

If appropriate, the discussion section is also where you state how the findings from your study revealed new gaps in the literature that had not been previously exposed or adequately described. This part of the paper or thesis is not strictly governed by objective reporting of information but, rather, it is where you can engage in creative thinking about issues through evidence-based interpretation of findings. This is where you infuse your results with meaning. The following are general rules of the thumb, that you should adopt when composing your discussion of the results;

- ✧ Do not be verbose or repetitive.
- ✧ Be concise and make your points clearly.
- ✧ Avoid using jargon.

- ✧ Follow a logical stream of thought.
- ✧ Use the present verb tense, especially for established facts; however, refer to specific works and references in the past tense.
- ✧ If needed, use subheadings to help organize your presentation or to group your interpretations into themes.

Content of the Discussion Section

The content of the discussion section of your paper most often includes:

- ➢ Explanation - comment on whether or not the results were expected and present explanations for the results; go into greater depth when explaining findings that were unexpected or especially profound. If appropriate, note any unusual or unanticipated patterns or trends that emerged from your results and explain their meaning.

- ➢ Comparison - compare your results with the findings from other studies, or use the studies to support a claim. This can include re-visiting key sources already cited in your literature review section, or, save them to cite later in the discussion section if they are more important to compare with your results than being part of the general research you cited to provide context and background information.

- ➢ Deduction - a claim for how the results can be applied more generally. For example, describing lessons learned, proposing recommendations that can help improve a situation, or recommending best practices.

- ➢ Hypothesize - Make an informed claim or possible conclusion arising from the results, which may be proved or disproved in subsequent research.

Organization and Structure of the Discussion Section in a Thesis

Keep the following sequential points in mind as you organize and write the discussion section of your research report;

(i) Think of your discussion as an inverted pyramid. Organize the discussion from the general to the specific, linking your findings to the literature, then to theory, then to practice if appropriate.

(ii) Use the same key terms, mode of narration, and verb tense [present] that you used when describing the research problem in the introduction.

(iii) Begin by briefly re-stating the research problem you were investigating and answer all of the research questions underpinning the problem that you posed in the introduction. You may start with a paragraph that orients the reader to the chapter, as some will jump straight to this section without reading the previous chapters or sections.

For example;

> *"Chapter five presents conclusions based on research findings from data collected on ….. As well as discussion and recommendations for policy action, followed by suggestions for further research."*

(iv) Describe the patterns, principles, and relationships shown by each major findings and place them in proper perspective. The sequencing of providing this information is important; first, state the answer, then the relevant results, then cite the work of others. If appropriate, refer the reader to a figure or table to help enhance the interpretation of the data. The order of interpreting each major finding should be in the same order as they were described in your results section.

(v) A good discussion section includes analysis of any unexpected findings. This paragraph should begin with a description of the unexpected finding, followed by a brief interpretation as to why you believe it appeared and, if necessary, it's possible significance in relation to the overall study. If more than one unexpected finding emerged during the study, describe each of them in the order they appeared as you gathered the data.

(vi) Before winding up the discussion, identify potential limitations and weaknesses. Comment on their relative importance in relation to your overall interpretation of the results and, if necessary, note how they may affect the validity of the findings. Avoid using an apologetic tone; however, be honest and self-critical.

(vii) Use of the first person is acceptable, but too much use of the first person may actually distract the reader from the main points.

(viii) In the summary section, include purpose of the study, research objectives, summary of literature reviewed, summary of the methodology and summary of the findings.

(ix) In the discussion section, highlight each conclusion as per the objectives, together with major findings. Present each briefly but do not stop there. Show how your conclusions compare or contrast with those of other researchers, explaining how each is important. This is where you come clear on what you have learnt from your research, highlighting any new material that could probably lead to a new theory/model, modify or nullify existing ones.

CHAPTER 9:

RECOMMENDATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

Your dissertation needs to include recommendations and suggestions for research that can be done by others at a later date, picking from where you left. Depending on requirements of your university, suggestions for future research can be either integrated into research limitations section or it can be a separate section. This chapter discusses the rationale for suggestions for further research, recommendations and how to write them.

