Analysing Factors in ICT Integration among Staff in Tertiary Academic Institutions

A Case Study of Machakos Town

Moses M. Owana

University of Nairobi

© 2015

Analysing Factors in ICT Integration among Staff in Tertiary Academic Institutions

A Case Study of Machakos Town

Moses M. Owana

I56/68809/2011

University of Nairobi

School of Mathematics

This Research Thesis Submitted to the School of Mathematics in Partial Fulfilment of the Requirement for the Award of the degree of Master of Science in Statistics of the University of Nairobi

Copyright © 2015

All rights reserved.

No part of this publication may be copied, photocopied, scanned, reproduced, translated, transmitted, recorded or stored on any electronic storage medium or retrieval system in any form or by any means now known or hereinafter invented, in whole or in part, without a written permission from the author or publisher.

TABLE OF CONTENTS

Declaration and Approval

I hereby declare that this research thesis is my original work and has not been presented in any university or any other institution of learning in the world for the award of any degree or diploma.

Presenter: Moses M. Owana

Signature: ……………………………..……………..

Date: ……………………………..……………..

This research thesis has been submitted to the School of Mathematics at the College of Biological and Physical Sciences of the University of Nairobi by the approval of the university research supervisors.

Supervisor: Dr Nelson Owuor

Signature: ……………………..……..……………..

Date: ……………………..……..……………..

Supervisor: Dr Vincent Oeba

Signature: ……………………..……..……………..

Date: ……………………..……..……………..

Dedication

My dedication for this research goes to my father Mzee Nelson Owana, who from my childhood age, has been a great source of inspiration to me to pursue higher levels in academics, and even in his sunset years as an octogenarian, has for the last decade, been encouraging me to proceed with education and pursue a masters degree course.

This dedication also extends to my son James as a source of inspiration to him to continue with the spirit of excellence in education which my father championed.

Acknowledgement

My profound acknowledgement extends to all those who assisted me in various ways in the course of this research in particular and in my masters degree studies in general that ultimately enabled me to complete the long journey and compile this final report.

My profound gratitude goes to all the statistics lecturers at the School of Mathematics of the University of Nairobi in general and in particular to my two supervisors Dr Nelson Owuor and Dr Vincent Oeba for their dedicated support and guidance that enabled me to accomplish this enormous task.

I also wish to acknowledge the contribution of all those who assisted me in the collection of data by administering questionnaires to the respondents. In this regard, special thanks go to Mr Lore Okemba of Machakos Teachers Training College who produced copies of the questionnaires and also administered them on my behalf in his institution. I also wish to recognise the contribution of Ms Linnet Maundu of Wote Technical Training Institute for administering for me the questionnaires in her institution.

My gratitude also extends to all the respondents who sacrificed their valuable time to fill the questionnaires from which I obtained the valuable data for analysis.

Special thanks also extend to my wife Caroline, who made unquantifiable sacrifice to ensure that I had a conducive environment as I went through this research.

Executive Summary

Organisations and the government have been trying to promote ICT integration among staff to enable them use ICT tools at the place of work. Enormous amount of resources have been spent to improve productivity by promoting ICT usage. However, general observation shows that with all this effort, ICT integration is still at a low level although the number of computer users appears to be on the increase. The purpose of this research therefore was to investigate the factors influencing ICT integration among staff in tertiary academic institutions. Specifically, the research sought to identify; clusters of variables which measure similar underlying dimensions in ICT integration, the individual factors in ICT integration, and the institutional factors in ICT integration.

The research was a case study of institutions in Machakos Town that used survey design procedure where multiple sampling techniques were used to cater for the heterogeneity of the target population. Data was collected through questionnaires administered directly to the staff.

The research used two statistical techniques in the analysis of data: principal component analysis to reduce the number of explanatory variables, and the ordinal logistic regression to extend further to estimate the parameters and to model the response variable to enable statistical inference be drawn on the data variables.

The following variables were found to be statistically significant in ICT integration: age, mode of ICT training, e-mail use, possession of a smart phone, access to a projector, and availability of all the required software. However, the following variables were found not to be statistically significant in ICT integration: gender, staff category, job satisfaction, access to a desktop PC, and availability of only some of the required software.

Stakeholders in ICT should allocate resources towards provision of facilities that enhance ICT integration such as application software required by staff, projectors, Internet access and local area networks, and promotion of formal training in ICT as formal training has a significant impact on ICT integration. Also more resources should be directed to training of the young staff. Policies and programmes should not focus on the gender as part of affirmative action to promote women, as gender was not found to be statistically significant. ICT skills should not be made a mandatory requirement by employers at the time of recruitment.

Abbreviations and Acronyms

ICT	Information and Communication Technology
CD	Compact Disk
DBMS	Database Management System
DVD	Digital Versatile Disk
HDD	Hard Disk Drive
GPS	Global Positioning System
GLM	Generalised Linear Model
IMF	International Monetary Fund
ISP	Internet Service Provider
MIS	Management Information System
NEPAD	New Partnership for Africa's Development
PCA	Principal Component Analysis
UN	United Nations
UNCTAD	United Nations Conference on Trade and Development

CHAPTER 1: INTRODUCTION

1.1 Background to the Study

Throughout the history of mankind, data in different forms has been processed, basically manually. But since the invention of the computer in the 1940s, man has consistently shifted from the manual data processing to electronic data processing involving the use of the computer and other digital devices. This has seen the computer become an integral and indispensable tool to the management of almost all organisations in the contemporary world. Without the use of the computer and its associated benefits, the so called blue chip companies would not be what they are today. Lucey (1991) notes that ICT is has a profound influence on all aspects of life, including organisations.

The use and teaching of ICT in Kenyan schools is still minimal, although the government has made a move through the NEPAD schools initiative in which eight public schools – seven secondary and one primary, were assisted to set up computer laboratories, and later in 2007, a programme was rolled out by the Ministry of Education in collaboration with the Directorate of e-Government, in which three schools were selected from each district, and one computer laboratory set up fully funded by the government. The challenge is that this selection was based on schools that already had the infrastructure in place – connection to the national electric grid and a building. In the field of education, Gathano (2009) however, notes that ICT is still viewed as a study area rather than as a tool in teaching. This attitude also extends to the workplace across almost all careers.

It is widely believed that poverty and phone or Internet connectivity are some of the greatest impediments to computer usage and ICT integration especially in Africa. Orodho (2004) states that half of this population live below US $1 a day which is the set international poverty line. He further notes that on average there are only 18 mainline telephones per 1000 people in Africa as compared to the developed world of 567 lines per 1000 people. However, the current proliferation of mobile phones in Kenya appears to have overtaken the hitherto shortage of landline phones.

The UN (2013) notes that since about 70% of the Kenyan population owns mobile phones, leading mobile network operators like Safaricom have taken a great step in offering services that meets the needs of citizens. Such services include M-Pesa and Kipokezi. This has even appealed to the majority of Kenyans, as it supports the branchless members of the society too, in doing normal and secure businesses via M-Pesa. The recent IMF report reveals that M-Pesa transactions in Kenya exceeded those carried out by Western Union worldwide.

The Government of Kenya through the Directorate of e-Government (2011) recognizes the role played by ICT and the Internet as a means of enhancing communication and exchange of information, delivering services and interacting better with citizens and consumers. It further recognizes the need for a consistent approach to the use of ICT in general and the Internet in particular, culminating in the formulation of its ICT Standards Document to ensure the citizens and all its customers are provided with high quality, consistent, accessible and useable on-line and off-line services. Through this, the Government of Kenya aims at minimising the effects of the "digital divide" by ensuring that Kenyans have equal access to all services and information.

According to a bi-annual e-Government survey conducted by the United Nations Public Administration Network, UN (2013) on e-Government Readiness in which it ranks countries of the world according to the state of e-government readiness and the extent of e-participation, and then constructing a model for the measurement of digitized services based on website assessment, telecommunication infrastructure and human resource endowment, Kenya does not appear among the top 50 countries in the world.

In this dynamic world, everything in life; school, college, home, workplace, entertainment, business and general communication, is going digital. This means that the tool used is either a computer or a computer device. In the academic realm, e-learning has created a mode of learning which is convenient as it takes place anytime, anywhere and at the pace of the learner, coupled with the e-library popularly known as the library without walls.

World Bank (2011) defines "e-Government" as the use by government agencies of information technologies (such as wide area networks, the Internet, and mobile computing) that have the ability to transform relations with citizens, businesses, and other arms of government. These technologies result in better delivery of government services to citizens, improved interactions with business and industry, citizen empowerment through access to information, and more efficient government management. The resulting benefits are reduced corruption, increased transparency, greater convenience, revenue growth, and cost reduction (Kamuti, 2009).

The Directorate of e-Government (2011) asserts that government websites should provide convenient channels that assist consumers to identify what services they require and proceed to obtain the services on-line. This should bring to an end the need to physically travel to government offices to transact business. In recognising the role of ICT, the Government of Kenya outlines the ICT standards in the e-government standards designed to give guidelines to improve public service delivery through the use of ICTs and the Internet. It further recognizes the opportunity presented by the social media such as Facebook and Twitter.

The Government of Kenya therefore recognises the need to place adequate and strategic attention to these new opportunities provided by ICT by ensuring that they are not limited and accessible to only the large corporations within national economies, but also to local public sector organisations. Successful use and management of ICT will continue to improve as well as enable Government to serve all its customers better (E-Government, 2011).

The Government of Kenya recently launched the Vision 2030 in which three out of the four main pillars involve objectives relating to technical skills, all of which are anchored on ICT integration, since ICT usage transcends all professions. ICT Integration is therefore the engine from which the force of propagation towards the attainment of this goal originates. In this regard ICT integration is an integral and indispensable ingredient in achieving this goal.

To achieve ICT Integration as envisaged in the Vision 2030, the country must achieve full computer literacy among the employed staff in organisations as well as the self employed citizens who offer supportive services to both organisations and individuals.

1.2 Statement of the Problem

General observation shows that a good proportion of staff in various organisations, especially the public sector, lack basic computer literacy. Although a good proportion is able to operate the computer, most of such people are only able to perform very basic operations, but are unable to use even 10% of the potential of the computer. This means that very few staff members are competent enough to use computers proficiently as a tool to perform various tasks undertaken in service delivery in their professions or duties. They are therefore not able to integrate the ICT skills with their professional skills to enable them use computers in dispensing services. Also those who use the computer to prepare documents, for example secretaries and clerical staff, use the computer only as a typewriter or a simple calculator by typing simple text documents and performing simple arithmetic calculations. In general, very few staff members make productive use of the capability of the computer. It has also been observed that exceptionally few staff members, outside the ICT department, are able to perform basic computer troubleshooting, which is expected at the level of a general user.

Organisations and governments have been trying to promote computer usage among staff especially professionals, using different methods such as workshops, seminars, sensitisation, internal training, granting study leave, incentives for innovation, industrial attachment, incorporating ICT in the curriculum as a subject in almost all courses at all academic levels, etc. In 2003, the Government of Kenya waived taxes on ICT facilities with an aim of bringing down the hitherto exorbitant prices of computers and component parts such as printers, storage media, modems, ink, projectors, etc. This saw an increase in the sale of computers, consequently leading to increased usage, but this increase still remains marginal.

Despite all these efforts, computer literacy levels in Kenya still appear low and fall far below expectation. This may be attributed to several factors both known and unknown including negative attitude and computer phobia both of which have almost escalated to the level of a syndrome. The e-government agenda cannot succeed without an ICT literate population.

It is widely believed that the usage of computer depends on age as old people tend to resist the use of new technology in general and ICT in particular. Likewise, it is also strongly believed that the usage of the computer depends on the level of education as highly educated people tend to embrace the use of ICT. Observation also points out variation in computer usage as based on area of residence with a dichotomy between urban and rural areas. Orodho (2004) concludes that the level of urbanisation in Africa is very low with less than 30% of the continent's population living in urban areas. Does this proportion have a bearing on the proportion using or integrating ICT at work?

It should be noted that so far, researches carried out in ICT integration in general, have not been carried out in Kenya. In Kenya, the existing research work has focused particularly on ICT integration in education. This points at an existing gap as computers are used across all fields and professions and transcend the teaching and learning sector.

The purpose of this research therefore is to verify these theories by investigating the factors and the level of ICT usage by staff at the place of work, and to find out the root causes of the problem of low level of ICT integration. Possible outcomes of this study include an understanding of the problem of low level of ICT integration, and the appropriate remedial measures on how to ameliorate the situation. The findings of this research will help the government, local and international organisations and other stakeholders in finding out the impediments to ICT integration and to formulate policies and plans of action, and to channel resources aimed at achieving ICT Integration in the right direction.

1.3 Objectives of the Study

General Objective

The goal of this research is to assess the current level of ICT integration and to find out the factors which influence ICT integration among staff in tertiary academic institutions, and including the factors which by themselves may not directly affect ICT integration, but along which the various integration levels are drawn.

Specific Objectives:

1. To identify clusters of variables which measure similar underlying dimensions in ICT integration among staff in tertiary academic institutions.
2. To identify the individual factors in ICT integration among staff in tertiary academic institutions.
3. To identify the institutional factors in ICT integration among staff in tertiary academic institutions.

1.4 Hypotheses of the Study

To achieve the above objectives, the following hypotheses have been formulated:

1. The variables which determine ICT integration among staff in tertiary academic institutions have no correlation and are all independent of each other.
2. There is no individual variable which is statistically significant in influencing ICT integration among staff in tertiary academic institutions.
3. There is no institutional variable which is statistically significant in influencing ICT integration among staff in tertiary academic institutions.

1.5 Justification of the Study

It has been observed that the Government of Kenya through various ministries, organisations (both public and private, national and international), NGOs and institutions of learning at all hierarchical levels, spend enormous amount of financial resources on ICT in areas encompassing hardware, software and live ware. These include:

Purchase and maintenance of ICT gadgets such as computers, printers, projectors, i-pads, modems, mobile phones, and communication equipment such as network switches, routers, firewalls and Wi-Fi equipment, and computer software such as operating systems, anti-virus and application software. Also included is the purchase of ICT consumables such as print cartridges, printing paper, Internet bandwidth, website hosting, off-site data backup, e-library services, phone air time, etc;

Setup and maintenance of local area networks and wide area networks to facilitate both internal and external communication electronically. Such networks involve both cabled and wireless media. Besides, is the setting up of ICT infrastructure nationally and internationally such as the laying of optical fibre cable both on land and under the sea, mounting of satellite dishes, terrestrial boosters, satellite systems in space, etc.

Computer training of students and staff in schools, colleges and at work through proficiency training, seminars, workshops, etc;

However, with all these logistical efforts and financial costs, ICT integration is still very low. Although general observation shows that a large proportion of Kenyans are able to use ICT facilities, a keen observation reveals that most of such people use only the very basic features of the current software in the market. It is therefore noted that all these efforts and resources appear to have made only a negligible impact on the ground. This research therefore carries out a scientific study to get to the root causes of this phenomenon by trying to identify the factors in ICT integration among staff in organisations.

1.6 Scope and Limitations of the Study

This research was carried out as a case study of five tertiary academic institutions in the Municipality of Machakos, namely Machakos Teachers Training College, Machakos University College, South Eastern University of Kenya – Machakos Campus, Scotts Theological College, and Wote Technical Training Institute. It collected data pertaining to the factors which are perceived to have a bearing on ICT integration at the workplace. The data was obtained from the staff members through administering hard copy questionnaires.

The study had certain limitations some of which are listed below:

(a) Due to both financial and time constraints, the research was limited to a case study of five academic institutions in one region. This might have compromised on the external validity of the research, and consequently limited the generalisations of its findings in that regard. In this light, the findings of this research may not apply fully to organisations in rural areas and multinational organisations.

(b) The data was obtained from primary sources and is predominantly qualitative. Some of the variables are based on personal opinion or judgement and may therefore not be very objective and accurate. The data was also collected from one geographical region - Machakos. This implies that majority of the workforce naturally come from the same geographical locality and therefore have a common socio-economic and cultural background. This according to the researcher, may compromise the external validity of the research because most variables in the research are socio-economic in nature.

(c) Due to time constraint, data was collected from only one private institution. This means that private institutions were not well represented in the sample. Consequently the findings from this research may apply better to tertiary public institutions than their private counterparts.

1.7 Assumptions in the Study

By taking all of the tertiary institutions in Machakos Town, the sample of staff taken was a true representative of the population of staff in tertiary academic institutions in Machakos Town.

The respondents understood the questions correctly as intended and answered them accurately and honestly without holding back information or giving false information.

1.8 Definition of Terms

ICT: Information and Communication Technology. The combination of hardware and software products and services and related equipment that people use to process, manage, store, access, communicate, and share information and resources. It therefore includes electronic media, telephone networks, computers, computer networks, and the Internet.

ICT integration: The effective usage of ICT facilities and equipment to perform productive work. In the context of this research, it means the effective and productive use of computers and related ICT facilities to perform ones duties in his area of responsibility for his organisation or private work.

M-Pesa: A mobile phone service that offers mobile banking that allows the subscribers to send and receive money electronically, and also to pay bills.

Kipokezi: A mobile phone service that allows subscribers to do online chatting and also exchange electronic mails via standard mobile phones.

MIS: Management Information System. This is a combination of computer science, management science, and operations research with a practical orientation towards developing systems and application software.

Digital divide: The presumably wide gap in knowledge, access and usage of ICT among different groups of people.

e-Government: Electronic Government. It is the use of ICTs to improve the service delivery and activities of public sector organisations. It consists of the digital interactions between a government and citizens (G2C), government and businesses/commerce (G2B), government and employees (G2E), and also between government and governments/agencies (G2G).

Smartphone: A mobile phone handset with computer capabilities and therefore has additional features for Internet communication and mini application software.

Firewall: This is a combination of software and hardware serving as a barrier in computer network security and placed between an organizational network and the Internet to control the incoming and outgoing network traffic by filtering information passing through it based on the configured criteria.

Router: This is a device that forwards data from one computer network to another based on the address information of its ultimate destination.

World Wide Web: A system with universally accepted standards for storing, retrieving, formatting, and displaying information in a networked environment.

Intranet: A computer network which is a subset of the Internet but owned and managed by an individual organisation or a consortium of organisations for use by its own members and customers.

Wi-Fi: Wireless fidelity is a network standard that creates wireless local area network with signals covering a short radius, but transmits data at speeds faster than advanced mobile phone networks. It is used in homes, offices, restaurants, hotels, PSV buses and airports, hot spots in colleges, etc.

Facebook: This is an on-line communication platform on the Internet in which uses interact socially and share information.

Twitter: This is an on-line social networking service that enables users to send and receive short messages of up to a maximum of 140 characters. Registered users can read and post messages, but unregistered users can only read them.

WhatsApp: This is an instant messaging application for smart phones that operates under a subscription business model. It uses the Internet to send text messages, images, video, user location and audio media messages.

Live ware: It is a term that refers to the people or human beings who use computers and other related ICT facilities, to operate the system of hardware and software.

Poverty line: It is a measure of poverty level that uses a set minimum income per day that enables a person to spend only on the very basic needs and no leisure. It may vary between countries, but the internationally set standard was US$ 1 a day, but revised by the World Bank in 2008 to US$ 1.25 a day.

CHAPTER 2: REVIEW OF LITERATURE

2.1 Introduction

This section is mainly concerned with the review and a brief analysis of related literature in the area of productive use of ICT resources among users in various fields and in various organisations with a view to understanding the factors in the ICT integration among both teaching and non-teaching staff in tertiary academic institutions. It also takes a critical look at past research work involving the relevant statistical analytic techniques available. It focuses on previous writings, research findings and conclusions pertaining to the understanding, usage and application of ICT in various fields, and the various technologies in ICT currently in use today.

2.2 Areas of Application of ICT

2.2.1 Communication and Mass Media

For two or more computers or devices to communicate with each other, they must be connected in a network. Such a network may cover a room, building, compound, town, country, continent or the whole world.

The Internet is the world's largest and widely used network. Loudon and Loudon (2001), defines the Internet as an international network of networks which are both privately and publicly owned. They further state that organisations and individuals can use the Internet to perform business transactions and exchange data regardless of whether they are located next door or on the other side of the globe. This is because computers and most ICT equipment can manipulate data in all forms – text, graphics, audio and video.

As technology transforms the world into a global village, telecommunication through satellites plays a crucial role in intercontinental and transcontinental communication. According to Hutchinson and Sawyer (1996), a satellite system uses solar-powered satellites in geostationary orbits above the earth to receive, amplify, and retransmit electromagnetic signals. This technology has enabled users to communicate worldwide, search for information from websites, perform electronic funds transfer worldwide, hold video conferences across continents, and receive live news broadcasts both locally and internationally.

Communication technology has also seen the emergence of a new concept called virtual organisations. Such organisations use computer networks to link people, assets and ideas, allying with suppliers and customers to create and distribute new products and services without being limited by traditional organisational boundaries or location (Loudon & Loudon, 2001). The academic arena has experienced the emergence of virtual universities with students spread all over the world linked together only by the Internet.

According to Shelly and Vermaat (2011), the wireless technology has made dramatic changes in the way computers communicate worldwide. Billions of home and business users use laptops, smart phones, and other mobile devices to access the Internet, send and receive electronic mail and chart on-line. The social media, especially Facebook, Twitter and WhatsApp have become the most popular means of communication especially among the youth.