Research Recommendations

Recommendations are arguably the most important part of the post data analysis phase of the research process. This is where you'll suggest specific interventions or strategies to address the issues and constraints identified in the research findings. Recommendations should directly respond to key findings arrived at through data collection, analysis and interpretation. Prioritization is essential to narrowing down findings and once this is done, recommendations should be developed to align with the most important findings as per by the research objectives. Recommendations should be brief, succinct and starting with an action verb e.g. create, introduce, fund, facilitate, coordinate etc. They should use a "SMART" format (Specific, Measurable, Attainable, Realistic and Timely). Each recommendation should be followed by a few sentences of explanatory text.

In addition to being "SMART," recommendations should be feasible and practical. Both operational feasibility and political feasibility i.e. political goodwill to implement should be considered when coming up with recommendations. Gauging stakeholder receptivity to proposed recommendations is essential at this juncture.

A final consideration for recommendations is timing. Often solutions to identified problems are sequential, building upon intermediate steps. A researcher may sometimes frame recommendations as short-term (e.g. within 6 to 12 months) and long term (e.g., 1 to 3 years). Recommendations are typically summarized and presented in full within the latter sections of a dissertation, usually as a short list of bullets following the findings summary. In some dissertations, there may be a section at the end of the document that consolidates recommendations and

highlights the linkages across core areas (e.g. service delivery and educational management, curriculum review etc.) prior to the conclusion section.

Since recommendations stem from research findings, the researcher should link each of the recommendations to the finding that supports it, in order to highlight the direct connection between assessment and action. You can show this link visually by using a two-columned table: the first column lists the finding, and the respective recommendation is listed adjacently in the second column.

In the recommendations section, provide a list of things that should be done by major stakeholders e.g., government, teachers, universities, parents etc. as follow up from your research findings, with the goal of solving or addressing the problem that necessitated your research. These should however be based on verifiable and irrefutable research evidence.

The following are useful questions that the researchers should ask themselves when formulating any recommendations;

- Is it feasible?
- Is there political will, or potential opposition?
- Does it fill an existing gap or need?
- What is the likely impact?
- Are resources available (individuals, expertise, funding)?
- Does it require policy reform?
- What is the estimated time frame to implement?
- Does it solve an existing societal issue and, if so, how?
- Will it remove a critical barrier or accelerate an opportunity to achieve an outcome of interest?
- Might it generate an unintended consequence?
- Is it informed by *effect size*?

How to Write Recommendations in a Research Report

Many students and novice researchers put in a lot of effort and write a good research report but fall short on giving proper recommendations. Recommendations in a research paper or thesis form a crucial part of any research. As a researcher, you need to demonstrate a deep understanding of the topic of research and your most important findings if you are to give good recommendations. The following are tips that may help you reach this milestone;

- Recommendations in the research paper should be the objective of the research. Therefore at least one of your latent objectives of the paper is to provide recommendations to stakeholders of the parties that will benefit from your research. For example, "to encourage higher school completion rate, the Ministry of Education should provide free meals to students in marginalized areas. Additionally, the parents should apply for bursaries to supplement current government subsidies."
- Recommendations in the research paper should come from your review and analysis. For example "It was observed that all head teachers that were interviewed had been stationed at their respective schools for the past 2-3 years only. This shows that the attrition rate of head teachers in the region is high and therefore, the Teachers Service Commission should work on reducing the turnover of head teachers".
- Recommendations in the research paper should also come from the data you have analyzed. For example, "the research found that students over 20 years of age are at greater risk of dropping out of school. Therefore, it is recommended that policies that are made for combating class repetition cases should target this specific group".
- Recommendations in the research paper should also come from observation. For example, "it is observed that staff turnover at the school is stable but academic performance has displayed a downward trend. Therefore, the school should analyze its academic practices and programmes."
- Recommendations in the research paper should be written in the order of priority. The most important recommendations for decision-makers should come first. However, if the recommendations are of equal importance then it should come in the sequence in which the topic is approached in the research.