2.2.2 Electronic Services

With communication technology, the Internet and intranets enable both staff and customers of an organisation to interact and get services on-line. These include e-mail, e-learning, e-commerce, e-business, e-government, e-library, etc.

A new form of service delivery by the government, called e-government, has also emerged and is completely revolutionising government service delivery. 'E-Government' refers to 'the utilization of information technology (IT), information and communication technologies (ICTs), and other web-based telecommunication technologies to improve and/or enhance on the efficiency and effectiveness of service delivery in the public sector.' (Jeong, 2007 in UN 2013).

By linking thousands of organisations and billions of individuals in a single network, the Internet creates a foundation for a vast electronic marketplace. Loudon and Loudon (2001), defines an electronic market as an information system that links together many buyers and sellers to exchange information, products, services and payments. Buyers and sellers can complete purchase and sale transactions digitally regardless of their physical location on the globe. Product selection, price verification, ordering of goods, and payment of bills is all performed electronically. Goods and services are advertised, sold and paid for world wide using the Internet as a global marketplace. Sellers create brochures, price lists, product manuals and order forms on the World Wide Web. Trading in the stock market has also not been left out.

All these combine to form e-commerce which, according to Loudon and Loudon (2001), is the process of buying and selling goods and services electronically with computerised business transactions, and include supportive activities such as marketing, advertisement, customer support, delivery and payment.

2.2.3 Research and Statistics

With the current technology in computing and the voluminous amount of data collected in most research studies of today, almost every research involves the use of the computer and the relevant statistical software as tools in the analysis of data.

Researchers in various disciplines have found computers to constitute an indispensable part of their research equipment (Kothari, 2011). Kremelberg (2011) notes that a statistical package enables the user to create variables, enter data, modify variables, create charts and graphs, and run complex analysis. There are several statistical packages currently in the market with varying abilities and features to suite the different types of data, types of analyses and types of output. Different packages also suite different users depending on their individual abilities.

Besides data analysis, specialised ICT gadgets are used in the electronic collection of data whose scientific nature and degree of accuracy render them impossible to collect manually. These include data where GPS devices are used to indicate latitude, longitude and altitude of certain points on the earth surface. GPS enabled devices are also implanted on wild animals on land and fish in the sea to enable the monitoring of their movements.

In some cases, a questionnaire is designed and posted on a website from where it is retrieved, filled on the screen and sent electronically by the respondent, hence no need data entry.

2.2.4 Management Decision Making

The management of today uses ICT both directly and indirectly to perform the day-to-day activities of data collection, storage, processing and communication, and also to assist in high level decision making. Loudon and Loudon (2001) state that contemporary systems bring about managerial changes – who has access to what information, about whom and when, and institutional core changes – what products and services are produced, under what conditions, and by who. Kothari (2006) notes that the management uses the computer to find the optimum solution from multiple options, schedule projects, manage inventory, etc.

2.2.5 Navigation Services

In the field of navigation, computers and other digital devices use the Global Positioning System (GPS) for navigation, positioning and timing. The GPS is a space-based satellite navigation system that provides location and time information anywhere on or near the earth where there is an unobstructed line of sight to at least four GPS satellites. The system provides critical capabilities to military, civil and commercial users around the world. It is freely accessible to anyone with a GPS receiver (Garmin, 2013).

The receiver computes its own location precisely then displays it on a map or in text form in terms of latitude and longitude, and may also include altitude. In addition, many GPS units also show direction and speed (Garmin, 2013).

The GPS navigation system is used in planes and ships by pilots, by security firms to track stolen vehicles which are subscribed, by companies owning a fleet of vehicles to keep track of their vehicles, by wildlife researchers to keep track of migration of animals, missiles to keep track of target, security intelligence to track criminals, and so on. All these could not be possible without the use of computers.

2.2.6 Simulation Systems and Computer Aided Design

Simulation is a technique of testing a model which resembles a real life situation, (Kothari, 2006). Simulation is used heavily in engineering in the design and testing of aircrafts, vehicles, rockets, ships, bridges, and also in businesses where processes are simulated to assist in decision making. An aircraft manufacturer, for instance, will always test a scale model of a new type of aircraft in an air tunnel before building a full-sized plane, (Kothari, 2006). He further notes that businesses processes are simulated using computers for testing the result of some business decisions before they are actually executed.

In mathematical simulation, the Monte Carlo technique, involves the use of random numbers which are generated using a computer or a calculator. In the transport sector, computers are used to train pilots (Kothari, 2011). Before handling a real aircraft, a trainee pilot is trained using an aircraft flight simulator installed in a computer system. Simulation in today’s world would not be possible without the use of a computer.

In engineering design the computer with CAD software is used for testing through simulation. This covers all branches of engineering. Without computers, man could not have landed on the moon, launched satellites, nor built 100 storied buildings or planes (Kothari, 2011).

2.2.7 Telecommuting

Through communication technology, intranets and the Internet enable employees and managers of organisations to render services pertaining to their work from their homes, or any where provided they have computers connected to the network. Information technologies such as e-mail, the Internet and video conferencing at the desktop permit tight coordination of geographically dispersed workers across time zones and cultures (Loudon & Loudon, 2001). They further observe that communication technology has eliminated distance as a factor for many types of work in many situations. Salespersons can spend more time in the field with customers and have more up-to-date information.

2.2.8 Accounting and Finance

Kothari (2011) outlines some of the uses of the computer in accounting and finance as handling payroll of personnel, office accounts, invoicing, sales analysis, stock control, financial forecasting, processing cheques, updating accounts, production of customer statements, and calculation of balance. Although a few organisations still perform these tasks manually, it is almost impossible for the large organisations.

2.3 Benefits of ICT Integration

2.3.1 Role of ICT

It has been observed that ICT has infiltrated every facet of human life by influencing virtually all the things we do, ranging from mode of learning in schools and colleges to gadgets used at home, tools at the place of work, and entertainment equipment. Life appears not to be possible without the use of ICT.

Kamuti (2009) observes that ICT is one of the driving forces of modern development and with advancement in ICT, one can live in the global village irrespective of distance and international boundaries. With ICT, interactions transcend geographical distance.

2.3.2 Benefits of ICT

It is worth noting that no organisation can claim its place in the contemporary world without the use of computers and other ICT facilities. Loudon and Loudon (2001), note that computing power which has been doubling every 18 months, has improved the performance of microprocessors 25,000 times since their invention in the early 1970's. With powerful easy-to-use software, the computer can process vast amounts of data, and simulate complex physical and logical processes.

Kamuti (2009) outlines the benefits of ICT integration to an organisation as cost reduction, improved productivity, support in decision making, and enhancement in organisational functions.

A government which has integrated ICT in its service delivery, by adopting e-government, is able to enhance service delivery to its citizens. World Bank (2011) notes that ICT technologies enable better delivery of government services to citizens, improved interactions with business and industry, citizen empowerment through access to information, or more efficient governance. The possible benefits are less corruption, increased transparency, greater convenience, growth in revenue, and reduction in cost.

2.4 Challenges in the Use of ICT

The use of ICT, like any other technology, has come as a blessing, but not without challenges and the challenges keep on changing in nature and scope as technology changes over time.

(a) **Rapid change in technology:** Technology around ICT changes so rapidly that computer hardware, computer software, and computer accessories and equipment are rendered obsolete very fast, and this in turn may lead to loss of capital to an organisation. Moving at the same pace with computer technology is generally very expensive. In 1965, Gordon Moore, a cofounder of Intel, predicted that the number of transistors on an integrated circuit would double every two years. His concept called Moore's Law has remained valid for more than 40 years (Shelly & Rosenblatt, 2010).

(b) **Security:** Organisations store a lot of data on the computer storage media, and such data form the nerve centre for their operations. In most cases, data does not originate where it is processed, and so it has to be collected at the point of origin and transmitted to the point of processing. In most cases, the result is also transmitted to an end user in a geographically different location. These necessitate the electronic transfer of data from one point to another.

Data stored in the computer and data on transmission, both need to be protected against the security threats to which they are vulnerable. Such threats encompass virus attack, hacking, electronic eavesdropping, terrorism, sabotage, environmental hazards and natural calamities. Computer hardware, software, data, storage media, transmission equipment and media, personnel, and the policies of an organisation all need protection.

The wireless technology has become very popular in data communication. Shelly and Vermaat (2011), note that although wireless access provides convenience to users, it also poses additional security risks. This is evidenced by one study in the USA which found out that 80% of the wireless networks have no security protection. The study concludes that some perpetrators connect to other peoples' wireless networks to gain free Internet access or confidential data of an organisation.

(c) **Changing trends:** Trends in the usage of ICT among customers of services and products has also been observed to be changing rapidly which poses a challenge in the ICT sector. In its annual report of 2007, IBM stated that software and services produced 77% of the total revenue while hardware accounted for only 23% of sales (Shelly & Rosenblatt, 2010).

(d) **Technology overload:** According to Shelly and Vermaat (2011), a person is perceived to have a problem with technology overload when the use of technology negatively impacts on his health, personal life and professional life. They further note that for some people, technology overload often leads to less time spent with family and has proven to be as potent a cause of divorce as gambling and substance abuse.

(e) **Poor infrastructure:** Kamuti (2009) observes that poverty, illiteracy, human capacity and poor governance are some of the challenges facing ICT usage in developing countries. These factors are at their worst in rural areas.

2.5 Factors Influencing the Use of ICT

2.5.1 Income

Any digital device or communication medium has a cost element associated with its acquisition and subscription. It is also noted that high customs duty on mobile telephones and computer equipment, as well as high cost of telephone services are deterrents to poor users - this includes both import duty and taxes on computer equipment and pricing schemes for communication services (World Bank, 2011).

2.5.2 Level of Education

Although any person with a positive attitude regardless of level of education should be able to use a computer and other ICT devices, it is no doubt that to use a computer proficiently and to integrate it into one's work, one needs some minimum level of education. Also the level of ICT integration presumably varies positively with education. In addition, Kamuti (2009) concludes that lack of staff training on computer use has a negative impact on ICT integration.

2.5.3 Gender

The concept of "gender" is sometimes confused with just improving the status and position of women as opposed to men. In fact the "gender" component to development suggests the improvement of women's social and economic status in relation to and alongside men's participation and involvement (World Bank, 2011). It further notes that women usually face higher barriers than men to the kinds of training that can make them computer literate or equip them with the skills needed for ICT-related employment.

ICTs are enabling tools for economic development and social change. Unfortunately, many women face constraints that impede them from taking advantage of the opportunities presented by the new technologies. If the dissemination of information technology proceeds without taking note of particular needs of women, they will continue to be excluded from the benefits of ICTs (World Bank, 2011).

When it involves paying for information access, women may not have the disposable income to do so or may hesitate to use family resources for information, instead of food, education or clothing (World Bank, 2011).

On socio-cultural issues, the World Bank (2011) found out the following on gender bias in attitudes towards women studying or using ICTs:

- Women often do not have equal access to ICT facilities. Frequently, information centres or cybercafés are located in places that women may not feel comfortable visiting, or that are culturally inappropriate for them to visit. Because most communication facilities in developing countries are in offices or in shared public access points, women also have problems of time-given gender-defined multiple roles, heavy domestic responsibilities, and scarce leisure hours among others.
- In a project for rural farmers in Peru, when women undertook ICT training with men, the men mocked them by saying that computers are for men, not women.
- In Uganda, girls did not get equal access to the limited number of computers installed in schools (under a WorldLinks Program) because of the socio-cultural norm that "girls do not run." Boys ran to the computers first and refused to give way for the girls.
- In India, aggressive boys pushed the girls away and prevented them from using the computers.

2.5.4 Infrastructure and Area of Residence

All ICT equipment, regardless of the design, whether desktop, laptop, smartphone or i-pad, use electric power as a source of energy. This means that ICT facilities cannot be used in an area where there is no electricity from the national grid or alternative source of power such as generator or solar system. Kasumbu (2007) in Kamuti (2009) notes that many of the benefits of ICT are found in urban areas, due to the comparatively well developed. Also generally the educated population tend to migrate from rural to urban areas due to several pull factors resulting in another divide between urban and rural areas based on education. This further exacerbates the gap. However, for this research, electricity is a constant and not a variable since the scope of the study is Machakos Town which by nature has electricity everywhere.

According to World Bank (2011), concentration of the ICT infrastructure in urban areas limits the accessibility of new technologies for women because most poor women in developing countries live in rural areas. Internet connectivity is available mainly in major towns in most developing countries, especially in Africa. It is observed that in rural areas, there are fewer relay stations for mobile phones compared to urban areas.

2.5.5 Government Policy

For several years, stakeholders in the ICT sector were operating in their own individual ways because there were no standards, policies or laws developed and enforced by the government concerning ICT equipment, networks, products, services, etc. In 2006, the Government of Kenya published its ICT Policy to ensure the availability of accessible, efficient, reliable and affordable ICT services to citizens.

As a result of liberalisation of the mobile cellular market by the Government, current statistics in the National ICT Policy indicates that there were 260,000 fixed telephone line subscribers and 3.0 million cellular mobile subscribers by June 2004, translating into a fixed tele-density of 0.75 per hundred inhabitants for fixed and 9.75 per hundred inhabitants for mobile phones against the world average of 19 and 21 respectively (Ministry of Information and Communications, 2006).

The Ministry further notes that there were 73 registered ISPs, 16 of which are active, approximately 1,030,000 users and over 1000 cyber cafes and telephone bureaus by June 2005, approximately 11,500 public phones installed in the country.

The Government has a responsibility to all who enquire and seek for its services, and therefore there is a need to minimize the effects of the "digital divide‟ by ensuring that Kenyans have equal access to all services and information. This means ensuring that information is available to all people, regardless of disability, bandwidth, location, reduced mobility and age or language barriers (Directorate of e-Government, 2011).

According to the Constitution of Kenya (2010), every citizen has the right of access to information held by the State, and also the State shall publish any important information affecting the nation. It is no doubt that ICT makes this easier, faster, cheaper and convenient as information posted on an Internet website is available to everybody anytime anywhere as long as there is Internet connectivity.

2.5.6 Organisational Policy

Mobile services such as banking, payments, money transfers and wireless Internet access offered by mobile telecommunication service providers in Kenya, have resulted in proliferation of ICT integration by citizens regardless of income and area of residence. According to UNCTAD

(2002), wireless technologies have made in-rods even in relatively low income areas where pre-paid cards allow almost universal access to subscribers.

2.6 Application of PCA and Logistic Regression

Mwai (2005) in his study of the application of principal component analysis in foreign exchange market variations used principal component analysis to analyse the exchange rates of different currencies at the Nairobi Stock Market. He used data comprising 16 currencies and through the PCA technique, reduced the dimension of the data to only two principal components.

The first two principal components extracted were able to account for up to 91.547% of the total variation in the data. The study has no specific finding as he states in part of his conclusion that although PCA achieves the objective of examining patterns among variables, it is not an end in itself.

Wambugu (2008) in her study of the utilisation of free primary education (FPE) in urban areas in Kenya used the logistic regression model to analyse data.

The response variable was utilization of FPE categorised as using and not using. The explanatory variables were province, sex of child, education level, poverty status and sex of household head.

She found out the following:

- The utilisation of FPE increases as children progress from standard 1 to standard 8.
- The children from poor families are 5.4 times more likely to use the FPE compared to those from rich backgrounds.
- The children from male headed households are 4% less likely to use the FPE compared to those from female headed households.
- The children from good backgrounds are 22% more likely to temporarily withdraw from school compared to those from poor backgrounds.
- The male children are 11% less likely to temporarily withdraw from school compared to female children.

Kamui (2008) in his study to identify the key determinants of academic performance in primary schools on the basis of school supply, used the binary logistic regression model to analyse data. The schools were classified into only two categories – performing and non-performing schools.

Response variable was academic performance (performing, non-performing) while the explanatory variables were many but focused on supply, composition of school management, meetings, etc.

He found out the following:

- The teacher factors that were presumed to affect school performance were all found as not significant.
- Among the physical infrastructural factors, schools with permanent classrooms perform better compared to those with temporary classrooms.
- A school with a small class size is 1.02 times more likely to perform better compared to a school with a large class size.
- In terms of regional factors, only province/district was found to be significant.
- A school in which the books purchased are recommended by the head teacher is 11.14 times more likely to perform better compared to a school in which the recommendation is made by the class teacher.
- A school with a feeding programme is 2.63 times more likely to perform better compared to a school without a feeding programme.

2.7 Existing Gaps in Past Research

In conclusion, the past research work on ICT integration has basically focused on the teaching and learning in schools in various parts of the country, and in the management of NGOs in one county - Embu. However, no meaningful research appears to have focused on ICT usage in general and in particular in non-teaching activities in tertiary institutions.

Also, past researches focused on the use of the computer only with no classification of the types, and therefore did not include the use of other ICT gadgets such as printers, projectors, mobile phones, i-pads, etc. Another gap is that the ICT infrastructure such as local area network, Internet access, Internet services and Internet bandwidth were not explicitly analysed.

Another existing gap identified is that the past researches did not focus on the use of specific computer software. They therefore did not take into account the large variety of software in the market today.

In terms of methodology, the past researches used the generalised linear models focused on binary logistic regression handling only two possible outcomes. The researches also did not specify how any possibility of multi-collinearity between the explanatory variables was resolved. This research went further to handle a situation in which the outcome variable has three possible values and they are ordinal in nature.

This research has therefore focused on bridging theses gaps stated above.

2.8 Conceptual Framework

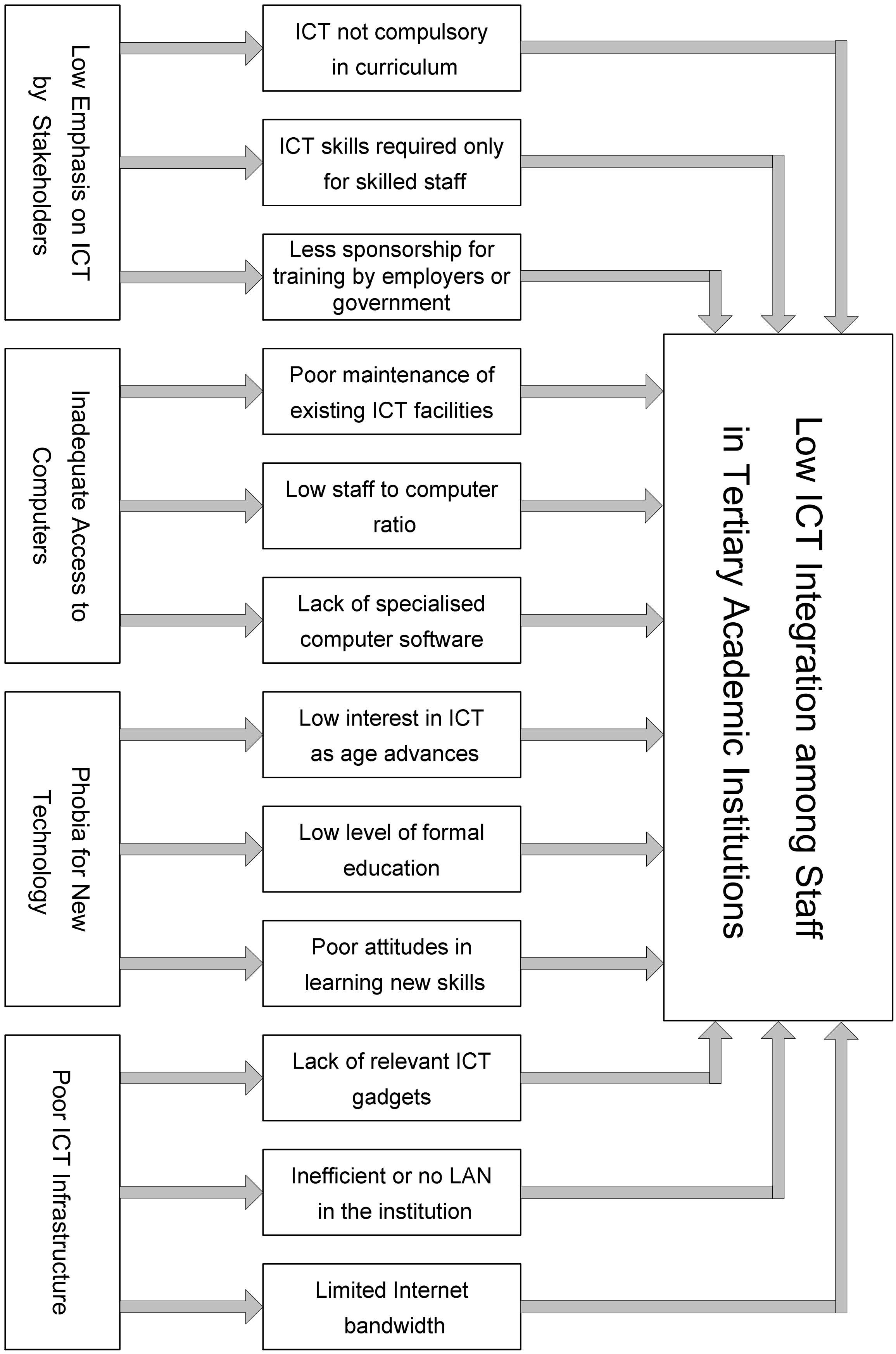

Figure 2.1: Conceptual Framework

2.9 Logical Framework Matrix

	Narrative Summary	Verifiable Indicators	Means of verification	Assumptions
Goal	To achieve functional ICT integration at the workplace in all professions by the year 2030.	Ability by all relevant staff to professionally integrate ICT by using the computer as a tool in their work. ICT integration becomes a way of life.	The staff exhibiting ability to competently use computers and ICT resources to prepare documents, communicate and perform office tasks.	No major technological change will occur in the intervening period. Funds will be available.
Purpose or Objective	Identification of the causes of low ICT integration. Identification of the remedial measures to enhance ICT integration.	Increment in the number of documents prepared using the computer. Almost all routine operations computerised. Improved efficiency in storage, retrieval, and transmission of data. Better policies in line with ICT Integration.	All documents prepared in the computer. All routine work performed on the computer. All staff making maximum use of all the available software.	Economic factors like taxes, electricity tariffs, curriculum, prices of ICT facilities will remain fairly constant.
Outputs	All data used by an institution stored on and accessed from the computer. Improved efficiency in the delivery of services. High quality of output documents. Improved accuracy of output. Fast communication: through e-mail.	More funding for ICT infrastructure. Faster service delivery. High quality output documents. Improved accuracy in outputs. Internal and external communication through LAN and Internet. Less paperwork in offices. Increased use of ICT resources.	ICT budget in institutions. Tasks performed faster and more efficiently. Accurate results obtained from information processing. Few documents in paper form. All skilled staff using ICT facilities to perform their tasks.	No new major threat to computer security. Communication networks will be functional. Funds will be made available for Internet services.
Activities	The storage and retrieval of data, preparation of documents, and processing of data carried out on the computer.	Data processed on, stored onto and retrieved from computer. Communication done electronically. Services delivered electronically. Documents coming from the computer.	Evidence of documents being processed on, stored on and retrieved from the computer. Evidence of data electronically processed as staff work. No hand written documents.	Approvals such as signed ones will be done as scheduled by the relevant officers. Procurement procedures will not change.