- Recommendations in a research paper if associated with different categories then you should categorize them. For example, you have separate recommendations for policymakers, educators, and administrators then you can categorize the recommendations.
- Recommendations in the research paper should come purely from your research. For example, you have written research on the impact on teaching strategies on motivation. However, nowhere you have discussed online teaching. Then you should not give recommendations for using online teaching to boost students' motivation.
- The use of bullet points offers better clarity rather than using long paragraphs. For example, this paragraph "It is recommended that Britannia Biscuit should launch and promote sugar-free options apart from their existing product range. Promotion efforts should be directed at creating a fresh and healthy image.
- A campaign that conveys a sense of health and vitality to the consumer while enjoying biscuit is recommended" can be written as:
    - The company should launch and promote sugar-free options
    - The company should work towards creating fresh and healthy image
    - The company should run a campaign to convey its healthy image

As earlier mentioned, the inclusion of an action plan along with recommendation adds more weightage to your recommendation. Recommendations should be clear and conscience and written using actionable words. Recommendations should display a solution-oriented approach and in some cases should highlight the scope for further research.

Suggestions for Further Research

Suggestions for further research section should propose 3 to 6 suggestions for further research, depending on the number of your research objectives, which should be done by other researchers in the near future, as continuation of your research. With valid reasons, you will need to propose 4 to 5 suggestions for future studies depending on the number of your research objectives and these can include but not limited to the following:

- **Building upon findings of your research** - Often, the findings from your dissertation research will highlight a number of new avenues that could be explored in future studies.

These could be findings that you did not anticipate since your dissertation will inevitably lead to findings that you did not anticipate from the start. These are useful when making future research suggestions because they can lead to entirely new avenues to explore in future studies. If this was the case, it is worth (a) briefly describing what these unanticipated findings were and (b) suggesting a research strategy that could be used to explore such findings in future.

The other aspect is factors that address unanswered aspects of your research questions. This is because sometimes, dissertations manage to address all aspects of the research questions that were set. However, this is seldom the case. Typically, there will be aspects of your research questions that could not be answered. This is not necessarily a flaw in your research strategy, but may simply reflect that fact that the findings did not provide all the answers you hoped for. If this was the case, it is worth (a) briefly describing what aspects of your research questions were not answered and (b) suggesting a research strategy that could be used to explore such aspects in future.

- **Addressing limitations of your research** - Your research will not be free from limitations and these may relate to formulation of research aim and objectives, application of data collection method, sample size, scope of discussions and analysis etc. You can propose future research suggestions that address the limitations of your study. In the Research Limitations section of your Conclusions chapter, you will have inevitably detailed the potential flaws (i.e., research limitations) of your dissertation. These may include; an inability to answer your research questions, Theoretical and conceptual problems, Limitations of your research strategy, Problems of research quality etc. Identifying what these research limitations were and proposing future research suggestions that address them is arguably the easiest and quickest ways to complete the Future Research section of your Conclusions chapter.
- **Replicating the same research in a new location** - It is most likely that you have addressed your research problem within the settings of specific context, location and/or culture. Accordingly, you can propose future studies that can address the same research problem in a different setting, context, location and/or culture. More often than not, you will not only have set out a conceptual framework (or theoretical model), as described in

the previous section, but you will also have examined (or tested) it in the field. When you do this, focus is typically placed on a specific context, location and/or culture.

If this is the case, the obvious future research suggestion that you could propose would be to examine your conceptual framework (or test the theoretical model) in a new context, location and/or culture. For example, perhaps you focused on school head teachers (rather than ordinary classroom teachers), or Kirinyaga county (rather than the whole Mt. Kenya region), or a more individualistic culture like the Maasai community (rather than a more collectivist culture like pastoral communities in Kenya). When you propose a new context, location and/or culture as your future research suggestion, make sure you justify the choice that you make.