Table 2.1: Logical framework matrix

2.10 Operational Framework

2.10.1 Factors Affecting ICT Integration – Explanatory Variables

2.10.1.1 Individual Employee Factors

(i) Demographic factors

- Gender
- Age of staff
- Staff category - teaching, non-teaching
- Duration in employment
- Level of education of staff
- Age at first usage of the computer
- Management level
- Profession by training

(ii) ICT Knowledge and usage factors

- Frequency of use of the computer
- Place of acquisition of computer skills
- Mode of acquisition of ICT skills
- usage Internet services
- Depth of knowledge of e-mail services
- Level of job satisfaction
- Mobile phone type owned
- Use of mobile phone for Internet services
- Compulsory inclusion of ICT during college training
- Competency level in general computer usage
- Application software installation competency level
- Basic computer troubleshooting competency level

2.10.1.2 Institutional Factors

(i) General Information

- Name of institution
- Type of institution - public, private
- Status of the institution - tertiary, university
- Requirement of ICT skills by employer at recruitment
- Access to Internet in the office
- Availability of software you require
- Anti-virus software installation

(ii) Hardware factors

- Unlimited access to desktop computer
- Unlimited access to computer printer
- Unlimited access to LCD projector

(iii) ICT Infrastructure and Capacity

- Highest level of training offered by institution
- Level of education of head of ICT
- Availability of local area network - LAN
- Optical fibre LAN communication medium
- Wireless LAN communication medium
- Internet connectivity
- Internet bandwidth
- Installation of institutional MIS
- Subscription to e-library resources
- Number of staff
- Number of computers/devices by category
- Initiator of ICT gadget acquisition
- Involvement in drawing ICT budget
- Number of professionally trained ICT staff in the institution
- ICT maintenance type
- Maintenance frequency if preventive
- Student enrolment (no of regular students)

2.10.1.3 Software usage indicators

- Word processing
- Spreadsheet
- Presentation
- Desktop publishing
- Database management system - DBMS
- Project management
- Accounting package
- Statistical package
- Computer aided design - CAD

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

This chapter outlines and spells out the scope of the study, the observational units in the study, the target population, the type of data collected, the methods used in the collection of data, the techniques and models used in the analysis of data and the corresponding output, and the various statistical inferences drawn from the analysis. It also extends to encompass the justification for the selection of each technique used at every stage of research and the parameters used in making statistical inference.

3.2 Research Design

This research is a case study of institutions in Machakos Town. The observational units are the members of staff (both teaching and non-teaching) in the tertiary academic institutions. The research collected and analysed data as they were observed in the field without any alteration or control of any attribute. It is therefore a survey research and not an experimental research.

Because of time, financial and logistical constraints in the research, data collection was not triangulated, but instead was collected at once using basically one method – questionnaire. To minimise bias in the data, samples were selected using a combination of multiple random sampling techniques. This ensures that the weaknesses of one technique are compensated by the strengths of the other. The main focus of attention is the factors in ICT integration by individual staff, but because some factors are institutional in nature, the institutions themselves also serve as observational units to enable analysis of factors at that level.

The data is basically primary in terms of source, but is both qualitative and quantitative in nature in order to cater for the various types of variables. This ensures that the analysis combines the strengths and compensates for the weaknesses of qualitative and quantitative analysis (Bryman, 2001 in Nsibirano, 2009).

3.3 Target Population

The research has its target population defined as all the employed staff in tertiary learning institutions in Machakos Town. Five institutions were selected through purposive sampling to form a case study. The research excluded the small private colleges because of their unstructured nature of recruitment and operations. The members of staff in this study included all levels along the staff hierarchical structure, and staff on casual terms if applicable – part-time lecturers. It was difficult to obtain a list of all the staff of an institution from the management as secondary data to form the sampling frame due to circumstances beyond the control of the researcher, and therefore the total number of employees was used to create a dummy list forming the sampling frame.

3.4 Collection of Data

Data was collected from primary sources through questionnaire administered by the researcher himself and assisted by enumerators. In this case, an individual staff member was both an observational unit and a sampling unit. There were two different questionnaires for the different staff categorised as teaching and non-teaching staff and ICT technical staff.

The questionnaire for the teaching and non-teaching staff collected data on attributes which are unique to the individual staff member while the questionnaire for the ICT technical staff collected data pertaining to the attributes of the entire institution as an entity.

A combination of multi-stage sampling and stratified random sampling techniques were used. At the first sampling stage, the institutions were selected through purposive sampling. All institutions selected were generally structured in their operations, while the small private colleges which are generally unstructured were excluded.

Within each institution, the staff was split into two strata - teaching staff and non-teaching staff. Within each staff category in each institution, a random sample was selected through proportional allocation, giving a sample size proportional to the size of that stratum. However, both random and quasi-random sampling techniques were used within a given stratum. For the ICT technical staff, purposive sampling was used since the questionnaire collected technical information on ICT which could not be adequately provided by non-ICT staff.

The sample size was determined from the population size using the Cochran's (1977) formula for categorical data.

The sample size is given by:

$$n = \frac{N}{1+N(\alpha)^2}$$

where N - population size

n - sample size

α - set level of significance (taken as 0.05)

Institution	Population size		Sample size	
	Teaching	Non-teaching	Teaching	Non-teaching
Institution 1	84	11	41	5
Institution 2	172	70	84	34
Institution 3	40	10	19	5
Institution 4	12	10	6	5
Institution 5	10	6	5	3
Total	318	107	155	52
Overall Total	425		207	

Table 3.1: Population and Sample Sizes

After determining the overall sample size as 207 from the population size of 425, stratified random sampling was used with a sampling fraction of 207/425 applied to the sub-populations of various institutions and categories of staff as shown in Table 3.1.

3.5 Analysis of Data

The research used Principal Component Analysis (PCA) and Generalised Linear Models (GLMs) in the analysis of data. This is because the response variable, which in this case is the level of ICT integration among staff, is a categorical variable with three levels - high integration, low integration and no integration. The explanatory variables are many – 39 variables and of different types with some categorical and others numerical. Therefore there was very high possibility of multi-collinearity among the variables and also need to reduce the number of variables by retaining only the significant ones.

Classification of respondents into any of the three levels of ICT integration is based on the following criteria:

High: Ability and usage of specialised software or both advanced and ordinary features of the general application packages to perform tasks related to one's responsibilities and duties at work in his profession. All or most of the tasks which can be computerised are performed using the computer. This also includes the usage of the Internet to browse and download information, get e-services, fill on-line forms, watch and read digital news, download software from the Internet, install software and backup documents on removable media (CDs, DVDs, memory cards, and external HDDs).

Low: Ability and usage of the computer is limited to very basic computer facilities such as creating and editing simple documents, basic data entry, printing of documents, sending and receiving e-mail, and using mobile phones for e-payments and mobile banking.

None: No usage of the computer as a tool to perform routine tasks at the place of work in one's profession. ICT usage is limited to very basic mobile phone services such as sending and receiving calls and, sending and receiving messages, and using phones for mobile money transfers.

The level of ICT integration as the response variable is by nature a latent variable, and therefore the data about the respondent collected in the questionnaire on ICT usage, assisted in the verification of the self classification of the respondents into one of these three levels.

The response variable denoted by Y is the level of ICT integration. The data was majorly analysed using two statistical techniques: principal component analysis and ordinal logistic regression.

3.5.1 Definition of explanatory variables used in the analysis

Employee demographic factors

X_1 - Gender (*male*, female)
X_2 - Age (no. of years)
X_3 - Staff category (*teaching*, non-teaching)
X_4 - Duration in employment (no. of years)
X_5 - Level of education (*secondary*, certificate, diploma, bachelor, masters/Ph.D.)
X_6 - Age at first usage of the computer (no of years)
X_7 - Management level (top, middle, *bottom*)

Employee ICT knowledge and usage factors

X_8 - Frequency of use of the computer (daily, occasionally, none)
X_9 - Place of ICT skills acquisition (school, college, workplace, home)
X_{10} - Mode of acquisition of ICT skills (formal, informal)

X_{11} - Internet e-mail usage (yes, no)

X_{12} - Internet browsing usage (yes, no)

X_{13} - Internet social media usage (yes, no)

X_{14} - Level of job satisfaction (v. dissatisfied, dissatisfied, moderate, satisfied, v. satisfied)

X_{15} - Mobile phone type owned (*ordinary*, smartphone, both)

X_{16} - Compulsory inclusion of ICT during college training (yes, *no*)

Institutional general information

X_{17} - Institution (name or code)

X_{18} - Type of institution (public, private)

X_{19} - Status of institution (tertiary, university)

X_{20} - ICT skills required at recruitment (yes, no)

X_{21} - Internet access in the office (yes, no)

X_{22} - Availability of all software required (yes, *no*, some)

X_{23} - Installation of anti-virus software in one's computer (yes, *no*)

Institutional infrastructure and capacity factors

X_{24} - Access to a desktop computer (yes, no)

X_{25} - Access to a computer printer (yes, no)

X_{26} - Access to an LCD projector (yes, no)

X_{27} - Highest level of training offered by institution (certificate, diploma, degree)

X_{28} - Level of education of head of ICT (certificate, diploma, bachelor, masters/Ph.D.)

X_{29} - LAN medium – optical fibre (yes, no)

X_{30} - LAN medium – wireless (yes, no)

X_{31} - Internet connectivity bandwidth (mbps)

X_{32} - Installation of institutional MIS (yes, no)

X_{33} - Subscription to e-library resources (yes, no)

X_{34} - Staff-computer ratio

X_{35} - Staff-printer ratio

X_{36} - Initiator of ICT gadget acquisition (end-user, ICT staff, administration)

X_{37} - Involvement in drawing ICT budget (yes, no)

X_{38} - Number of trained ICT support staff

X_{39} - Student enrolment

The questionnaire also captured additional variables designed to assist in verifying the self classification by a user in terms of level of ICT integration.

3.6 Principal Component Analysis

3.6.1 Introduction

Principal component analysis (PCA) is a statistical technique for identifying groups or clusters of variables by reducing a large data set with several variables to a more manageable size while retaining as much of the original information as possible. It solves the problem of multi-collinearity by combining variables which are collinear into one factor. The data reduction is achieved by identifying the variables which correlate highly within a group but do not correlate with variables outside that group (Field, 2009). This means that there is high correlation between variables within a cluster forming a factor but no correlation between clusters or factors. In this context, a factor is defined as a cluster or group of linearly correlated variables which measure aspects of the same underlying dimension.

Principal component analysis was initially introduced as a means of fitting planes by orthogonal least squares, but the use was later extended to analysing correlation structures (Mwai, 2005). It is therefore concerned with explaining the structure of the variance-covariance matrix and the correlation matrix of a set of multivariate data through a set of linear combination of these variables.

PCA uses an orthogonal linear transformation to convert a set of observations of highly correlated variables into a new set of linearly uncorrelated variables called principal components. As a dimension reduction technique, the number of principal components is less than (or equal to) the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance by accounting for as much of the variability in the data as possible, and each succeeding component in turn has the next largest possible variance subject to the constraint that it is orthogonal to (uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed (Wikipedia, 2014). PCA is sensitive to the relative scaling of the original variables and therefore the correlation matrix is preferred over the covariance matrix in the determination of principal components.

Principal component analysis is a data analytic method which provides a specific set of projections which represent a given data set in fewer dimensions (Hewson, 2009). Therefore given a set of linearly correlated variables $x_1, x_2, x_3, \ldots, x_k$ a new set of linearly uncorrelated variables $z_1, z_2, z_3, \ldots., z_q$ are obtained each of which is a linear combination of the x variables, and $q \leq k$.

Hewson (2009), outlines the reasons for the use of principal component analysis:

- Dimension reduction by providing a specific set of projections which represent a given data set in fewer dimensions.
- PCA enables the transformation of correlated variables into a new set of uncorrelated factors.
- PCA finds linear combinations of data which have relatively large variability i.e. maximum variance.

Mwai (2005) outlines the uses of principal component analysis as:

- Data reduction: Although a full set of correlated variables are required to produce the total variability, often much of this variability can be accounted for by a new set of fewer but uncorrelated variables each of which is a linear combination of the original variables.
- Interpretation: PCA often reveals relationships that could not easily be noticed among the variables.
- Monitoring quality and stability: Where the multivariate data is normally distributed, a control ellipse with its major and minor axes corresponding to largest and second largest principal components respectively can be fitted so that points lying outside the contour are considered to be unstable.
- Visual aid for grouping variables: By plotting the component loadings, variables which are highly correlated with each other tend to cluster suggesting a natural grouping.

3.6.2 Eigen values and Eigen vectors

Two vectors $\underline{a}$ and $\underline{b}$ are orthogonal if

$$\underline{a}^T \underline{b} = a_1 b_1 + a_2 b_2 + a_3 b_3 + \ldots + a_k b_k = 0$$

A vector $\underline{a}$ is said to be normalised if $\underline{a}^T \underline{a} = 1$

Two vectors $\underline{a}$ and $\underline{b}$ are are said to be orthogonal and of unit length if

$\underline{a}^T \underline{b} = \underline{b}^T \underline{a} = 0$

$\underline{a}^T \underline{a} = 1$ and $\underline{b}^T \underline{b} = 1$

Therefore the vectors $\underline{a}$ and $\underline{b}$ are said to be orthonormal (orthogonal and normalised).

In general, a vector $\underline{a}$ is orthonormal if

$$\underline{a}_i^T \underline{a}_j = \begin{cases} 0 & \text{for} \quad i \neq j \\ 1 & \text{for} \quad i = j \end{cases}$$

An eigen vector $\underline{e}$ of a square matrix **A** is said to be normalised if

$\underline{e}^T \underline{e} = 1$ and $\underline{e} = \dfrac{\underline{a}}{\sqrt{\underline{a}^T \underline{a}}}$

A square matrix $\mathbf{E}$ comprising the eigen vectors such that $\mathbf{E} = (\underline{e}_1, \underline{e}_2, \underline{e}_3, \ldots.. \underline{e}_k)$, therefore has its columns normalised and mutually orthogonal.

$\mathbf{E}$ is therefore an orthogonal matrix such that

$\mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}$ where $\mathbf{I}$ is the corresponding identity matrix.

It can therefore be observed $\mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}$, but from the laws of matrices, we have $\mathbf{E}^{-1}\mathbf{E} = \mathbf{I}$

Therefore $\mathbf{E}^{\mathrm{T}} = \mathbf{E}^{-1}$

In general, if a square matrix **A** is orthogonal, then $\mathbf{A}^{\mathrm{T}} = \mathbf{A}^{-1}$, hence the inverse and the transpose of the matrix are equal.

Given a square matrix **A** of size k X k, a scalar λ and a non-zero vector $\underline{x}$ can be determined such that:

$$|\mathbf{A} - \lambda\, \mathbf{I}| = 0$$

is a polynomial equation of degree k in λ and λ_i's are the roots of the equation and also represent the eigen values.

For each eigen value λ_i, there is a corresponding eigen vector $\underline{e}_i$ which can be determined by solving the equation:

$\mathbf{A}\, \underline{e}_i = \lambda_i\, \underline{e}_i$ or simply

$$(\mathbf{A} - \lambda_i \mathbf{I})\, \underline{e}_i = \underline{0}$$

The eigen vectors $\underline{e}_i$ are both orthogonal and normalised and correspond to unequal eigen values λ_i

Also $trace\,(\mathbf{A}) = \sum_{i=1}^{k} \lambda_i$ and $|\mathbf{A}| = \prod_{i=1}^{k} \lambda_i$

3.6.3 Derivation of Principal Components

Derivation of the principal components of the observations proceeds as follows:

Given a random vector $\underline{x}$, a vector of transformed random variables $\underline{z}$ is given by

$$\underline{z} = \mathbf{E}^{\mathrm{T}} \left(\underline{x} - \underline{\bar{x}} \right)$$

Therefore we need to find matrix $\mathbf{E}$ such that $var(\underline{z})$ is maximised.

In this case $\mathbf{E}$ must be orthogonal and normalised by satisfying the condition $\mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}$.

$$var\,(\underline{z}) = var\,(\underline{e}^T \underline{x}) = \mathrm{E}\,(\underline{e}^T \underline{x})^2$$ where $\mathbf{E}$ is a matrix of all $\underline{e}$'s.

$$= \frac{1}{n-1} \Sigma \left(\underline{z}_i - \underline{\bar{z}} \right)^2 = \frac{1}{n-1} \Sigma\, \underline{e}^T \left[\left(\underline{x}_i - \underline{\bar{x}} \right) \left(\underline{x}_i - \underline{\bar{x}} \right)^T \right] \underline{e}$$

$$= \underline{e}^T\, \hat{\Sigma}\, \underline{e} \quad \Leftrightarrow \quad \mathrm{E}^{\mathrm{T}} \hat{\Sigma}\, \mathrm{E}$$

$$= Diag\left(\hat{\lambda}_1,\ \hat{\lambda}_2,\ \hat{\lambda}_3, \ldots ,\hat{\lambda}_k \right)$$

To determine $\underline{e}$ which maximises the variance, we compute

$$max\,(\underline{e}^T\, \hat{\Sigma}\, \underline{e}) \qquad \text{subject to} \qquad \underline{e}^T\, \underline{e} = 1$$

where Σ is the variance-covariance matrix.

Applying a Lagrange multiplier λ in the form,

$$\frac{d}{dx_i} f(\underline{x}) - \lambda \frac{d}{dx_i}\, g(\underline{x}) = 0$$ at a stationary point

$$\varphi(x) = f(x) - \lambda\{g(x) - 1\}$$

Applying this in the PCA by substituting $\underline{e}$ for x and then differentiating with respect to $\underline{e}$ gives:

For the first principal component, we have φ

$$\varphi\left(\underline{e}_1\right) = \underline{e}_1^{\,T}\,\mathbf{\Sigma}\,\underline{e}_1 - \lambda_1\left(\underline{e}_1^{\,T}\,\underline{e}_1 - 1\right)$$

$$\frac{d\varphi_1}{d\underline{e}_1} = 2\,\mathbf{\Sigma}\,\underline{e}_1 - 2\,\lambda_1\,\underline{e}_1$$

Equating the derivative to zero and rearranging gives

$$\mathbf{\Sigma}\,\underline{e}_1 = \lambda_1\,\underline{e}_1$$

$$(\mathbf{\Sigma} - \lambda_1\mathbf{I})\,\underline{e}_1 = \underline{0}$$

If $\underline{e}_1^{\,T}\,\underline{e}_1 = 1$ but $\underline{e}_1 \neq 0$, then the matrix $\mathbf{\Sigma} - \lambda_1\mathbf{I}$ is singular and therefore

$$|\,\mathbf{\Sigma} - \lambda_1\mathbf{I}\,| = 0$$

Therefore λ_1 is found as the largest eigen value of $\mathbf{\Sigma}$ and $\underline{e}_1$ as the corresponding eigen vector of the first principal component.