For example, there may be little value in future studies looking at different cultures if culture is not an important component underlying your conceptual framework (or theoretical model). If you are not sure whether a new context, location or culture is more appropriate, or what new context, location or culture you should select, a review the literature will often help clarify where you focus should be.

- **Expanding a conceptual or theoretical framework -** Assuming that you have set out a conceptual framework (or theoretical model) and examined (or tested) it in the field, another series of future research suggestions comes out of expanding that conceptual framework (or theoretical model). We talk about a series of future research suggestions because there are so many ways that you can expand on your conceptual framework (or theoretical model). For example, you can do this by;
    (i) Examining constructs (or variables) that were included in your conceptual framework (or theoretical model) but were not focused.
    (ii) Looking at a particular relationship aspect of your conceptual framework (or theoretical model) further.
    (iii) Adding new constructs (or variables) to the conceptual framework (or theoretical model) you set out (if justified by the literature).

    It would be possible to include one or a number of these as future research suggestions. Again, make sure that any suggestions you make have are justified, either by your findings or the literature.

- **Re-evaluating a conceptual or theoretical framework -** With the dissertation process at the undergraduate and master's level lasting between 3 and 9 months, a lot a can happen in between. For example, a specific event (e.g., 9/11, the economic crisis) or some new theory or evidence that undermines (or questions) the literature (theory) and assumptions underpinning your conceptual framework (or theoretical model). Clearly, there is little you can do about this. However, if this happens, reflecting on it and re-evaluating your conceptual framework (or theoretical model), as well as your findings, is an obvious source of future research suggestions.

BIBLIOGRAPHY

Anderson, G & Arsenault, N. (1998). *Fundamental of Educational Research*, London: Falmer Press

Ary, D, Jacobs, L.C, Sorenson, C, & Razavieh, A. (2010). *Introduction to Research in Education* (8th Ed.) Wordsworth: Cenage Learning

Bailey, R.A. (2008). Design of Comparative Experiments. *Journal of the Royal Statistical Society: Series A General, 111 (3), 181-211*

Cohen J. (1988). *Statistical Power Analysis.* (2nd edition). Hillsdale NJ: Erlbaum

Cohen J. (1990). Things I have learned (so far). *American Journal of Psychology. 45(1), 1304-1312*

Cortina, J.M., (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology, 78(1), 98-104.*

Creswell, J.W. & Plano, C. (2011). *Designing and Conducting Mixed Methods Research*. Los Angeles, CA: Sage

Delice, A. (2010). The Sampling Issues in Quantitative Research. E*ducation Sciences: Theory and Practice, 10 (4), 2001-2018*

DeRue, S. (2012). "A Quasi-Experimental Study of After-Event Reviews". *Journal of Applied Psychology, 97 (5), 681-689*

Frank, S. F. (1962). *Theory and Practice of Psychological Testing*. New Delhi:Mohan Primiani.

Gene V. G. & Julian C. S. (1970). *Statistical Methods in Education and Psychology*. New Jersey: Prentice-Hall.

George, D. & Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 Update*. (4th Edition). Boston: Allyn & Bacon.

Guest, G. (2013). Describing Mixed Methods Research: An Alternative to Topologies. *Journal of Mixed Methods Research, 7 (2), 141-151*

Herert, J. K. & Richard E. (1971). *Learning and Human Abilities: Educational Psychology*. New York: Harper and Row Publishers.

Kahn, J.V. & Best, J.W. (1998). *Research in Education.* (8th Edition). Boston: Allyn and Bacon Publishers.

Kuder, G.F., & Richardson, M.W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2(3), 151–160.*

Pearl, J. (2015). Generalizing Experimental Findings. *Journal of Causal Inference, 3 (2), 259-266*

Teddlie, C. and Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. London: Sage.