The second principal component, in addition to assuming orthogonality of the eigen vectors $\underline{e}_1$ and $\underline{e}_2$, it is also assumed that the two variables $\underline{z}_1$ and $\underline{z}_2$ are uncorrelated i.e.

$$cov\left(\underline{z}_1,\ \underline{z}_2\right) = 0 \quad\Leftrightarrow\quad cov\left(\underline{e}_2^{\,T}\underline{x},\ \underline{e}_1^{\,T}\underline{x}\right) = 0,\quad \text{so}$$

$$\underline{e}_2^{\,T}\,\mathbf{\Sigma}\,\underline{e}_1 = \underline{e}_1^{\,T}\,\mathbf{\Sigma}\,\underline{e}_2 = 0 \quad\Leftrightarrow\quad \underline{e}_2^{\,T}\,\underline{e}_1 = \underline{e}_1^{\,T}\,\underline{e}_2 = 0$$

$$\varphi\left(\underline{e}_2\right) = \underline{e}_2^{\,T}\,\mathbf{\Sigma}\,\underline{e}_2 - \lambda_2\left(\underline{e}_2^{\,T}\,\underline{e}_2 - 1\right) - \mu\left(\underline{e}_1^{\,T}\,\mathbf{\Sigma}\,\underline{e}_2\right)$$

$$\Rightarrow\quad \frac{d\varphi_2}{d\underline{e}_2} = 2\,\mathbf{\Sigma}\,\underline{e}_2 - 2\,\lambda_2\,\underline{e}_2 - \mu\,\mathbf{\Sigma}\,\underline{e}_1$$

Equating the derivative to zero and pre-multiplying throughout by $\underline{e}_2$ gives

$$2\,\underline{e}_2^{\,T}\,\mathbf{\Sigma}\,\underline{e}_2 - 2\,\underline{e}_2^{\,T}\,\lambda_2\,\underline{e}_2 - \mu\,\underline{e}_2^{\,T}\,\mathbf{\Sigma}\,\underline{e}_1 = 0 \qquad \text{but} \quad \underline{e}_2^{\,T}\,\mathbf{\Sigma}\,\underline{e}_1 = 0$$

Rearranging gives

$$\mathbf{\Sigma}\, \underline{e}_2 = \lambda_2\, \underline{e}_2$$

$$(\mathbf{\Sigma} - \lambda_2 \mathbf{I})\, \underline{e}_2 = \underline{0}$$

If $\underline{e}_2^{\,T}\, \underline{e}_2 = 1$ but $\underline{e}_2 \neq 0$, then the matrix $\mathbf{\Sigma} - \lambda_2 \mathbf{I}$ is singular and therefore

$$|\,\mathbf{\Sigma} - \lambda_2 \mathbf{I}\,| = 0$$

Therefore λ_2 is found as the second largest eigen value of $\mathbf{\Sigma}$ and $\underline{e}_2$ as the corresponding eigen vector of the second principal component.

The i th principal component, in addition to assuming orthogonality of the eigen vectors $\underline{e}_i$ and $\underline{e}_j$, it is also assumed that the two variables $\underline{z}_i$ and $\underline{z}_j$ are uncorrelated i.e.

$$cov\left(\underline{z}_i\,,\ \underline{z}_i\right) = 0 \ \Leftrightarrow \quad cov\left(\underline{e}_i^{\,T}\underline{x}\,,\ \underline{e}_j^{\,T}\underline{x}\right) = 0\,, \quad \text{for } i \neq j \quad \text{so}$$

$$\underline{e}_i^{\,T}\,\mathbf{\Sigma}\,\underline{e}_j = \underline{e}_j^{\,T}\,\mathbf{\Sigma}\,\underline{e}_i = 0 \qquad \Leftrightarrow \qquad \underline{e}_i^{\,T}\,\underline{e}_j = \underline{e}_j^{\,T}\,\underline{e}_i = 0$$

Applying the Lagrange multiplier, differentiating w.r.t. $\underline{e}_i$ and equating to zero gives

$$\mathbf{\Sigma}\, \underline{e}_i = \lambda_i\, \underline{e}_i$$

$$(\mathbf{\Sigma} - \lambda_i \mathbf{I})\, \underline{e}_i = \underline{0}$$

If $\underline{e}_i^{\,T}\, \underline{e}_i = 1$ but $\underline{e}_i \neq 0$, then the matrix $\mathbf{\Sigma} - \lambda_i \mathbf{I}$ is singular and therefore

$$|\,\mathbf{\Sigma} - \lambda_i \mathbf{I}\,| = 0$$

Therefore λ_i is found as the i th largest eigen value of $\mathbf{\Sigma}$ and $\underline{e}_i$ as the corresponding eigen vector of the i th principal component.

This process continues until the k th eigen value and the k th eigen vector are determined.

Formulating the sample principal components gives:

The first principal component, z_1 is the linear combination

$$z_1 = e_{11}x_1 + e_{21}x_2 + e_{31}x_3 + \ldots. + e_{k1}x_k = \underline{e}_1^T \underline{x}$$

where sample variance is maximum among all such linear combinations.

A limit is set on the variance of z_1, by setting a normalising constraint on the co-efficients $e_{11}, e_{21}, e_{31}, \ldots, e_{k1}$

of x which gives the condition

$$\underline{e}_1^T \underline{e}_1 = 1$$

The second principal component, z_2 is the linear combination

$$z_2 = e_{12}x_1 + e_{22}x_2 + e_{32}x_3 + \ldots. + e_{k2}x_k = \underline{e}_2^T \underline{x}$$

where sample variance is maximum among all such linear combinations subject to the two conditions.

$$\underline{e}_2^T \underline{e}_2 = 1$$

$$\underline{e}_2^T \underline{e}_1 = \underline{e}_1^T \underline{e}_2 = 0$$

The third principal component, z_3 is the linear combination

$$z_3 = e_{13}x_1 + e_{23}x_2 + e_{33}x_3 + \ldots. + e_{k3}x_k = \underline{e}_3^T \underline{x}$$

where sample variance is maximum among all such linear combinations subject to the two conditions.

$$\underline{e}_3^T \underline{e}_3 = 1$$

$$\underline{e}_3^T \underline{e}_2 = \underline{e}_2^T \underline{e}_3 = 0$$

The i th principal component, z_i is the linear combination

$$z_i = e_{1i}x_1 + e_{2i}x_2 + e_{3i}x_3 + \ldots. + e_{ki}x_k = \underline{e}_i^T \underline{x} \qquad \text{for } i = 1, 2, 3, \ldots, k$$

where sample variance is maximum among all such linear combinations subject to the two conditions.

$$\underline{e}_i^T \underline{e}_i = 1$$

$$\underline{e}_i^T \underline{e}_j = \underline{e}_j^T \underline{e}_i = 0 \qquad \text{for } i \neq j$$

3.6.4 Covariance and Correlation Matrices in Principal Component Analysis

Given the variance-covariance matrix $\mathbf{\Sigma}$, the vectors $\underline{e}_1, \underline{e}_2, \underline{e}_3, \ldots, \underline{e}_k$ represent the eigen vectors corresponding to the eigen values $\lambda_1, \lambda_2, \lambda_3, \ldots., \lambda_k$ and

$$\text{total sample variance} = S_1^2 + S_2^2 + S_3^2 + \cdots + S_k^2 = \sum_{i=1}^{k} S_i^2 = trace\,(\mathbf{\Sigma})$$

$$= \sum_{i=1}^{k} \lambda_i \qquad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots. \geq \lambda_k \geq 0$$

where S_i^2 is the sample variance of x_i

$$\text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} S_{11}^2 & S_{12} & \ldots. & S_{1k} \\ S_{21} & S_{22}^2 & \ldots. & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{k1} & S_{k2} & \ldots & S_{kk}^2 \end{bmatrix}$$

$$cov(x_i, x_j) = \boldsymbol{S}_{ij} = \lambda\, e_{ij} \qquad \text{for } i \neq j$$

In general, given the variance-covariance matrix $\mathbf{S}$ of size k X k with eigen value – eigen vector pairs $(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), (\lambda_3, \underline{e}_3), \ldots., (\lambda_k, \underline{e}_k)$,

the i th sample principal component is given by the linear combination

$$z_i = \underline{e}_i^T \underline{x} = e_{i1}x_1 + e_{i2}x_2 + e_{i3}x_3 + \ldots. + e_{ik}x_k \quad \text{for } i = 1, 2, 3, \ldots, k$$

has a sample mean $\underline{e}_i^T \bar{x}$, sample variance $\underline{e}_i^T \mathbf{S}\, \underline{e}_i$ and sample covariance $\underline{e}_i^T \mathbf{S}\, \underline{e}_j$ for the two variables x_i and x_j for $i \neq j$

Since the variance-covariance matrix **S** is in the same units and magnitude of measurement as those of the original variables, it is more appropriate to use the correlation matrix **R** which is standardised and has no bearing on the units and magnitude of the original data.

Also, if variables are recorded on widely differing scales, a principal component analysis of the covariance matrix will largely reflect the variables with the numerically greatest variance (Hewson, 2009).

Given the correlation matrix **R**, the vectors $\underline{e}_1, \underline{e}_2, \underline{e}_3, \ldots\ldots \underline{e}_k$ represent the eigen vectors corresponding to the eigen values $\lambda_1, \lambda_2, \lambda_3, \ldots. , \lambda_k$

$$\text{total standardised sample variance} = \sum_{i=1}^{k} r_{ii} = r_{11} + r_{22} + r_{33} + \cdots + r_{kk}$$

$$= \sum_{i=1}^{k} \lambda_i = trace\,(\mathbf{R}) = k \qquad \text{for } i = 1, 2, 3, \ldots, k$$

where r_{ij} is the sample correlation co-efficient between x_i and x_j for $i \neq j$

$$\text{and} \quad \mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \ldots. & r_{1k} \\ r_{21} & r_{22} & \ldots. & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \ldots & r_{kk} \end{bmatrix}$$

$$\text{and} \quad cor(x_i\,,\, x_j) = r_{ij} = e_{ij}\sqrt{\lambda_i} \qquad \text{for} \quad i \neq j$$

Also $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots. \geq \lambda_k \geq 0$ since **R** is positive definite and of full rank.

If $\underline{z}_1, \underline{z}_2, \underline{z}_3, \ldots\ldots \underline{z}_k$ are the standardised observations with correlation matrix **R** of size k X k and the corresponding eigenvalue–eigenvector pairs are

$$(\lambda_1, \underline{e}_1),\ (\lambda_2, \underline{e}_2),\ (\lambda_3, \underline{e}_3), \ldots. , (\lambda_k, \underline{e}_k),$$

then the i th sample principal component is given by

$$z_i = \underline{e}_i^{\,T}\, \underline{x} = e_{i1}x_1 + e_{i2}x_2 + e_{i3}x_3 + \ldots + e_{ik}x_k \quad \ldots\ldots(3.1)$$

$$\text{Proportion of the standardized sample variance due to the } i \text{ th principal component} = \frac{\lambda_i}{k} \qquad \text{for } i = 1, 2, 3, \ldots, k$$

$$cov(z_i\,,\, z_j) = 0 \qquad \text{for} \quad i \neq j$$

In performing this analysis, we take the unbiased estimator of the covariance matrix as

$$\hat{\Sigma} = S = \frac{1}{n-1}\sum_{i=1}^{k}\left(\underline{x}_i - \underline{\bar{x}}\right)^T\left(\underline{x}_i - \underline{\bar{x}}\right)$$

where $\left(\underline{x}_i - \underline{\bar{x}}\right)$ is a matrix of the mean centred data.

3.6.5 Spectral Decomposition and Principal Components

Given a symmetric square matrix **A** of size k X k,

We have $A = C\,D\,C^{\mathrm{T}}$ where $D = diag(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_k)$ and C is the matrix of the normalised vectors.

Also $D = C^{\mathrm{T}}\,A\,C$

$$A^{\frac{1}{2}} = C\,D^{\frac{1}{2}}\,C^{\mathrm{T}} \qquad D^{\frac{1}{2}} = diag\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3}, \dots, \sqrt{\lambda_k}\right)$$

$$A^{-1} = C\,D^{-1}\,C^{\mathrm{T}} \qquad D^{-1} = diag\left(\lambda_1^{-1}, \lambda_2^{-1}, \lambda_3^{-1}, \dots, \lambda_k^{-1}\right)$$

$$A^{2} = C\,D^{2}\,C^{\mathrm{T}} \qquad D^{2} = diag\left({\lambda_1}^2, {\lambda_2}^2, {\lambda_3}^2, \dots, {\lambda_k}^2\right)$$

Given the random vector $\underline{x}$ and the standardised random vector $\underline{z}$, we define $\underline{z}$ as

$$z_i = \frac{x_i - \mu}{\sigma_{ii}} \qquad \text{where } \underline{x} \sim N\left(\underline{\mu}, \Sigma\right)$$

If the diagonal of the variance matrix and corresponding standard deviation matrix is given by

$$V = \begin{bmatrix} S_{11}^2 & 0 & \dots. & 0 \\ 0 & S_{22}^2 & \dots. & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & S_{kk}^2 \end{bmatrix} \quad \text{and} \quad V^{\frac{1}{2}} = \begin{bmatrix} S_{11} & 0 & \dots. & 0 \\ 0 & S_{22} & \dots. & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & S_{kk} \end{bmatrix}$$

$\Sigma = V^{\frac{1}{2}}\,R\,V^{\frac{1}{2}}$ where $\mathbf{R}$ is the correlation matrix of the original variables x.

Therefore $\boldsymbol{R} = \left(V^{\frac{1}{2}}\right)^{-1}\,\boldsymbol{\Sigma}\,\left(V^{\frac{1}{2}}\right)^{-1}$

and $$E\left[\underline{z}\right] = E\left[\left(V^{\frac{1}{2}}\right)^{-1}\left(\underline{x}-\underline{\mu}\right)\right]$$

$$= \left(V^{\frac{1}{2}}\right)^{-1} E\left(\underline{x}-\underline{\mu}\right)$$

$$= 0 \qquad \text{since } E\left(\underline{x}-\underline{\mu}\right) = E\left[\underline{x}\right] - \underline{\mu} \quad \text{but} \quad E\left[\underline{x}\right] = \underline{\mu}$$

$$cor\left[\underline{z}\right] = E\left[\left(V^{\frac{1}{2}}\right)^{-1}\left(\underline{x}-\underline{\mu}\right)\left(\underline{x}-\underline{\mu}\right)^{T}\left(V^{\frac{1}{2}}\right)^{-1}\right]$$

$$= \left(V^{\frac{1}{2}}\right)^{-1} E\left[\left(\underline{x}-\underline{\mu}\right)\left(\underline{x}-\underline{\mu}\right)^{T}\right]\left(V^{\frac{1}{2}}\right)^{-1}$$

$$= \left(V^{\frac{1}{2}}\right)^{-1} \Sigma \left(V^{\frac{1}{2}}\right)^{-1}$$

$$= R$$

3.6.6 The Geometry of Principal Components

There are two types of rotation in principal component analysis: orthogonal and oblique rotation.

- Orthogonal rotation is used in situations when the underlying factors are assumed independent or uncorrelated. The factor loading is therefore taken as the Pearson's correlation co-efficient between the variable and the factor. The same values are the regression co-efficients of the factors. The values of the regression co-efficients and correlation co-efficients are therefore equal in such situations.
- Oblique rotation is used in situations when the underlying factors are assumed related or correlated. The Pearson's correlation co-efficient between the variable and the factor is not the same as the regression co-efficient of the factor. The values of the regression co-efficients and correlation co-efficients are therefore not equal in such situations.

3.6.7 Communality in PCA

The total variance of a variable in a factor or principal component has two components:

- Common variance which is shared with other variables or measures in that factor.
- Unique variance which is specific to that variable.

Communality is the proportion of common variance present in a variable. Therefore a variable that has no specific variance would have a communality of 1, while a variable that shares none of its variance would have a communality of 0. Principal component analysis assumes that all the variance within a factor is common and therefore the communality of every variable in a factor is 1 (Field, 2009). He further points out that because some information is discarded through discarding some factors, the factors retained will therefore not explain all of the variance in the data, and the communalities after extraction will always be less than 1. The closer the communalities are to 1, the better the factors are at explaining the original data. The more factors are retained, the greater the communalities will be and vice versa.

Communality is computed using the squared multiple correlation (SMC) of each variable with all others. This is basically the co-efficient of determination R^2 between the specific variable and the factor. Communality is therefore a measure of the proportion of the variance explained by the extracted factors (Field, 2009).

The factors retained after extraction will not map perfectly onto the original variables because they merely represent the common variance in the data.

3.6.8 Properties of Principal Components

(a) If $\underline{x}$, is a mean centred random vector, with covariance matrix $\mathbf{\Sigma}$, the principal components are given by:

$$\underline{x} \rightarrow \underline{z} = \mathbf{E}^T \left(\underline{x} - \underline{\mu}\right)$$

The transformation requires mean-centred values. $\mathbf{E}$ is orthogonal, $E^T \Sigma\, E$ is diagonal and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots . \geq \lambda_k \geq 0$ provided $\mathbf{\Sigma}$ is positive definite.

(b) $E(z_i) = 0$ and $var(z_i) = \lambda_i$

Therefore $var(z_1) \geq var(z_2) \geq var(z_3) \geq \ldots . \geq var(z_k)$

and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots . \geq \lambda_k \geq 0$.

No standard linear combination of $\underline{x}$ has a larger variance than λ.

Because of orthogonal projections, $cov\left(z_i\, ,\, z_j\right) = 0$ for $i \neq j$

This implies that the principal components are uncorrelated and are derived in decreasing order of variance.

(c) The trace of the variance-covariance matrix $\mathbf{\Sigma}$, is equal to the sum of its eigen values.

$$\sum_{i=1}^{k} \operatorname{var}(y_i) = trace\,(\mathbf{\Sigma}) = \sum_{i=1}^{k} \lambda_i$$

The trace of the correlation matrix $\mathbf{R}$, is equal to the number of variables.

$$trace\,(\mathrm{R}) = \sum_{i=1}^{k} \lambda_i = k$$

Therefore $\underline{x} \sim N_k\left(\underline{\mu}\,,\ \Sigma\right)$, $x_i \sim N\,(\mu_i\,,\ \sigma^2)$ and $z_i \sim N\,(0\,,\ \lambda_i)$

(d) The generalised variance which is the determinant of the covariance matrix Σ, can be expressed as the product of its eigen values.

$$|\Sigma| = \prod_{i=i}^{k} \lambda_i$$

The generalized variance and the sum of variances are unchanged by the principal component transformation.

(e) The first q principal components have a smaller mean squared deviation from the population or sample variables than any other k dimensional subspace. As a dimension reduction technique, the principal component analysis does not require any distributional assumptions on the variables.

3.6.9 Proportion of Total Variance Explained

The proportion of total variance explained by each linear combination of the i th principal component is given by

$\dfrac{\lambda_i}{\sum \lambda_i}$ for variance-covariance matrix $\mathbf{S}$ and $\dfrac{\lambda_i}{k}$ for correlation matrix $\mathbf{R}$.

The proportion of total variance explained by the first q principal components is given by

$\sum_{i=1}^{q} \lambda_i \Big/ \sum_{i=1}^{k} \lambda_i$ for variance-covariance matrix $\mathbf{S}$ and

$\dfrac{1}{k}\sum_{i=1}^{q} \lambda_i$ for correlation matrix $\mathbf{R}$.

3.6.10 Determining the Principal Components to be Retained

By default, there are as many factors or principal components as there are variables in a data set. However, a decision has to be made using established criteria on the extraction and retention of components which are statistically important. Each principal component has an associated eigen value and a corresponding eigen vector. Since the eigen value associated with a component indicates the substantive importance of that component, and by extension the corresponding eigen vector, the general criterion is to retain only the components with large eigen values. This is simplified by the fact that eigen values are generated in descending order of size i.e.

$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots. \geq \lambda_k \geq 0.$ So the first of the required number is retained.

But how large enough should an eigen value be in order to represent a meaningful component which should be retained? The following are some of the established criteria for determining the number of principal components to be retained.

(a) **Cattell Scree Plot**

In this technique proposed by Cattell in 1966, a graph is plotted with the eigen values along the vertical axis and the corresponding principal components along the horizontal axis. The graph is called a scree plot. In a typical case, few components will have high eigen values but most components will have relatively low eigen values. The graph is negatively sloped with sharp descent in the curve for the first few components followed by a tailing off for the remaining components. The cut-off point is at the point of inflexion in the curve i.e. the elbow of the curve. The components to the left of the point of inflexion are retained.

(b) **Kaiser's Criterion**

Kaiser's criterion of 1960 recommends the retention of all components with eigen values greater than 1.0. This criterion is based on the idea that the eigen values represent the amount of variation explained by a component and that an eigen value of 1.0 represents a substantial amount of variation (Field, 2009). Therefore all the components whose eigen values satisfy the criterion $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots. \geq \lambda_q \geq 1.0$ for $q < k$ are retained. It is further supported by the idea that the mean of all the eigen values is 1.0 as illustrated in 3.6.8 part (c).

Kaiser's criterion is viewed as accurate when the number of variables in the data set is less than 30 and the resulting communalities after extraction are all greater than 0.7. It is also viewed as accurate when the sample size exceeds 250 and the average communality is greater than or equal to 0.6.

(c) **Jolliffe's Criterion**

According to Jolliffe (1986) in (Field, 2009), Kaiser's criterion is too strict, and he therefore recommends the option of retaining all components with eigen values greater than 0.7. This criterion is based on the idea that the eigen values represent the amount of variation explained by a component and that an eigen value of 1.0 represents a substantial amount of variation (Field, 2009). Therefore all the components whose eigen values satisfy the criterion $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq .... \geq \lambda_q \geq 0.7$ for $q < k$ are retained.

According to Jolliffe (1986) in (Hewson, 2009), Kaiser's criterion involves using the *broken stick* approach. This approached involves computing the expected values of the proportion of explained variation, and then plotting the expected values alongside the corresponding actual values of the proportion of explained variation. If the expected value of this proportion is greater than the actual value, then the corresponding component is extracted. This is because such components account for more variation than would be expected purely by chance.

The expected value of the proportion of explained variation is given by the length

$$l_i = \frac{1}{k}\sum_{i=q}^{k}\frac{1}{i}$$

where l_i is the length of variation of the i th longest component and q is the number of extracted components.

(d) **Proportion of Variance Explained Criterion**

Some statisticians propose the rule of retaining just enough components to explain some specified large percentage of total variation of the original variables - proportions between 70% and 90% are usually used, but 75% is most popular. Therefore all the components whose eigen values satisfy the criterion

$\lambda_1 + \lambda_2 + \lambda_3 + ... + \lambda_q \approx 0.75$ for $q < k$ are retained.

3.7 Ordinal Logistic Regression

3.7.1 Derivation of the Logistic Regression

The simplest way to model data with one response variable Y and one explanatory or predictor variable X is to make the response variable a linear function of all the explanatory variables and express it in the form:

$$Y = \beta_0 + \beta_1 X$$

However, in most situations, there are multiple explanatory variables and the model is therefore expanded as a linear function of all the explanatory variables to become:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_k X_k$$

In this model, the coefficients are estimated using the ordinary least squares (OLS) technique which maximises the sum of squared deviations of the actual values of the response variable from their corresponding values along the regression line.

In this research, the response variable Y is categorical and therefore the interest is in the probability of Y, and the model therefore becomes a linear probability model given by:

$$P(Y) = p = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_k X_k$$

$$= \beta_0 + \sum \beta_i X_i$$

where X_i is the i th covariate while β_i is the regression co-efficient of X_i and β_0 is the regression intercept of the model.

However, probability of Y can only lie in the range 0 to 1 while the expression on the RHS can take any value in the range zero to infinity.

To solve the problem of this inconsistency, a restriction must be imposed on the co-efficients.

First step is to transform the response from a probability into the odds which is actually the proportion of the probability of success to the probability of failure in a binomial case.

This is given by:

$$odds = \frac{p}{1-p} = \frac{P(event)}{1-P(event)}$$

When the odds is greater than 1, then the event of interest is more likely to occur, but when the odds is less than 1, then the event of interest is less likely to occur.

The proportion of the odds of one event to another event is called the odds ratio. The odds are less restricted and can take any value in the range zero to infinity.

Next step is to take the natural logarithm of the odds to compute the logit or log odds. This is given by:

$$logit(p) \; = \; log\left(\frac{p}{1-p}\right)$$

Through this transformation, the lower limit of zero is removed and the function now has unlimited range i.e. $-\infty \; < \; logit(p) \; < \; \infty$

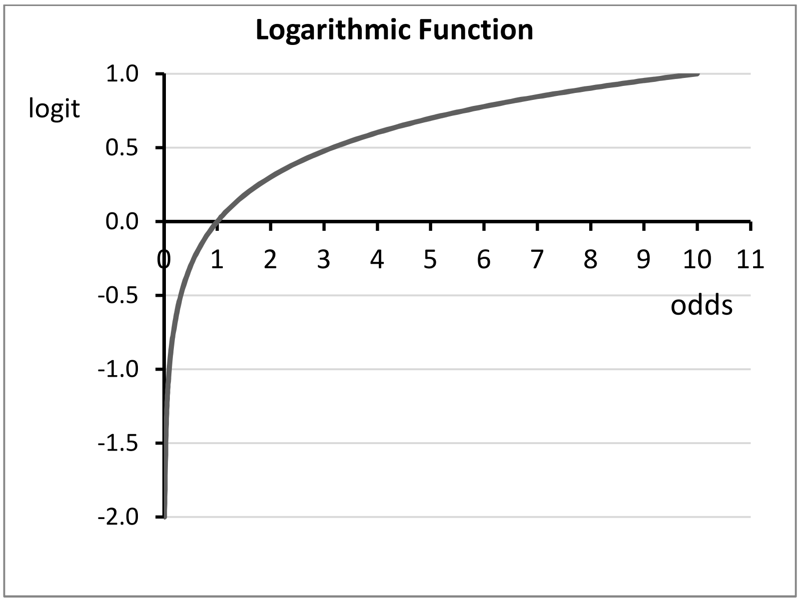

Figure 3.1: Logarithmic transformation of the odds

From the hypothetical graph and the formula, it is observed that:

- as probability $p \to 0$, the odds $\to 0$ and the logit $\to -\infty$
- as probability $p \to 1$, the odds $\to \infty$ and the logit $\to \infty$.

Suppose Y_i is a response variable with only two possible outcomes success and failure with probabilities p and 1 - p respectively, then Y_i has a binomial distribution defined by:

$$Y_i \sim binom(n_i \,, p_i).$$

By taking the logit of the odds as a linear function of explanatory variables gives:

$$logit(p) = log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_k X_k$$

$$\left(\frac{p}{1-p}\right) = exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots . + \beta_k X_k)$$

$$\left(\frac{p}{1-p}\right) = exp(\beta_0 + \sum \beta_i X_i)$$

$$p = (1-p)\, exp(\beta_0 + \sum \beta_i X_i)$$

$$p = [exp(\beta_0 + \sum \beta_i X_i) - p\, exp(\beta_0 + \sum \beta_i X_i)]$$

$$p\,[1 + exp(\beta_0 + \sum \beta_i X_i)] = exp(\beta_0 + \sum \beta_i X_i)$$

$$p = \frac{exp(\beta_0 + \sum \beta_i X_i)}{1 + exp(\beta_0 + \sum \beta_i X_i)}$$

$$p_j = \frac{exp(\beta_{0j} + \sum \beta_{ij} X_{ij})}{1 + exp(\beta_{0j} + \sum \beta_{ij} X_{ij})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.2)$$

where p_j is the probability of *j* th category in a single trial in the binomial distribution. Success is one of the values of the categorical variable i.e. Y_j.

Each regression co-efficient β_i defines the size of the contribution of the explanatory variable X_i to the response variable, with a larger value implying a stronger influence on the probability of the response variable, and vice versa. Also a positive value of the regression co-efficient β_i means that its corresponding predictor variable X_i increases the probability of the response variable, while a negative value of the regression co-efficient β_i means that its corresponding predictor variable X_i decreases the probability of the response variable. However, in considering any one variable X_i, it assumes that all the other variables $X_i s$ are held constant.

This approach extends further to include situations in which the response variable Y has more than two possible outcomes. This forms the basis of the logistic regression model upon which other models such as the multinomial logistic regression, ordinal logistic regression and ordinal probit regression models are based. In the case of multiple ordered outcomes, ordinal regression computes the probabilities of a set of outcomes in a cumulative approach.

3.7.2 The Ordinal Logistic Regression Model

The response variable Y has three levels categorised and measured on a three-point Likert scale as high, low and none. According to Stevens (1946), ordinal measurements describe order, but not relative size or degree of difference between the items measured and the numbers assigned to objects or events represent the rank order of the entities assessed. He defines a Likert scale as a type of ordinal scale and may also use names with an order. Therefore taking the three levels of Y as ordinal since they are actually hierarchical; the levels are ordered as shown below:

$$high > low > none$$

Assuming a proportional odds model, with the level **none** taken as the reference category, and using a set of explanatory variables $X_1, X_2, X_3, \ldots X_k$, two ordinal logistic regression models are fitted simultaneously onto the data.

The two models are:

$$ln\left[\frac{P(Y = none)}{P(Y = low,\ high)}\right] = \beta_{01} + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_k X_k \text{ ....(3.3)}$$

$$ln\left[\frac{P(Y = none,\ low)}{P(Y = high)}\right] = \beta_{02} + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_k X_k \text{ ....(3.4)}$$

Odds Ratio for variable X_i

$$OR = \frac{odds\ at\ X_i = 1}{odds\ at\ X_i = 0} = e^{\beta_i}$$

The odds of an event occurring is defined as the probability of an event occurring divided by the probability of that event not occurring (Field, 2009).

The individual employee factors and the institutional factors were analysed separately.

where x_i is assigned the value 0 for the reference category and assigned the value 1 for the other category if it represents a categorical variable. The category in italics represents the reference category.

x_i is assigned its actual value captured in the questionnaire if it represents a numeric variable.

k_1 = the reduced number of individual employee variables

k_2 = the reduced number of institutional variables

3.7.3 Assumptions in the Ordinal Logistic Regression Model

1. Measurement scales: The response variable is classified as ordinal among the measurement scales developed by Stevens (1946). However, an individual explanatory variable is either categorical or numerical.
2. Linearity: There is a linear relationship between each predictor variable and the logarithm (logit) of the response variable.
3. Independence of errors: Cases or data for observational units should not be related. Same data is not collected from same respondents at different times.
4. No perfect multi-collinearity: Each observation is independent. No perfect linear relationship exists between multiple predictors. That the explanatory variables should be independent from each other.
5. Proportional odds. The assumption of proportional odds means that each explanatory variable has an identical effect at each cumulative split of the ordinal dependent variable. This therefore implies that the:
   - Intercept β_{0i} depends on the category level (response variable).
   - Predictor variable X_i does not depend on the category level.
   - Co-efficient β_i does not depend on the category level.

The approach in the ordinal regression used by this research is the cumulative odds ordinal logistic regression with proportional odds, which uses cumulative categories.

Based on the above assumptions, the overall odds of any event may differ, but the effect of the predictors on the odds of an event occurring in every subsequent response category (level) is the same for every category. Therefore the intercept differs between models, but the regression co-efficients are the same across the fitted models.

3.7.4 Model Fitting and Selection

The model should be fitted correctly without over-fitting or under-fitting. That is only the meaningful variables should be included, but also all meaningful variables should be included. A good approach to ensure this is to use a stepwise method to estimate the logistic regression (Emerald, 2014).

The explanatory variables were analysed separately in two groups - individual and institutional factors. Within each group, *backward stepwise method* is used during analysis. According to Field (2009), this method first involves having the model with all the predictors included, then using the log-likelihood ratio statistic, the model is reduced step by step by having the predictors which do not have a substantial effect on how well the model fits the observed data removed, starting with the predictor with the least impact – largest p-value. For a variable to be removed its p-value must be larger than 0.05. Each variable is examined to evaluate if removing it from the model improves the model fit. The final model in the end is a reduced model of the original full model. The backward stepwise method is preferred over the forward stepwise method because of the suppressor effects which occur when a predictor has a significant effect, but only when another variable is held constant (Field, 2009).

The log-likelihood ratio test statistic is given by

$$\mathrm{LR} = -2\ln\left(\frac{L_R}{L_F}\right) = -2\left(\ln L_R - \ln L_F\right) = 2\left(\ln L_F - \ln L_R\right) \quad \text{and}$$

where $0 \leq \left(\frac{L_R}{L_F}\right) \leq 1$

The log-likelihood ratio statistic approximately follows a chi-square distribution, hence

$\mathrm{LR} \sim \mathcal{X}^2{}_q$ where q = no. of restrictions

L_R – Log-likelihood of restricted model (with intercept only)

L_F – Log-likelihood of full model (with all explanatory variables)

In checking the contribution of an additional variable X_i in the model, its regression co-efficient β_i is evaluated.

To achieve this, the research formulates and tests the hypothesis

H_0: $\beta_i = 0$, vs H_1: $\beta_i \neq 0$,

for $i = 1, 2, 3, \dots, k$ excluding the intercept β_0 in the model

If $\mathcal{X}^2_{\text{calc}} < \mathcal{X}^2_{0.05,\, q}$ then H_0 is not rejected and X_i is excluded in the model.

3.7.5 Probabilities of the ICT Integration Levels

Probabilities of a staff belonging to the various categories are as outlined below:

Probability that a staff member belongs to the category of **none** integration

$$P(Y = \text{none}) = \frac{exp(\beta_{01}+\beta_1X_1+\beta_2X_2+\beta_3X_3+\dots+\beta_kX_k)}{1+exp(\beta_{01}+\beta_1X_1+\beta_2X_2+\beta_3X_3+\dots+\beta_kX_k)}$$

Probability that a staff member belongs to the category of **low** integration

$$P(Y = \text{low}) = \frac{exp(\beta_{02}+\beta_1X_1+\beta_2X_2+\dots+\beta_kX_k)}{1+exp(\beta_{02}+\beta_1X_1+\beta_2X_2+\dots+\beta_kX_k)} - \frac{exp(\beta_{01}+\beta_1X_1+\beta_2X_2+\dots+\beta_kX_k)}{1+exp(\beta_{01}+\beta_1X_1+\beta_2X_2+\dots+\beta_kX_k)}$$

Probability that a staff member belongs to the category of **high** integration

P(Y = high) = 1 – {P(Y = none) + P(Y = low)}

$$= 1 - \left\{\frac{exp(\beta_{02}+\beta_1X_1+\beta_2X_2+\beta_3X_3+\dots+\beta_kX_k)}{1+exp(\beta_{02}+\beta_1X_1+\beta_2X_2+\beta_3X_3+\dots+\beta_kX_k)}\right\} \quad \text{.................................(3.5)}$$

3.7.6 Interpretation and Statistical Inference

3.7.6.1 Interpretation of the Odds ratio Exp(β_i)

Taking an explanatory variable X_i and the corresponding odds ratio OR = e^{β_i}. Then this means that holding other explanatory variables constant:

- the odds of that factor falling in the *none* category as opposed to the *low* or *high* categories is OR. This value of OR implies that the explanatory variable category is 100(OR – 1)% more likely or is 100(1 – OR)% less likely to fall in the *low* or *high* integration category.

- the odds of that factor falling in the *none* or *low* categories as opposed to the *high* category is OR. This value of OR implies that the explanatory variable category is 100(OR – 1)% more likely or is 100(1 – OR)% less likely to fall in the *high* integration category.

If $\beta_i > 0$, then OR = $e^{\beta_i} > 1$

Therefore the other category of interest is 100(OR – 1)% more likely to have the characteristic of interest in X_i than the reference category. This also means that X_i increases the odds that Y_i equals a specified value by a factor equal to the odds ratio.

If $\beta_i < 0$, then OR = $e^{\beta_i} < 1$

Therefore the other category of interest is 100(1 – OR)% less likely to have the characteristic of interest in X_i than the reference category. This also means that X_i decreases the odds that Y_i equals a specified value by a factor equal to the odds ratio.

If $\beta_i = 0$, then OR = $e^{\beta_i} = 1$

Therefore the other category of interest and the reference category are equally likely and therefore the characteristic of interest in X_i does not influence the response variable. This also means that X_i does not change the odds that Y_i equals a specified value.

For continuous explanatory variables, such as age and duration of employment in years, and Internet bandwidth in megabytes per second (mbps), the research interpreted how much a unit increase or decrease in that variable, for example a one year increase or decrease in age, a one mbps increase or decrease in Internet bandwidth, is associated with the odds of ICT integration having a higher or lower value.

Also generally, the more the odds ratio deviates from 1, the stronger the relationship between the values of X_i and Y_i.

3.7.6.2 Statistical Inference on the Regression Co-efficient and the Odds ratio

To test for significance of the regression co-efficient β_i the research formulates the hypothesis

$$\mathrm{H_0}: \beta_i = 0, \quad \text{vs} \quad \mathrm{H_1}: \beta_i \neq 0,$$

The confidence limits for the co-efficient β_i is therefore given by

$$\beta_i = \hat{\beta}_i \pm Z_{\propto/2} S_{\hat{\beta}_i}, \quad \text{therefore}$$

95% confidence limits become $\beta_i = \hat{\beta}_i \pm 1.9600\, S_{\hat{\beta}_i}$

To test for the significance of the odds ratio e^{β_i}, the research formulates the hypothesis

$$\mathrm{H}_0:\ e^{\beta_i} = 1, \qquad \text{vs} \qquad \mathrm{H}_1:\ e^{\beta_i} \neq 1,$$

The confidence limits for the odds ratio e^{β_i} is therefore given by

$$e^{\beta_i} = e^{\hat{\beta}_i \pm Z_{\propto/2} S_{\hat{\beta}_i}}, \qquad \text{therefore}$$

95% confidence limits become $\quad e^{\beta_i} = e^{\hat{\beta}_i \pm 1.9600\, S_{\hat{\beta}_i}}$

Where **z** is the Wald's test statistic given by

$$z = \frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}} \qquad \text{and} \qquad z \sim N(0, 1)$$

Also $\quad \mathrm{z}^2 = \left(\frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}\right)^2 \quad$ and $\quad \mathrm{z}^2 \sim \mathcal{X}^2_{(1)}$

If the confidence interval for the odds ratio e^{β_i} includes 1, then the corresponding explanatory variable X_i does not make a statistically significant contribution to the determination of the response variable Y and therefore H_0: e^{β_i} = 1 is not rejected, otherwise variable X_i makes a significant contribution in determining Y and therefore H_0: e^{β_i} = 1 is rejected on the basis of the evidence provided by the data.

3.7.7 Checking Adequacy of the Model

In checking how well the model fits the observed data, the actual values of the response variable are compared with the corresponding predicted values. This is to determine how well the values predicted by the model compare with those of the actual.

The deviance value was used to test for goodness of fit for the model.

Deviance = Null deviance – Residual deviance

= Chi-square calculated ($\mathcal{X}^2_{\text{calc}}$)

Chi-square critical = $\mathcal{X}^2_{\text{critical}} = \mathcal{X}^2_{0.05,\,\mathrm{V}}$ at α = 0.05 level of significance

with v degrees of freedom

If $\mathcal{X}^2_{\text{calc}} > \mathcal{X}^2_{\text{critical}}$ then the research concludes that the fit is significant.

In checking the goodness of the model as a measure of how well the model fits the observed data, the **Pseudo $\mathbf{R^2}$** was used. The following was computed:

(a) McFadden's Pseudo R^2 given by:

$$\text{Pseudo } R^2 = 1 - \frac{\ln \hat{L}\,(M_{full})}{\ln \hat{L}\,(M_{intercept})}$$

where $\ln \hat{L}\,(M_{full})$ – log-likelihood of full model of interest

$\ln \hat{L}\,(M_{intercept})$ – log-likelihood with all co-efficients except that of intercept set to zero.

(b) Nagelkerke's Pseudo R^2 is used. The Pseudo R^2 is given by:

$$R_N^2 = \frac{R_{CS}^2}{1 - exp\left\{\frac{2}{\mathrm{n}}(\mathrm{LL}_{intercept})\right\}}$$

where R_{CS}^2 – Cox and Snell's R^2 given by

$$R_{CS}^2 = 1 - exp\left(-\frac{2}{\mathrm{n}}\mathrm{LL}_{full} - \mathrm{LL}_{intercept}\right)$$

The model was accessed based primarily on the Nagelkerke's Pseudo R^2.

Pseudo $\mathbf{R^2}$ is interpreted as follows:

$R^2 \rightarrow \pm 1.0 \Rightarrow$ a better fit of the model on the data

$R^2 \rightarrow \pm 0.5 \Rightarrow$ a moderate fit of the model on the data

$R^2 \rightarrow \pm 0.0 \Rightarrow$ a poorer fit of the model on the data

3.8 Data Source and Data Management

The study uses data collected from primary sources through administering questionnaires to various respondents in the form of prospective computer users who are staff in learning institutions at tertiary level. The questionnaires were administered to individuals to capture individual attributes and a special questionnaire administered to an ICT specialist serving as the head of the ICT support department or section to collect information of a technical but general nature about the institution. Data verification and cleaning was done manually by perusing all the responses in all the questionnaires, and the data keyed in a template designed in IBM SPSS Statistics.

Data was collected from five institutions in Machakos area with a total response of 157 respondents against a planned sample size of 207 yielding a response rate of 75.8%. This excludes 5 cases which were discarded because the questionnaires were incomplete and missing several crucial variables.

The data was captured using IBM SPSS Statistics software. The data items which the research initially perceived as variables but were found to be constants were removed from the list of variables and excluded from analysis. A constant in this context is a variable whose value is the same across all respondents (cases). The data set used by this research is predominantly categorical with only a few numerical variables. To facilitate analysis, the categorical variables were coded using general discrete form for nominal variables, and using a suitable Likert scale for ordinal variables. The analysis was predominantly performed in IBM SPSS and R software, but supplemented by using Microsoft Excel.