Wilcox, R.R. (2010). *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*. (Second Edition). London: Springer

INTERNET SOURCE LINKS

https://usq.pressbooks.pub/statisticsforresearchstudents/chapter/correlation-assumptions/
http://www.nus.edu.sg/celc/research/books/cwtuc/chapter04.pdf
https://academicguides.waldenu.edu/writingcenter/writingprocess/conclusions
https://courses.lumenlearning.com/olemiss-writ250/chapter/results-discussion-and-conclusion/
https://dissertation.laerd.com/principles-of-research-ethics.php
https://files.eric.ed.gov/fulltext/EJ1196755.pdf
https://libguides.usc.edu/writingguide/discussion
https://library.sacredheart.edu/c.php?g=29803&p=185902
https://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n571.xml
https://online.maryville.edu/blog/data-analysis-techniques/
https://pressbooks.bccampus.ca/jibcresearchmethods/chapter/12-4-getting-in-and-choosing-a-site/
https://research-methodology.net/research-methodology/reliability-validity-and-repeatability/research-validity/
https://research-methodology.net/research-methodology/selecting-research-area/
https://scientific-publishing.webshop.elsevier.com/manuscript-preparation/what-background-study-and-how-should-it-be-written/amp/
https://statistics.laerd.com/spss-tutorials/ancova-using-spss-statistics-2.php
https://towardsdatascience.com/sampling-techniques-a4e34111d808
https://uca.edu/psychology/files/2013/08/Ch7-Sampling-Techniques.pdf
https://www.alchemer.com/resources/blog/secondary-data-analysis/
https://www.djsresearch.co.uk/glossary/item/Target-Population
https://www.editage.com/insights/the-basics-of-writing-a-statement-of-the-problem-for-your-researchproposal?utm_source=TrendMD&utm_medium=cpc&utm_campaign=Editage_Insights_TrendMD_0
https://www.enago.com/academy/pilot-study-defines-a-good-research-design/
https://www.intellspot.com/primary-data-vs-secondary-data/
https://www.scribbr.com/methodology/reliability-vs validity/#:~:text=Validity%20refers%20to%20how%20accurately,the%20physical%20or%20social%20world.
https://www.simplilearn.com/data-analysis-methods-process-types-article
https://www.simplypsychology.org/pilot-studies.html
https://www.tutorialspoint.com/excel_data_analysis/data_analysis_process.htm

Swaen, B. & George, T. (2022, August 02). What Is a Conceptual Framework? | Tips & Examples.Scribbr. Retrieved November 13, 2022, from https://www.scribbr.com/methodology/conceptual-framework/

Vinz, S. (2022, October 10). What Is a Theoretical Framework? | Guide to Organizing. Scribbr. Retrieved November 12, 2022, from https://www.scribbr.com/research-paper/theoretical-framework

APPENDICES

Appendix 1: The T-Test Distribution Table

cum. Prob	*t*.50	*t*.75	*t*.80	*t*.85	*t*.90	*t*.95	*t*.975	*t*.99	*t*.995	*t*.999	*t*.9995
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
DF											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
Z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Source: statistical Inference

Appendix 2: The Chi-Square Distribution Table

D.F	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89
32	15.13	16.36	18.29	20.07	22.27	42.58	46.19	49.48	53.49
34	16.50	17.79	19.81	21.66	23.95	44.90	48.60	51.97	56.06
38	19.29	20.69	22.88	24.88	27.34	49.51	53.38	56.90	61.16
42	22.14	23.65	26.00	28.14	30.77	54.09	58.12	61.78	66.21
46	25.04	26.66	29.16	31.44	34.22	58.64	62.83	66.62	71.20
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15
55	31.73	33.57	36.40	38.96	42.06	68.80	73.31	77.38	82.29
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38
65	39.38	41.44	44.60	47.45	50.88	79.97	84.82	89.18	94.42
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43
75	47.21	49.48	52.94	56.05	59.79	91.06	96.22	100.84	106.39
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33
85	55.17	57.63	61.39	64.75	68.78	102.08	107.52	112.39	118.24
90	59.20	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12
95	63.25	65.90	69.92	73.52	77.82	113.04	118.75	123.86	129.97
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81

Source: statistical Inference

Appendix 3: The F-Test Table at 0.05 Alpha

Denominator DF	Numerator DF 1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
31	4.160	3.305	2.911	2.679	2.523	2.409	2.323	2.255	2.199	2.153
32	4.149	3.295	2.901	2.668	2.512	2.399	2.313	2.244	2.189	2.142
33	4.139	3.285	2.892	2.659	2.503	2.389	2.303	2.235	2.179	2.133
34	4.130	3.276	2.883	2.650	2.494	2.380	2.294	2.225	2.170	2.123
35	4.121	3.267	2.874	2.641	2.485	2.372	2.285	2.217	2.161	2.114
36	4.113	3.259	2.866	2.634	2.477	2.364	2.277	2.209	2.153	2.106
37	4.105	3.252	2.859	2.626	2.470	2.356	2.270	2.201	2.145	2.098
38	4.098	3.245	2.852	2.619	2.463	2.349	2.262	2.194	2.138	2.091
39	4.091	3.238	2.845	2.612	2.456	2.342	2.255	2.187	2.131	2.084
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077

Source: statistical Inference

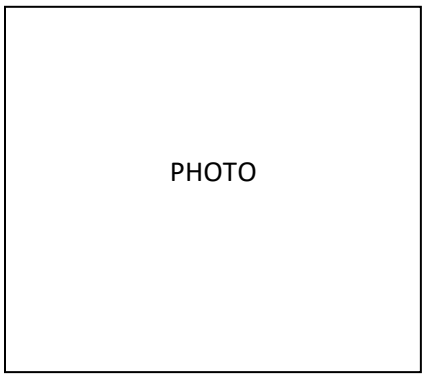
Appendix 4: The Krejcie and Morgan Table

N	S	N	S	N	S
10	10	220	140	1200	291
15	14	230	144	1300	297
20	19	240	148	1400	302
25	24	250	152	1500	306
30	28	260	155	1600	310
35	32	270	159	1700	313
40	36	280	162	1800	317
45	40	290	165	1900	320
50	44	300	169	2000	322
55	48	320	175	2200	327
60	52	340	181	2400	331
65	56	360	186	2600	335
70	59	380	191	2800	338
75	63	400	196	3000	341
80	66	420	201	3500	346
85	70	440	205	4000	351
90	73	460	210	4500	354
95	76	480	214	5000	357
100	80	500	217	6000	361
110	86	550	226	7000	364
120	92	600	234	8000	367
130	97	650	242	9000	368
140	103	700	248	10000	370
150	108	750	254	15000	375
160	113	800	260	20000	377
170	118	850	265	30000	379
180	123	900	269	40000	380
190	127	950	274	50000	381
200	132	1000	278	75000	382
210	136	1100	285	1000000	384

N = Population size, S = Sample size

Source: statistical Inference

ABOUT THE AUTHOR



The author (pictured) is a seasoned educational consultant, researcher and currently a lecturer at Murang'a University of Technology. He has previously taught in several universities, colleges and high schools across the country and has a Doctor of Philosophy in Science Education (Chemistry option), from which his teaching philosophy of "modernizing chemistry teaching and learning in Kenya" was born. Moreover, he has supervised and mentored postgraduate students from different universities across the republic of Kenya. Additionally, the author is a renowned playwright and drama producer, who has written and directed award winning plays at the Kenya National Secondary Schools and Colleges Drama Festival, like the famous "Persona Non-Grata", "Pitfalls of Vanity", and the "*Omwombashi*" narrative, all being Chavakali High School productions.

Motivational quote for upcoming researchers:

“Our deepest fear is not that we are inadequate. Our deepest fear is that we are powerful beyond measure. We ask ourselves, “Who am I to be brilliant, gorgeous, talented, fabulous?” Actually, who are you not to be? We were born to make manifest of the Glory of God that is within us. And as we let our own light shine, we unconsciously give other people permission to do the same”

…… MARIANNE WILLIAMSON