CHAPTER 4: RESULTS AND DISCUSSIONS

4.1 Clustering of Variables Using PCA

4.1.1 Extraction of Principal Components for Individual Factors

4.1.1.1 Correlation Matrix for Individual Staff Variables

Correlation matrix **R** of the variables pertaining to individual staff

	Gender	Age of staff	Staff category	Employment duration	Education level	Age at first use	Management Level	Freq. of computer use	Place ICT skills acquired	Mode of ICT training	E-mail on Internet	Browsing on Internet	Social media use	Job satisfaction level	Type of mobile phone	ICT compulsory in college
Gender	1.0000	0.1316	0.0616	0.1218	0.0120	0.2272	0.0729	-0.0790	-0.0404	0.0871	-0.0833	-0.1713	-0.0902	-0.0635	0.0573	-0.0039
Age of staff	0.1316	1.0000	0.4206	0.8592	0.2599	0.8081	0.2877	-0.1878	0.2553	-0.1351	0.0996	-0.0801	-0.1712	-0.0585	-0.0620	-0.1147
Staff category	0.0616	0.4206	1.0000	0.3692	0.4901	0.2952	0.1643	-0.1665	0.1224	-0.1271	0.1034	0.0707	-0.0628	-0.0275	0.0225	0.0646
Employment duration	0.1218	0.8592	0.3692	1.0000	0.2945	0.7107	0.2931	-0.2006	0.2659	-0.2126	0.1278	-0.0498	-0.1749	0.0593	-0.0597	-0.1717
Education level	0.0120	0.2599	0.4901	0.2945	1.0000	0.1353	0.3377	0.0182	-0.1087	-0.1234	0.3075	0.1862	-0.0194	0.0839	0.0660	0.0449
Age at first use	0.2272	0.8081	0.2952	0.7107	0.1353	1.0000	0.1831	-0.2168	0.2995	-0.1162	0.0042	-0.2087	-0.1632	-0.0823	-0.1102	-0.2270
Management Level	0.0729	0.2877	0.1643	0.2931	0.3377	0.1831	1.0000	-0.0086	0.0641	0.0497	0.0927	0.1724	-0.1308	0.0841	0.0172	-0.0405
Freq. of computer use	-0.0790	-0.1878	-0.1665	-0.2006	0.0182	-0.2168	-0.0086	1.0000	-0.1264	0.1250	0.1051	0.1298	0.1103	0.1273	0.2077	0.1854
Place ICT skills acquired	-0.0404	0.2553	0.1224	0.2659	-0.1087	0.2995	0.0641	-0.1264	1.0000	0.1344	-0.0293	0.0096	-0.0620	0.0042	-0.1955	-0.0462
Mode of ICT training	0.0871	-0.1351	-0.1271	-0.2126	-0.1234	-0.1162	0.0497	0.1250	0.1344	1.0000	-0.0429	0.0797	0.0959	-0.0651	0.0706	0.2035
E-mail on Internet	-0.0833	0.0996	0.1034	0.1278	0.3075	0.0042	0.0927	0.1051	-0.0293	-0.0429	1.0000	0.3818	0.0342	0.0135	-0.0174	0.2118
Browsing on Internet	-0.1713	-0.0801	0.0707	-0.0498	0.1862	-0.2087	0.1724	0.1298	0.0096	0.0797	0.3818	1.0000	0.0886	0.1345	0.1169	0.1365
Social media use	-0.0902	-0.1712	-0.0628	-0.1749	-0.0194	-0.1632	-0.1308	0.1103	-0.0620	0.0959	0.0342	0.0886	1.0000	0.0851	0.2325	0.1448
Job satisfaction level	-0.0635	-0.0585	-0.0275	0.0593	0.0839	-0.0823	0.0841	0.1273	0.0042	-0.0651	0.0135	0.1345	0.0851	1.0000	0.1007	0.1815
Type of mobile phone	0.0573	-0.0620	0.0225	-0.0597	0.0660	-0.1102	0.0172	0.2077	-0.1955	0.0706	-0.0174	0.1169	0.2325	0.1007	1.0000	0.1268
ICT compulsory in college	-0.0039	-0.1147	0.0646	-0.1717	0.0449	-0.2270	-0.0405	0.1854	-0.0462	0.2035	0.2118	0.1365	0.1448	0.1815	0.1268	1.0000

a. Determinant = 0.010

Table 4.1: Correlation matrix of individual staff variables

The determinant of the correlation matrix **R** in Table 4.1 is 0.010 which is greater than the recommended minimum of 0.00001 (Field, 2009). So the correlation matrix is significantly different from a singular matrix. Also the correlation coefficients have a wide range comprising both high values and low values. This renders the data highly suitable for principal component analysis.

4.1.1.2 Test for Sampling Adequacy and Sphericity

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.679
Bartlett's Test of Sphericity	Approx. Chi-Square	688.816
	df	120
	Sig.	0.000

The Kaiser-Meyer-Olkin measure of sampling adequacy has a value of 0.679 which is greater than the acceptable minimum value of 0.5. So the sample size was significantly adequate.

Since the Bartlett's test for sphericity shows a significant value, it means that the correlation matrix in Table 4.1 is significantly different from an identity matrix.

4.1.1.3 Communalities of Variables for Individual Staff

Communalities

	Initial	Extraction
Gender	1.000	0.681
Age of staff	1.000	0.889
Staff category	1.000	0.728
Employment duration	1.000	0.855
Education level	1.000	0.716
Age at first use	1.000	0.838
Management Level	1.000	0.760
Freq. of computer use	1.000	0.636
Place ICT skills acquired	1.000	0.734
Mode of ICT training	1.000	0.780
E-mail on Internet	1.000	0.777
Browsing on Internet	1.000	0.635
Social media use	1.000	0.703
Job satisfaction level	1.000	0.886
Type of mobile phone	1.000	0.698
ICT compulsory in college	1.000	0.788

Extraction Method: Principal Component Analysis.

Table 4.2: Communalities of variables for individual staff

From the communalities in Table 4.2, it is observed that all the variables have communalities with high values closer to the ideal value of 1.0. Field (2009) recommends an ideal minimum value of 0.6 for all communalities. From the data in the communalities table, this condition has been satisfied by all the variables analysed.

4.1.1.4 Eigen Values and Total Variance Explained

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.448	21.552	21.552	3.448	21.552	21.552	2.891	18.071	18.071
2	2.099	13.116	34.668	2.099	13.116	34.668	1.597	9.983	28.054
3	1.302	8.137	42.804	1.302	8.137	42.804	1.384	8.651	36.704
4	1.264	7.897	50.702	1.264	7.897	50.702	1.368	8.548	45.252
5	1.104	6.901	57.602	1.104	6.901	57.602	1.288	8.052	53.304
6	1.033	6.454	64.056	1.033	6.454	64.056	1.259	7.867	61.171
7	0.944	5.897	69.953	0.944	5.897	69.953	1.171	7.319	68.490
8	0.911	5.696	75.649	0.911	5.696	75.649	1.145	7.159	75.649
9	0.796	4.976	80.625						
10	0.706	4.413	85.038						
11	0.649	4.055	89.093						
12	0.551	3.441	92.534						
13	0.497	3.105	95.640						
14	0.351	2.193	97.833						
15	0.247	1.542	99.375						
16	0.100	0.625	100.000						

Extraction Method: Principal Component Analysis.

Table 4.3: Proportion of variance explained

4.1.1.5 Factor Extraction from Individual Staff Variables

A brief observation of the data analysis shows that if the component extraction is based on the Cattell scree plot, then using the point of inflexion as the criteria, only the first two factors will be extracted, accounting for only 34.668% of the total variation. This proportion is generally too low and hence the criterion was dropped.

Using the Kaiser's criterion of extracting principal components with eigen values greater than 1.0, implies that only the first 6 components will be extracted, accounting for only 64.056% of the total variation. This proportion is also not adequate and hence the criterion was dropped.

Using the Jolliffe's criterion of extracting components with eigen values greater than 0.7, implies that only the first 10 components will be extracted accounting for up to 85.038% of the total variation. This criterion results in the extraction of too many components. It was therefore dropped on this basis.

This research therefore used the general criterion where a maximum cumulative variance of 75% is applied resulting in the extraction of 8 components as shown in Table 4.3.

The eigen values are relatively small which this research attributes to the small range of values assigned to categorical variables on a scale of 0 to 5. However, the eigen values meet the criteria of being in descending order and all are positive. Therefore the correlation matrix **R** in Table 4.1 is positive definite.

The Cattell Scree Plot

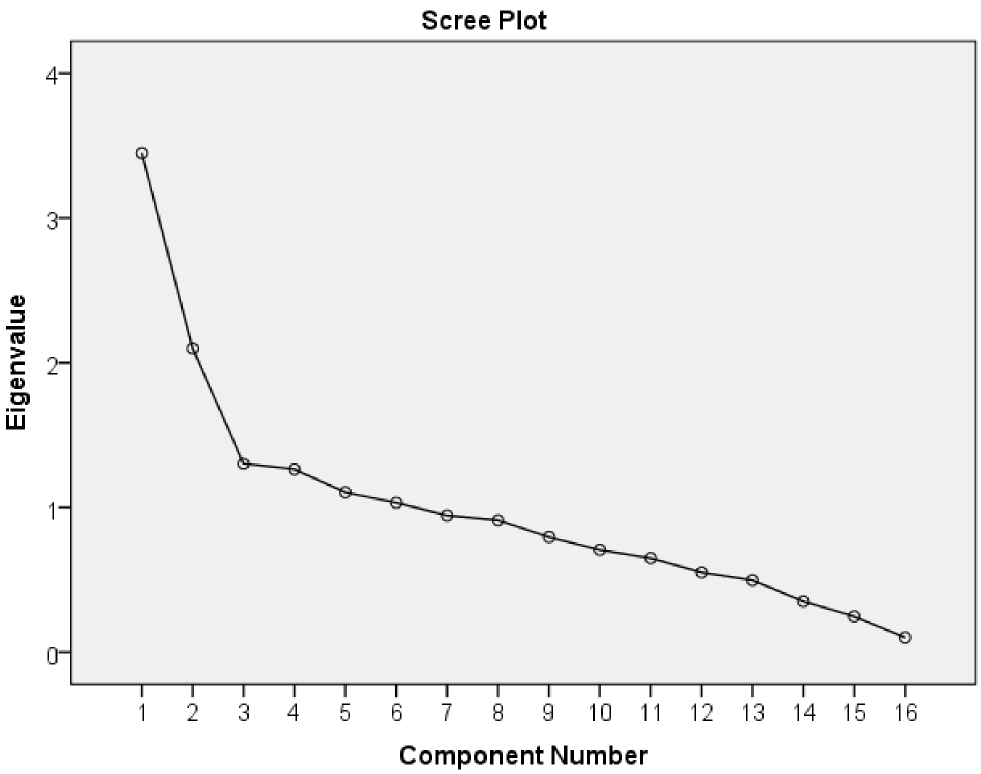

Figure 4.1: Scree plot of factors of individual staff

The Cattell scree plot in Figure 4.1 has a clear and well defined elbow or point of inflection and conforms well the general theory.

The correlation matrix of the principal components is an identity matrix, with all elements along the leading diagonal equal to 1.00 while the off-diagonal elements are equal to zero. This means that the principal components satisfy the two conditions of non-collinearity.

$$cor\left(z_i, z_j\right) = \begin{cases} 0 & \text{for} \quad i \neq j \\ 1 & \text{for} \quad i = j \end{cases}$$

4.1.1.6 Clustering of Individual Staff Variables into Factors

The clusters of the original variables forming principal components based on orthogonal rotation and sorting co-efficients are therefore shown in Table 4.4

Rotated Component Matrix - Unfiltered

Variable	Component							
	1	2	3	4	5	6	7	8
Age of staff	0.912	0.062	0.184	-0.033	0.104	-0.004	0.078	-0.033
Age at first use	0.891	-0.092	0.076	-0.072	0.007	0.025	0.128	-0.094
Employment duration	0.888	0.045	0.164	-0.058	0.151	-0.085	0.003	0.065
E-mail on Internet	0.103	0.861	0.045	-0.062	0.008	-0.069	-0.103	-0.067
Browsing on Internet	-0.159	0.535	0.045	0.117	0.382	0.144	-0.375	0.037
ICT compulsory in college	-0.223	0.480	0.171	0.097	-0.281	0.308	0.297	0.454
Staff category	0.296	0.127	0.778	0.065	0.096	-0.029	0.065	0.027
Freq. of computer use	-0.101	0.392	-0.570	0.298	0.099	-0.044	0.111	0.186
Education level	0.132	0.354	0.557	0.107	0.422	-0.257	0.063	0.062
Mobile phone type	-0.037	0.002	-0.073	0.796	0.150	-0.083	0.156	0.061
Social media use	-0.111	0.016	0.092	0.721	-0.273	0.151	-0.253	0.028
Management Level	0.189	0.031	0.100	-0.048	0.831	0.098	0.050	0.085
Mode of ICT training	-0.199	0.049	-0.122	0.137	0.150	0.787	0.224	-0.108
Place ICT skills acquired	0.409	-0.106	0.041	-0.217	-0.064	0.627	-0.320	0.088
Gender	0.145	-0.118	0.034	-0.005	0.050	0.070	0.796	-0.065
Job satisfaction level	-0.004	-0.039	-0.054	0.066	0.122	-0.073	-0.110	0.919

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 15 iterations.

Table 4.4: Clusters of variables for individual staff - highlighted

Figure 4.2 shows a graphical presentation of the results in Table 4.4 using a scatter plot with the first two principal components.

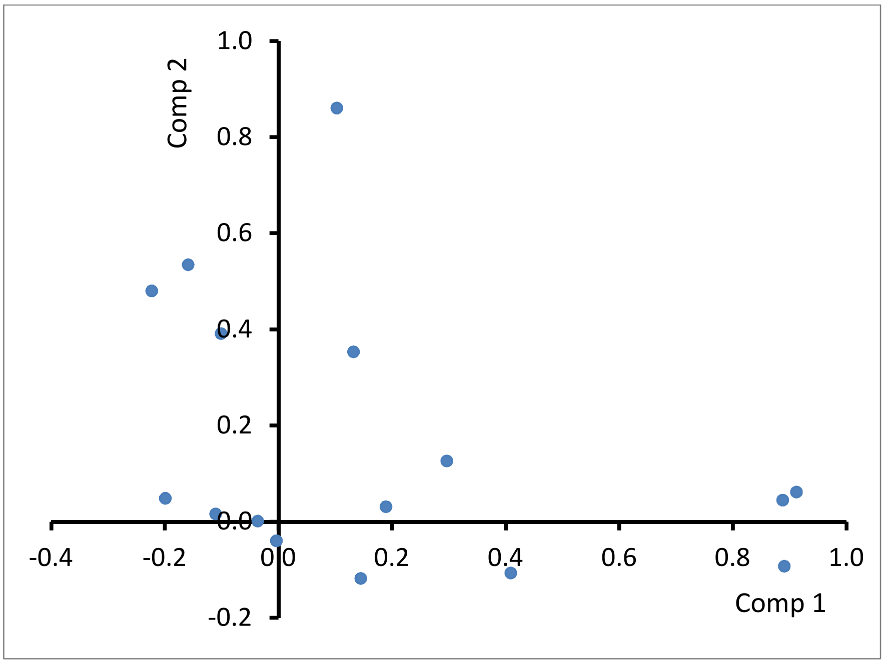

Figure 4.2: Scatter plot of rotated scores - individual

Rotated Component Matrix - Filtered

Variable	Component							
	1	2	3	4	5	6	7	8
Age of staff	0.912							
Age at first use	0.891							
Employment duration	0.888							
E-mail on Internet		0.861						
Browsing on Internet		0.535						
ICT compulsory in college								
Staff category			0.778					
Freq. of computer use			-0.570					
Education level			0.557					
Mobile phone type				0.796				
Social media use				0.721				
Management Level					0.831			
Mode of ICT training						0.787		
Place ICT skills acquired						0.627		
Gender							0.796	
Job satisfaction level								0.919

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 15 iterations.

b. Values less than 0.5 have been excluded.

Table 4.5: Clusters of variables for individual staff - filtered

An observation of Table 4.5 on rotated component matrix reveals a very clear grouping or clustering pattern of the variables. This is manifested by the values retained and the natural grouping formed after the values less than 0.5 have been excluded and the clusters are summarised the Table 4.6. This exclusion of values less than 0.5 is on the basis their corresponding variables are assumed as not significant in the linear models formed by the extracted principal components.

Extracted Components and Variables

	Component or Factor	**Variables forming the Cluster**
1	Demographic factors	• *Age of staff* • Age at first use • Employment duration
2	ICT literacy	• *e-mail on Internet* • Browsing on Internet
3	Employment factors	• *Staff category* • Frequency of computer use • Education level
4	Mobile phone usage	• *Type of mobile phone* • Social media use
5	Progress in employment	• *Management level*
6	ICT history	• *Mode of ICT training* • Place ICT skills acquired
7	Social factor	• *Gender*
8	Psychological factor	• *Job satisfaction level*

Table 4.6: Principal components for individual staff

To retain the analysis of the original data variables rather than the principal components, the variable with the highest loading in a cluster was selected to represent the group of variables comprising that component. The highest loading means that the variable has the highest correlation with the principal component for its cluster. The selected variable in a cluster is indicated in italics in Table 4.6.

4.1.1.7 Principal Components of Individual Staff Attributes

Component Score Coefficient Matrix

Variable	Component							
	1	2	3	4	5	6	7	8
Gender	0.005	-0.020	0.004	-0.022	0.024	0.049	0.675	-0.016
Age of staff	0.348	0.073	-0.056	0.068	-0.044	0.008	0.027	-0.017
Staff category	-0.039	-0.009	0.601	0.072	-0.046	0.034	0.038	0.016
Employment duration	0.338	0.042	-0.075	0.040	0.000	-0.056	-0.033	0.074
Education level	-0.082	0.117	0.366	0.052	0.236	-0.163	0.057	-0.002
Age at first use	0.357	0.004	-0.119	0.055	-0.090	0.025	0.063	-0.047
Management Level	-0.042	-0.117	-0.024	-0.041	0.695	0.107	0.015	0.041
Freq. of computer use	0.124	0.289	-0.543	0.150	0.070	-0.090	0.124	0.085
Place ICT skills acquired	0.164	-0.074	0.016	-0.104	-0.066	0.511	-0.299	0.109
Mode of ICT training	-0.076	0.009	-0.027	0.079	0.180	0.625	0.177	-0.137
E-mail on Internet	0.083	0.629	-0.098	-0.119	-0.144	-0.071	-0.025	-0.152
Browsing on Internet	-0.063	0.258	-0.010	0.036	0.278	0.124	-0.300	-0.079
Social media use	0.057	-0.069	0.151	0.587	-0.235	0.126	-0.259	-0.059
Job satisfaction level	0.023	-0.159	-0.045	-0.045	0.071	-0.069	-0.070	0.839
Type of mobile phone	0.065	-0.097	-0.069	0.619	0.125	-0.070	0.084	-0.039
ICT compulsory in college	-0.069	0.312	0.193	-0.068	-0.341	0.229	0.320	0.376

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 4.7: Component score co-efficient matrix for individual staff

The principal components are therefore derived from Table 4.7 and given by:

$z_1 = 0.005x_1 + 0.348x_2 - 0.039x_3 + 0.338x_4 - 0.082x_5 + \ldots + 0.065x_{15} - 0.069x_{16}$

$z_2 = -0.020x_1 + 0.073x_2 - 0.009x_3 + 0.042x_4 + 0.117x_5 + \ldots - 0.097x_{15} + 0.312x_{16}$

$z_3 = 0.004x_1 - 0.056x_2 + 0.601x_3 - 0.075x_4 + 0.366x_5 + \ldots - 0.069x_{15} + 0.193x_{16}$

$z_4 = -0.022x_1 + 0.068x_2 + 0.072x_3 + 0.040x_4 + 0.052x_5 + \ldots + 0.619x_{15} - 0.068x_{16}$

$z_5 = 0.024x_1 - 0.044x_2 - 0.046x_3 + 0.236x_5 - 0.090x_6 + \ldots + 0.125x_{15} - 0.341x_{16}$

$z_6 = 0.049x_1 + 0.008x_2 + 0.034x_3 - 0.056x_4 - 0.163x_5 + \ldots - 0.070x_{15} + 0.229x_{16}$

$z_7 = 0.675x_1 + 0.027x_2 + 0.038x_3 - 0.033x_4 + 0.057x_5 + \ldots + 0.084x_{15} + 0.320x_{16}$

$z_8 = -0.016x_1 - 0.017x_2 + 0.016x_3 + 0.074x_4 - 0.002x_5 + \ldots - 0.039x_{15} + 0.376x_{16}$

where the variable x_i is as follows:

i	x_i *variable*
1.	Gender
2.	Age of staff
3.	Staff category
4.	Employment duration
5.	Education level
6.	Age at first use
7.	Management Level
8.	Freq. of computer use
9.	Place ICT skills acquired
10.	Mode of ICT training
11.	E-mail on Internet
12.	Browsing on Internet
13.	Social media use
14.	Job satisfaction level
15.	Type of mobile phone
16.	ICT compulsory in college

4.1.2 Extraction of Principal Components for Institutional Factors

4.1.2.1 Correlation Matrix for Institutional Variables

Correlation matrix **R** of the variables pertaining to the institutions initially had very low values when all the variables were included rendering clustering of variables almost impossible, yet this is the essence of principal component analysis. Also the determinant of the matrix **R** had a value of 0.00 which is less than the ideal minimum value of 0.00001 recommended by Field (2009). This implies that the correlation matrix **R** is a singular matrix or is not significantly different from a singular matrix.

The correlation matrix was also initially found not to be positive definite. To proceed with the analysis of institutional factors, the variables which are common to all individuals in an institution were excluded from further analysis because of high level of multi-collinearity. Only those institutional variables which differ even between respondents in the same institution and are there able to discriminate the ICT integration among the individuals were analysed.

	ICT required at employment	Internet access at work	Software availability	Access to desktop PC	Access to printer	Usage of projector
ICT required at employment	1.000	0.088	0.012	0.039	0.171	-0.010
Internet access at work	0.088	1.000	0.228	0.263	0.181	0.135
Software availability	0.012	0.228	1.000	0.199	0.097	0.038
Access to desktop PC	0.039	0.263	0.199	1.000	0.144	0.118
Access to printer	0.171	0.181	0.097	0.144	1.000	0.437
Usage of projector	-0.010	0.135	0.038	0.118	0.437	1.000

Determinant = 0.655

Table 4.8: Correlation matrix of institutional variables

In Table 4.8, the determinant of the correlation matrix **R** is 0.655 which exceeds the recommended minimum of 0.00001 (Field, 2009). So the correlation matrix is significantly different from a singular matrix. The matrix is also positive definite.

The values of the correlation coefficients are generally low with all less than 0.5, which means that the variables are free from multi-collinearity.

4.1.2.2 Test for Sampling Adequacy and Sphericity

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.582
Bartlett's Test of Sphericity	Approx. Chi-Square	69.831
	df	15
	Sig.	0.000

The Kaiser-Meyer-Olkin measure of sampling adequacy has a value of 0.582 which is greater than the acceptable minimum value of 0.5 (Field, 2009).

Since the Bartett's test for sphericity shows a significant value, it means that the correlation matrix above is significantly different from an identity matrix.

4.1.2.3 Communalities of Institutional Variables

Communalities

Variable	Initial	Extraction
ICT required at employment	1.000	0.953
Internet access at work	1.000	0.533
Software availability	1.000	0.950
Access to desktop PC	1.000	0.804
Access to printer	1.000	0.733
Usage of projector	1.000	0.779

Extraction Method: Principal Component Analysis.

Table 4.9: Communalities of institutional variables

From the communalities in Table 4.9, it is observed that all the variables have communalities with high values closer to the ideal value of 1.0. The data in Table 4.9 confirms the ideal minimum value of 0.6 for all communalities recommended by Field (2009) is satisfied by all the variables analysed.

4.1.2.4 Eigen Values and Total Variance Explained

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.776	29.604	29.604	1.776	29.604	29.604	1.437	23.955	23.955
2	1.169	19.484	49.088	1.169	19.484	49.088	1.246	20.769	44.724
3	1.009	16.822	65.911	1.009	16.822	65.911	1.043	17.383	62.107
4	0.797	13.286	79.196	0.797	13.286	79.196	1.025	17.089	79.196
5	0.725	12.079	91.275						
6	0.524	8.725	100.000						

Extraction Method: Principal Component Analysis.

Table 4.10: Total variance explained by institutional variables

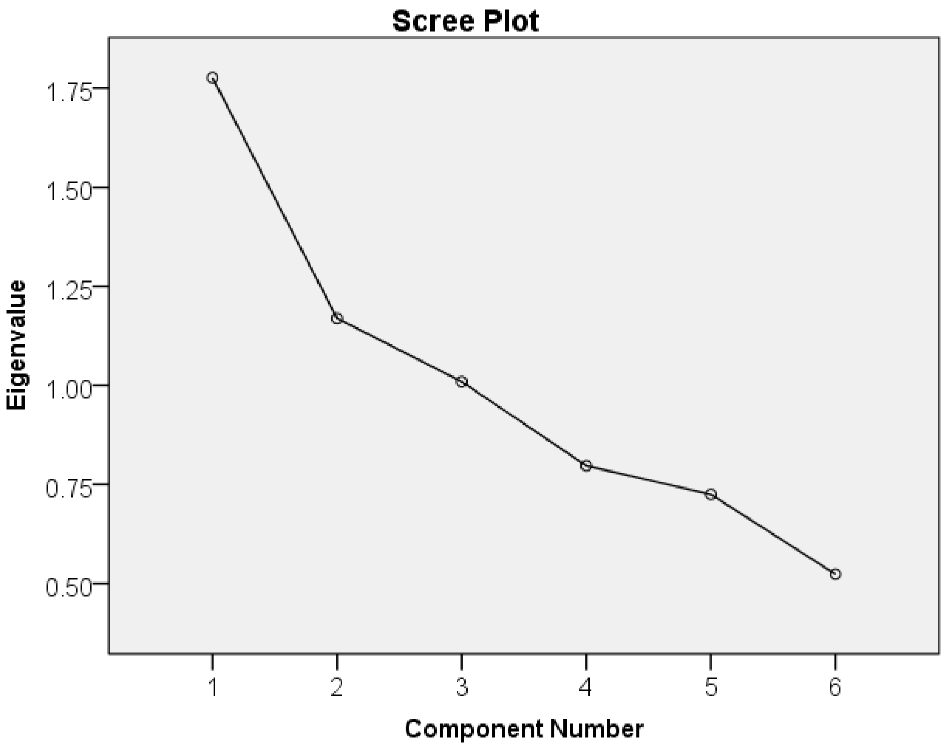

Figure 4.3: Scree plot for institutional factors

4.1.2.5 Factor Extraction from Institutional Variables

An observation of the Cattell scree plot in Figure 4.3 shows that there is no unique and distinct point of inflexion to be used as the criteria for extraction. Any point used would therefore be decided arbitrarily. This criterion was therefore dropped on that basis.

Using the Kaiser's criterion of extracting principal components with eigen values greater than 1.0, means that only the first 3 components will be extracted, accounting for only 65.911% of the total variation. This proportion is relatively low and the criterion was therefore dropped.

Using the Jolliffe's criterion of extracting components with eigen values greater than 0.7, means that the first 5 components will be extracted, accounting for up to 91.275% of the total variation. This proportion is too high and the extracted components would be too many.

This research therefore used the general criterion where a maximum cumulative variance of 75% is applied resulting in 4 extracted components as shown in Table 4.10.

The proportion of variance represented by the eigen values are relatively larger compared to those of individual staff variables. The eigen values meet the criteria of being in descending order and all are positive, hence the correlation matrix **R** in Table 4.8 is positive definite.

4.1.2.6 Clustering of Institutional Variables into Factors

The clusters of the original variables forming principal components based on orthogonal rotation and sorting are therefore shown in Table 4.11.

Rotated Component Matrix

Variable	Component 1	2	3	4
Usage of projector	0.851	0.060	-0.038	0.063
Access to printer	0.833	0.114	0.087	-0.038
Access to desktop PC	0.040	0.863	-0.024	0.118
Internet access at work	0.145	0.689	0.254	-0.081
Software availability	0.027	0.147	0.969	0.033
ICT compulsory in college	0.023	0.036	0.028	0.990

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Table 4.11: Clusters of institutional variables - highlighted

Figure 4.4 shows a graphical presentation of the results in Table 4.11 using a scatter plot with the first two principal components.

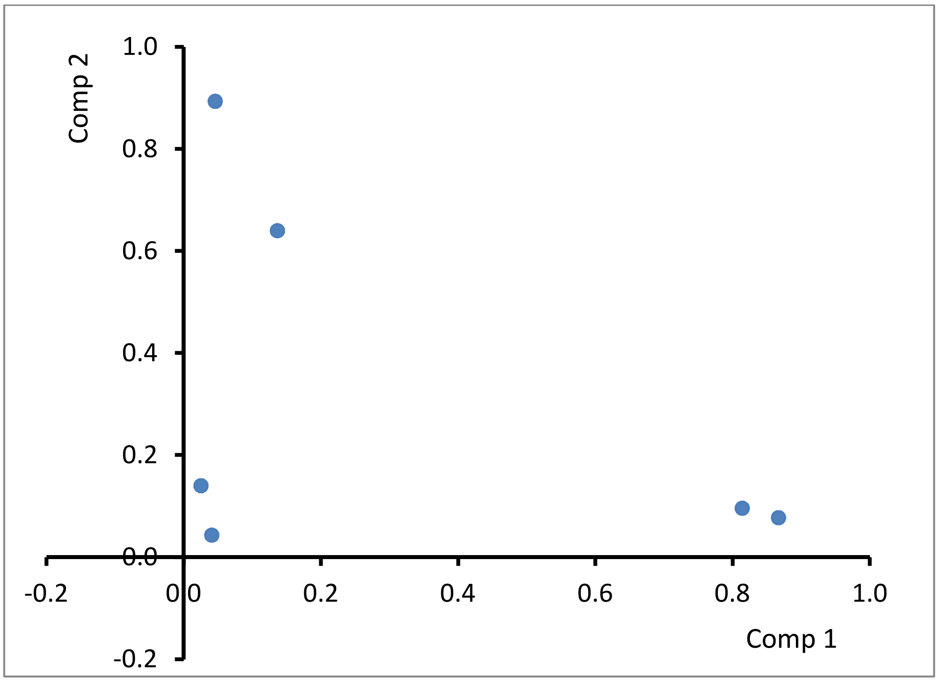

Figure 4.4: Scatter plot of rotated scores - institutional

Rotated Component Matrix

Variable	Component			
	1	2	3	4
Usage of projector	0.867			
Access to printer	0.814			
Access to desktop PC		0.894		
Internet access at work		0.640		
ICT required at employment			0.974	
Software availability				0.964

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.
b. Values less than 0.50 have been excluded.

Table 4.12: Clusters of institutional variables - filtered

An observation of the rotated component matrix in Table 4.12 reveals the clusters of institutional variables indicated by the values retained after those less than 0.50 have been excluded are as summarised in the Table 4.13. The natural groupings of the variables and the corresponding components are also indicated in the table.

Extracted Components and Variables

	Component or Factor	**Variables forming the Cluster**
1	Hardware accessories	• *Usage of projector* • Access to printer
2	Accessibility	• *Access to desktop PC* • Internet access at work
3	Employment factors	• *ICT required at employment*
4	Software	• *Software availability*

Table 4.13: Principal components for institutional factors

4.1.2.7 Principal Components of Institutional Attributes

Component Score Coefficient Matrix

	Component			
	1	2	3	4
ICT required at employment	-0.044	-0.017	0.942	-0.034
Internet access at work	-0.015	0.476	0.081	0.143
Software availability	-0.030	-0.121	-0.037	0.981
Access to desktop PC	-0.098	0.809	-0.098	-0.253
Access to printer	0.564	-0.072	0.162	0.043
Usage of projector	0.635	-0.046	-0.203	-0.076

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 4.14: Eigen vectors of institutional factors

From Table 4.14, the principal components for institutional factors are therefore given by:

$$z_1 = -0.044x_1 - 0.015x_2 - 0.030x_3 - 0098x_4 + 0.564x_5 + 0.635x_6$$

$$z_2 = -0.017x_1 + 0.476x_2 - 0.121x_3 + 0.809x_4 - 0.072x_5 - 0.046x_6$$

$$z_3 = 0.942x_1 + 0.081x_2 - 0.037x_3 - 0.098x_4 + 0.162x_5 - 0.203x_6$$

$$z_4 = -0.034x_1 + 0.143x_2 + 0.981x_3 - 0.253x_4 + 0.043x_5 - 0.076x_6$$

where the variable $\boldsymbol{x_i}$ is as follows:

i	x_i *variable*
1.	ICT required at employment
2.	Internet access at work
3.	Software availability
4.	Access to a desktop PC
5.	Access to a printer
6.	Usage of a projector

The correlation matrix of the principal components is an identity matrix with all elements along the leading diagonal equal to 1.00 while the other elements are zero. This means that the principal components satisfy the two conditions of non-collinearity.

$$cor\left(z_i,\ z_j\right) = \begin{cases} 0 & \text{for} \quad i \neq j \\ 1 & \text{for} \quad i = j \end{cases}$$

4.1.3 Extraction of Significant Variables

Individual Staff and Institutional Variables

From each principal component, one explanatory variable with the highest loading is extracted to represent the cluster, and is the one used in subsequent analysis in the ordinal logistic regression.

Individual variables

Age of staff
e-mail on Internet
Staff category
Type of mobile phone
Management level
Mode of ICT training
Gender
Job satisfaction level

Institutional variables

Access to projector
Access to desktop PC
ICT required at employment
Software availability

4.2 Analysis of ICT Integration Using Logistic Regression

4.2.1 Estimation of Parameters and Modelling of Individual Factors

Summarising all the values of the variables and their respective frequencies gives:

Case Processing Summary

		N	Marginal Percentage
Level of ICT integration	None	5	3.2%
	Low	88	56.1%
	High	64	40.8%
Gender	Male	84	53.5%
	Female	73	46.5%
Staff category	Non-teaching	50	31.8%
	Teaching	107	68.2%
Management Level	Bottom	49	31.2%
	Middle	102	65.0%
	Top	6	3.8%
Mode of ICT training	Informal	34	21.7%
	Formal	123	78.3%
E-mail on Internet	No	23	14.6%
	Yes	134	85.4%
Job satisfaction level	V dissatisfied	5	3.2%
	Dissatisfied	5	3.2%
	Moderate	57	36.3%
	Satisfied	79	50.3%
	V satisfied	11	7.0%
Mobile phone type	Ordinary	25	15.9%
	Smart	104	66.2%
	Both	28	17.8%
Valid		157	100.0%
Missing		0	
Total		157	

Table 4.15: Case summary of individual factors

Fitting the logistic regression model with the 8 extracted explanatory variables and analysing the model below by estimating the parameters gives:

Model: IntegLevel $\sim$ Gender + Age + StaffCateg + MngtLevel + TrainMode + EmailUse + JobSatis + PhoneType

	Estimates					Odds Ratio			
	Value	Std. Error	t value	95% confidence interval		$\exp(\hat{\beta}_i)$		95% confidence interval	
Coefficients:	$\hat{\beta}_i$			Lower	Upper	Absolute	Percent	Lower	Upper
Intercept - model 11	-2.8816	1.5804	-1.8233	-5.9792	0.2160	0.0560	5.6%	0.0025	1.2411
Intercept - model 12	1.6186	1.5591	1.0382	-1.4372	4.6744	5.0460	504.6%	0.2376	107.1721
Gender: Male	0.4977	0.3578	1.3909	-0.2036	1.1989	1.6449	164.5%	0.8158	3.3166
Age	-0.0555	0.0220	-2.5249	-0.0986	-0.0124	0.9460	94.6%	0.9061	0.9876
StaffCateg: Teaching	0.0604	0.4155	0.1453	-0.7541	0.8748	1.0622	106.2%	0.4705	2.3984
MngtLevel: Middle	0.4108	0.4059	1.0122	-0.3847	1.2063	1.5080	150.8%	0.6807	3.3412
MngtLevel: Top	-0.6768	1.1480	-0.5896	-2.9269	1.5733	0.5082	50.8%	0.0536	4.8224
TrainMode: Formal	1.0815	0.4669	2.3162	0.1663	1.9967	2.9491	294.9%	1.1809	7.3646
EmailUse: Yes	1.2516	0.5142	2.4342	0.2438	2.2593	3.4958	349.6%	1.2761	9.5764
JobSatis: Dissatisfied	0.4679	1.3767	0.3398	-2.2305	3.1662	1.5966	159.7%	0.1075	23.7179
JobSatis: Moderate	-0.6366	0.9913	-0.6422	-2.5794	1.3063	0.5291	52.9%	0.0758	3.6924
JobSatis: Satisfied	-0.2428	0.9792	-0.2480	-2.1620	1.6764	0.7844	78.4%	0.1151	5.3460
JobSatis: V satisfied	1.6396	1.2404	1.3218	-0.7917	4.0708	5.1529	515.3%	0.4531	58.6023
PhoneType: Smart	1.3452	0.5409	2.4871	0.2851	2.4052	3.8388	383.9%	1.3299	11.0810
PhoneType: Both	1.2654	0.6383	1.9826	0.0144	2.5164	3.5445	354.4%	1.0145	12.3834

Table 4.16: Parameter estimates and odds ratios of individual factors

Residual Deviance: 214.3781
AIC: 244.3781

The two ordinal logistic regression models for individual staff factors are therefore fitted onto the data as derived from Table 4.16 and given by:

Model 11:

$$ln\left[\frac{P(Y=none)}{P(Y=low,\ high)}\right] = -2.8816 + 0.4977X_1 - 0.0555X_2 + 0.0604X_3 + 0.4108X_4 - 0.6768X_5$$
$$+ 1.0815X_6 + 1.2516X_7 + 0.4679X_8 - 0.6366X_9 - 0.2428X_{10}$$
$$+1.6396X_{11} + 1.3452X_{12} + 1.2654X_{13} \qquad \text{.......................................(4.1)}$$

Model 12:

$$ln\left[\frac{P(Y=none,\ low)}{P(Y=high)}\right] = 1.6186 + 0.4977X_1 - 0.0555X_2 + 0.0604X_3 + 0.4108X_4 - 0.6768X_5$$
$$+ 1.0815X_6 + 1.2516X_7 + 0.4679X_8 - 0.6366X_9 - 0.2428X_{10}$$
$$+1.6396X_{11} + 1.3452X_{12} + 1.2654X_{13} \qquad \text{.......................................(4.2)}$$

4.2.2 Checking Adequacy of Model for Individual Staff

Using the chi-square goodness of fit

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	230.237	289	0.995
Deviance	208.833	289	1.000

Link function: Logit.

Deviance = Chi-square calculated (χ^2_{calc}) = 208.833

Chi-square critical = $\chi^2_{critical}$ = $\chi^2_{0.05,\,289}$ with 289 d.f. at $\alpha = 0.05$

= 329.648935

Since χ^2_{calc} = 208.833 < $\chi^2_{critical}$ = 329.648935 therefore the research concludes that the fit is a fair fit.

Using the Nagelkerke Pseudo R-squared

Pseudo R-Square

Cox and Snell	0.209
Nagelkerke	0.262
McFadden	0.147

Link function: Logit.

The Nagelkerke Pseudo R-squared of 0.262 although relatively low implies that the model fits the data being analysed albeit not strong.

This means that the explanatory variables in the model account for 26.2% in determining the value of the response variable – ICT integration, and therefore the remaining 73.8% is unexplained variation which is accounted for by factors not included in the model.

4.2.3 Estimation of Parameters and Modelling of Institutional Factors

Case Processing Summary

		N	Marginal Percentage
Level of ICT integration	None	5	3.2%
	Low	88	56.1%
	High	64	40.8%
Usage of projector	No	121	77.1%
	Yes	36	22.9%
Access to desktop PC	No	14	8.9%
	Yes	143	91.1%
ICT required at employment	No	95	60.5%
	Yes	57	36.3%
	Not sure	5	3.2%
Software availability	No	60	38.2%
	Yes	40	25.5%
	Some	47	29.9%
	Not sure	10	6.4%
Valid		157	100.0%
Missing		0	
Total		157	

Table 4.17: Case summary of individual factors

Fitting the logistic regression model with the 4 extracted explanatory variables and analysing the model below by estimating the parameters gives:

Model: IntegLevel ~ Projector + Desktop + ICTrequired + SoftwAvail

	Estimates					**Odds Ratio**			
	Value	**Std. Error**	**t value**	**95% confidence interval**		$\exp(\hat{\beta}_i)$		**95% confidence interval**	
Coefficients:	$\hat{\beta}_i$			**Lower**	**Upper**	**Absolute**	**Percent**	**Lower**	**Upper**
Intercept – model 21	-1.8723	0.7589	-2.4671	-3.3597	-0.3849	0.1538	15.4%	0.0347	0.6805
Intercept – model 22	2.6239	0.7690	3.4120	1.1167	4.1311	13.7894	1378.9%	3.0546	62.2488
Projector: Yes	1.3101	0.4353	3.0095	0.4569	2.1633	3.7065	370.7%	1.5792	8.6997
Desktop: Yes	1.3416	0.7603	1.7646	-0.1486	2.8318	3.8252	382.5%	0.8619	16.9758
ICTrequired: Yes	0.2251	0.3922	0.5739	-0.5436	0.9938	1.2524	125.2%	0.5806	2.7015
ICTrequired: Not sure	-0.7933	1.0877	-0.7293	-2.9252	1.3386	0.4523	45.2%	0.0537	3.8137
SoftwAvail: Some	0.5626	0.4404	1.2773	-0.3006	1.4258	1.7552	175.5%	0.7404	4.1611
SoftwAvail: Yes	1.7576	0.4926	3.5677	0.7921	2.7231	5.7985	579.9%	2.2080	15.2274
SoftwAvail: Not sure	-0.9658	0.8936	-1.0808	-2.7173	0.7857	0.3807	38.1%	0.0661	2.1938

Table 4.18: Parameter estimates and odds ratios of institutional factors

Residual Deviance: 209.5733

AIC: 227.5733

The two ordinal logistic regression models for institutional factors are therefore fitted onto the data as derived from Table 4.18 and given by:

Model 21:

$$ln\left[\frac{P(Y = none)}{P(Y = low,\ \ high)}\right] = -1.8723 + 1.3101X_1 + 1.3416X_2 + 0.2251X_3 - 0.7933X_4 + 0.5626X_5 + 1.7576X_6 - 0.9658X_7 \quad \text{………………………………………………………………(4.3)}$$

Model 22:

$$ln\left[\frac{P(Y = none,\ \ low)}{P(Y = high)}\right] = 2.6239 + 1.3101X_1 + 1.3416X_2 + 0.2251X_3 - 0.7933X_4 + 0.5626X_5 + 1.7576X_6 - 0.9658X_7 \quad \text{………………………………………………………………(4.4)}$$

4.2.4 Checking Adequacy of Model for Institution

Using the chi-square goodness of fit

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	72.514	35	0.000
Deviance	33.901	35	0.521

Link function: Logit.

Deviance = Chi-square calculated ($\mathcal{X}^2_{calc}$) = 33.901

Chi-square critical = $\mathcal{X}^2_{critical}$ = $\mathcal{X}^2_{0.05,\,35}$ with 35 d.f. at $\alpha = 0.05$

= 49.801850

Since $\mathcal{X}^2_{calc}$ = 33.901 < $\mathcal{X}^2_{critical}$ = 49.801850 therefore the research concludes that the fit is a poor fit.

Using the Nagelkerke Pseudo R-squared

Pseudo R-Square

Cox and Snell	0.233
Nagelkerke	0.292
McFadden	0.166

Link function: Logit.

The Nagelkerke Pseudo R-squared of 0.292 although lies between moderate and poor, it implies that the model fits the data being analysed albeit not strong.

This means that the explanatory variables in the model account for 29.2% in determining the value of the response variable – ICT integration, and therefore the remaining 70.8% is unexplained variation which is accounted for by factors not included in the model.

4.3 Findings and Discussions on ICT Factors

4.3.1 Findings on Individual Staff Factors

4.3.1.1 Interpretation of the Odds ratios of Explanatory Variables

From Table 4.16, the research derived the following interpretations on the level of ICT integration from estimates of the odds ratios of the variables pertaining to individual staff.

Gender - male: OR = e^{β_i} = 1.6449

Therefore a male employee is 64.5% more likely to integrate ICT compared to a female employee. Male employees therefore integrate ICT more than their female counterparts. This concurred with a report by World Bank (2011) that found out gender to be a crucial factor in the usage of ICT with females using ICT less than their male counterparts.

Age: OR = e^{β_i} = 0.9460

This means that for every additional year in age, an employee is 5.4% less likely to integrate ICT. The research found out that age reduces ICT integration among employees. In general, ICT integration is higher among the young than the old employees. This concurred with the popular belief that the young people use ICT more than the old.

Staff category - teaching: OR = e^{β_i} = 1.0622

A teaching staff member is 6.2% more likely to integrate ICT than the non-teaching staff. The research found out that the teaching staff members integrate ICT more than their non-teaching counterparts. Generally, the teaching staff members tend to have a higher level of education compared to their non-teaching counterparts. Therefore taking the level of education as a latent variable in this case, this finding concurred with Kamuti (2009) that ICT integration varies positively with level of education.

Management level - middle: OR = e^{β_i} = 1.5080

This means that an employee in middle management level is 50.8% more likely to integrate ICT compared to an employee at the bottom management level.

Management level - top: OR = e^{β_i} = 0.5082

An employee in top management level is 41.2% less likely to integrate ICT compared to an employee at the bottom management level. This was attributed to the fact that the top management of an organisation is responsible for strategic decisions of a long term nature where planning and policy formulation is the main responsibility. Such activities may require the computer less intensively compared to the bottom level where the work is predominantly clerical and routine in nature and therefore requires the computer more.

Training mode - formal: OR = e^{β_i} = 2.9491

Therefore an employee who acquired ICT skills through formal training is 2.949 times more likely to integrate ICT than an employee who acquired ICT skills through informal training. The research found out that formal training in ICT increases ICT integration among employed staff. Naturally ICT is a technical area and therefore formal training is superior in disseminating the skills required to use the ICT facilities proficiently in terms of breadth and depth.

E-mail use - yes: OR = e^{β_i} = 3.4958

An employee that uses e-mail is 3.496 times more likely to integrate ICT compared to an employee that does not use e-mail. The research found out that the use of e-mail by an employed staff increases his ICT integration. To use e-mail automatically involves the use of a computer or an embedded computer in the form of a digital device such as a phone or i-pad.

Job satisfaction - dissatisfied: OR = e^{β_i} = 1.5966

An employee who is dissatisfied with his job is 59.7% more likely to integrate ICT compared to an employee who is very dissatisfied with his job.

Job satisfaction - moderate: OR = e^{β_i} = 0.5291

An employee who is moderately satisfied with his job is 47.1% less likely to integrate ICT compared to an employee who is very dissatisfied with his job.

Job satisfaction - satisfied: OR = e^{β_i} = 0.7844

An employee who is satisfied with his job is 21.6% less likely to integrate ICT compared to an employee who is very dissatisfied with his job.

Job satisfaction – very satisfied: OR = e^{β_i} = 5.1529

Therefore an employee who is very satisfied with his job is 5.153 times more likely to integrate ICT compared to an employee who is very dissatisfied with his job.

The research found out that job satisfaction increases ICT integration at the lowest and highest levels, but reduces ICT integration at the middle levels. This was accounted for by the fact that job satisfaction is very subjective, too personal and very difficult to quantify, resulting in lack of consistency in the findings.

Phone type - smart: OR = e^{β_i} = 3.8388

This means that an employee who has a smart phone is 3.839 times more likely to integrate ICT compared to an employee who has an ordinary phone. Possession of a smart phone by an employee increases ICT integration. This could be explained by the fact that a person who acquires a smart phone does so because he is probably interested in the extra features of a mobile phone beyond just sending and receiving calls and messages.

4.3.1.2 Statistical Inference on the Co-efficients and Odds ratios

From Table 4.16, the research drew the following inferences on the level of ICT integration from estimates of the odds ratios of the variables pertaining to an individual staff.

Testing of hypothesis was carried out at the 5% level of significance.

To test for the significance of the odds ratio e^{β_i}, the research formulates the hypothesis

$$H_0: \ e^{\beta_i} = 1, \qquad \text{vs} \qquad H_1: \ e^{\beta_i} \neq 1,$$

The confidence limits for the odds ratio e^{β_i} is therefore given by

$$e^{\beta_i} = e^{\widehat{\beta}_i \pm Z_{\propto/2} S_{\widehat{\beta}_i}}, \text{ and} \qquad \text{95\% confidence limits} \qquad e^{\beta_i} = e^{\widehat{\beta}_i \pm 1.96\, S_{\widehat{\beta}_i}}$$

If the confidence interval (C.I.) for OR e^{β_i} includes 1, then $H_0: e^{\beta_i} = 1$ is not rejected,

otherwise $H_0: e^{\beta_i} = 1$ is rejected.

Gender: 95% confidence interval for OR = [0.8158, 3.3166] which includes 1.

Therefore the null hypothesis H_0 was not rejected at the 5% level. The research concluded that gender does not make a significant contribution in determining the level of ICT integration. The difference observed was therefore attributed to chance errors of sampling.

This was explained by the fact that at this level of tertiary institutions, most employees are at a level where gender factors have been overtaken, and female employees have an equal time on the computer as their male counterparts.

Age: 95% confidence interval for OR = [0.9061, 0.9876] which excludes 1.

Therefore the null hypothesis H_0 was rejected, and the research concluded that age makes a significant contribution in determining the level of ICT integration among employees.

This is probably because young employees tend to have more interest in ICT and they adventure more as opposed to the old who always tend to be contented with the status-quo.

Staff category: 95% confidence interval for OR = [0.4705, 2.3984] which includes 1.

Therefore the null hypothesis H_0 was not rejected, and the research concluded that staff category does not make a statistically significant contribution in determining the level of ICT integration between the teaching and non-teaching staff. The difference observed was therefore attributed to chance errors of sampling.

This is because both teaching and non-teaching staff need ICT to perform most of the tasks. There could be a difference within each category but not between the categories.

Training mode: 95% confidence interval for OR = [1.1809, 7.3646] which excludes 1.

Therefore the null hypothesis H_0 was rejected, and the research concluded that the mode of ICT training makes a significant contribution in determining the level of ICT integration.

This was attributed to the fact that formal training is naturally superior to informal training in terms of acquisition of ICT skills by the trainee. Formal training is usually structured with a specific set of objectives outlined in a curriculum, and a series of examinations for evaluation. Past researches in the literature review did not capture this variable, and therefore to this research, this was a new finding in ICT integration.

E-mail use: 95% confidence interval for OR = [1.2761, 9.5764] which excludes 1.

Therefore the null hypothesis H_0 was rejected, and the research concluded that e-mail use makes a significant contribution in the determination of the level of ICT integration.

This was probably because for any person to use e-mail, he must use an ICT gadget whether phone, i-pad or computer which must be connected to a communication network. This was also a new finding in ICT integration since the variable has not been included in past studies.

Phone type - smart: 95% confidence interval for OR = [1.3299, 11.0810] which excludes 1.

Therefore H_0 was rejected, and the research concluded that the possession of a smart phone makes a significant contribution in the determination of the level of ICT integration.

Generally, it is observed that those who posses smart phones are intensive and advanced users of the mobile phone beyond the basic functionality of sending and receiving calls and messages. To this research, this was a new finding as past studies did not include the mobile phone as a variable in the model.

Job satisfaction – dissatisfied: 95% C.I. for OR = [0.1075, 23.7179] includes 1.
Job satisfaction – moderate: 95% C.I. for OR = [0.0758, 3.6924] includes 1.
Job satisfaction – satisfied: 95% C.I. for OR = [0.1151, 5.3460] includes 1.
Job satisfaction – V. satisfied: 95% C.I. for OR = [0.4531, 58.6023] includes 1.

All these four confidence intervals include 1. Therefore H_0 was not rejected for each of the above levels of job satisfaction. The research concluded that job satisfaction does not make a significant contribution in the determination of the level of ICT integration among employees. This was also a new finding to this research as past researches did not include job satisfaction.

4.3.2 Findings on Institutional Factors

4.3.2.1 Interpretation of the Odds ratios and Explanatory Variables

From Table 4.18, the research derived the following interpretations on the level of ICT integration from estimates of the odds ratios of the variables pertaining to an institution.

Projector: OR = e^{β_i} = 3.7065

Therefore an employee who has access to an LCD projector is 3.707 times more likely to integrate ICT compared to an employee who does not have access to an LCD at the place of work. The research found out that access to a computer projector increases ICT integration by an employee. A projector is merely used to display the contents of a computer screen to a much larger audience, so to use a projector you must use a computer.

Desktop PC: OR = e^{β_i} = 3.8252

This means that an employee who has access to a desktop computer is 3.825 times more likely to integrate ICT compared to an employee who does not have access to a desktop computer at the place of work. The research therefore found out that access to a desktop computer in the office increases ICT integration by an employee. With an unlimited access to a computer in the office, there is naturally an irresistible urge to use it.

ICT skills requirement at employment – yes: OR = e^{β_i} = 1.2524

An employee to whom ICT skills was a mandatory requirement at the time of employment is 25.2% more likely to integrate ICT compared to an employee to whom ICT skills was not a mandatory requirement. The research found out that ICT skills requirement by the employer at the time of employment increases ICT integration among employees. This would be explained by the fact that any person who has ICT skills naturally has a tendency to use it.

ICT requirement at employment – not sure: OR = e^{β_i} = 0.4523

Therefore an employee who is not sure of whether ICT skills was a mandatory requirement at employment is 54.8% less likely to integrate ICT compared to an employee to whom ICT skills was not a mandatory requirement.

Software availability - some: OR = e^{β_i} = 1.7552

An employee to whom some of the required software is made available is 75.5% more likely to integrate ICT compared to an employee to whom the required software is not made available.

Software availability - yes: OR = e^{β_i} = 5.7986

Therefore an employee to whom all the required software is made available is 5.799 times more likely to integrate ICT compared to staff to whom the required software is not availed.

The research found out that making available all or some of the software required by an employee increases ICT integration. Software is an integral part of a computer system as without software the hardware is meaningless. Therefore the larger the variety of software installed in the computer, the more versatile the computer becomes.

Software availability – not sure: OR = e^{β_i} = 0.3807

An employee who is not sure of whether all the required software is made available is 61.9% less likely to integrate ICT compared to an employee to whom the required software is not made available.

4.3.2.2 Statistical Inference on the Co-efficients and Odds ratios

From Table 4.18, the research drew the following inferences on the level of ICT integration from estimates of the odds ratios of the variables pertaining to an institution.

Testing of hypothesis was carried out at the 5% level of significance.

To test for the significance of the odds ratio e^{β_i}, the research formulates the hypothesis

H_0: $e^{\beta_i} = 1$, vs H_1: $e^{\beta_i} \neq 1$,

The confidence limits for the odds ratio e^{β_i} is therefore given by

$e^{\beta_i} = e^{\widehat{\beta}_i \pm Z_{\propto/2} S_{\widehat{\beta}_i}}$, and 95% confidence limits $e^{\beta_i} = e^{\widehat{\beta}_i \pm 1.96\, S_{\widehat{\beta}_i}}$

If the confidence interval for OR e^{β_i} includes 1, then H_0: $e^{\beta_i} = 1$ is not rejected,

otherwise H_0: $e^{\beta_i} = 1$ is rejected.

Projector: 95% confidence interval for OR = [1.5792, 8.6997] which excludes 1.

Therefore the null hypothesis H_0 was rejected, and the research concluded that access to a projector makes a significant contribution in the determination of the level of ICT integration.

This is naturally because to use a projector, it must be connected to a computer, at least at the place of work. This could be for teaching, training, meetings, exhibitions, etc.

Desktop PC: 95% confidence interval for OR = [0.8619, 16.9758] which includes 1.

Therefore H_0 was not rejected, and the research concluded that access to a desktop computer does not make a significant contribution in the determination of the level of ICT integration. Any serious computer user will most likely acquire one if not provided by the institution.

ICT required - yes 95% confidence interval for OR = [0.5806, 2.7015] which includes 1.

Therefore the null hypothesis H_0 was not rejected, and the research concluded that ICT skills set as a mandatory requirement at employment does not make a significant contribution in the determination of the level of ICT integration.

This was probably because ICT usage or integration is more inclined towards personal interest and attitudes than an employment issue. Those who integrate ICT tend to do so out of personal interest and intrinsic satisfaction derived from it rather than as a requirement by the employer.

Software avail. - some: 95% confidence interval for OR = [0.7404, 4.1611] includes 1.

Therefore the null hypothesis H_0 was not rejected, and the research concluded that availability of only some of the software required by an employee does not make a significant contribution in the determination of the level of ICT integration.

Software avail. - yes: 95% confidence interval for OR = [2.2080, 15.2274] excludes 1.

Therefore the null hypothesis H_0 was rejected, and the research concludes that availability of all of the software required by an employee makes a significant contribution in the determination of his level of ICT integration.

This was attributed to the fact that those who integrate ICT generally have an affinity to use certain specialised software. Such software are generally too expensive for a user to purchase at an individual level, hence total reliance on the institution.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter focused on the conclusions drawn by the research based on the statistical analysis of the data, findings and inferences. It also tried to evaluate the external validity of the research by looking at the scope, the data collected, tools of analysis used and finally the conclusions drawn from analysis and recommendations made.

5.2 Conclusions

Although the level of multi-collinearity among variables was perceived to be high, it was found to be relatively low with only a small number of variables measuring similar dimensions in ICT integration. This meant that the number of variables in a cluster or factor still remained relatively low. The clusters were however found to be very distinct in composition.

The factors that affect the level of ICT integration among employed staff are both individual and institutional. Most of the factors perceived to influence ICT integration were found to be significant with only a few exceptions. Gender and job satisfaction which initially appeared too obvious were found not to be statistically significant. Gender was attributed to the fact that at the level of professionals, women are less overburdened by domestic chores which would significantly hinder them from integrating ICT. More individual factors were found to be significant in determining the level of ICT integration than the institutional factors.

5.3 Recommendations

5.3.1 Recommendations to the Stakeholders in ICT

Stakeholders include academic institutions, national and international organisations, government agencies, government and financiers:

Promotion of formal training in ICT since it has a significant impact on ICT integration. This can be achieved through incorporating ICT training in school and college curriculum, organising short ICT proficiency training for staff, and conducting seminars and workshops in ICT for proficiency training. Also more resources should be directed to training of the young employees as they integrate ICT more than their old counterparts.

Policies and programmes aimed at promoting ICT integration among staff should not focus on the gender element as part of affirmative action to promote women. As a factor, gender was not found to be statistically significant, although male employees were found to integrate ICT more than their female counterparts.

Provision of all facilities that enhance ICT integration to enable employees adopt ICT usage in a productive way. These include procurement of all application software required by staff, projectors, Internet access, probably cabled and wireless, etc.

Institutions should not make ICT skills a mandatory requirement at employment as the research found out that its influence in ICT integration exists, but is not significant.

5.3.2 Recommendations for Further Research

The external validity of this research on a wider scale could not be fully ascertained, except if applied in relatively similar situations like secondary schools in urban areas, and research organisations since they are academically inclined. This is because the scope was limited to one geographical area, an urban environment, and most respondents were from public institutions, and also the working environment, level of education, and ICT infrastructure were relatively similar across the institutions in the study.

Further research on a comparative analysis of PCA involving numerical and categorical data should be done to evaluate the difference, if any, on the number of extracted variables between categorical and numerical data, and to assess the difference in model adequacy.

Further research on ICT integration covering a wider geographical area, a larger number of institutions, a larger variety of institutions, and both rural and urban areas of residence. This will enable collection of data from a more heterogeneous population.

The models used in the analysis did not take into account the possibility of interaction between variables. Whether some variables may have significant impact only when combined with others by introducing interaction terms in the model was not ascertained by this research. Further research using data analysis involving models with interaction between variables to evaluate the impact of introducing interaction should be carried out.

References

Berk, K. N., & Carey, P. (2010). *Data Analysis with Microsoft Excel 2007.* Boston: Brooks/Cole.

Dobson, A. J. (2002). *An Introduction to Generalised Linear Models.* New York: Chapman & Hall/CRC.

E-Government, D. o. (2011). *ICT Standards and Guidelines, Version 0.1.* Nairobi: Office of the President.

Emerald, I. (2014). *Assumptions of Logistic Regression.* Retrieved from Statistics Solutions: https://www.statisticssolutions.com/assumptions-of-logistic-regression. Retrieved 20/12/2014

Field, A. (2009). *Discovering Statistics Using SPPS.* London: Sage Publications Ltd.

Garmin. (2013). *Global Positiong System.* Wikipedia http://www.garmin.com/en-US. Retrieved 10/07/2013.

Gathano, C. W. (2009). *An Ivestigation of Factors Influencing the Use of ICT in Teaching and Learning in Sec Schools in Thika District.* Nairobi: MA Research Thesis, University of Nairobi.

Government, K. (2010). *The Constitution of Kenya.* Nairobi: National Council for Law Reporting.

Halloran, S. O. (Retrieved 04/10/2013). Probit/Logit. *Sustainable Development*.

Hewson, P. J. (2009). *Multivariate Statistics with R.*

Hewson, P. J. (2009). *Multivariate Statistics with R.*

Kamui, M. M. (2008). *Use of logistic Regression to Identify key Determinants of Academic Performance in primary School.* University of Nairobi: Unpublished MSc Thesis.

Kamuti, S. N. (2009). *Evaluation of the Factors Influencing ICT Use on Management of NGOs in Embu District.* University of Nairobi: Unpublished MSc Thesis.

Kothari, C. R. (2006). *Quantitative Techniques.* New Delhi: Vikas Publishing House.

Kothari, C. R. (2011). *Research Methodology.* New Delhi: New Age International.

Kremelberg, D. (2011). *Practical Statistics.* Carlifonia: Sage Publications.

Loudon, K. C., & Loudon, J. P. (2001). *Essentials of Management Information Systems.* Prentice Hall.

Lucey, T. (1991). *Management Information Systems.* London: BookPower.

Ministry, I. C. (2006). *National Information and Communications Technology (ICT) Policy.* Nairobi: Government Press.

Mwai, P. K. (2005). *Application of Principal Component Analysis in Forex Market Variations.* Nairobi: University of Nairobi.

Nagler, J. (1994). Interpreting Probit Analysis. 1-15.

Nsibiro, R. (2009). *Gender Differentials in ICT Uptake.* Makerere University: International Journal of education and development using ICT, Vol 5, Issue 5, pp 33-42 .

Orodho, J. A. (2004). *Techniques of Writing Research Proposals and Reports in Education and Social Sciences.* Nairobi: Masola Publishers.

Oz, E., & Andy, J. (2006). *Management Information Systems.* Andover: Course Technology.

Shelly, G. B., & Rosenblatt, H. J. (2010). *Systems Analysis and Design. 8th Edition.* Boston: Course Technology.

Shelly, G. B., & Vermaat, M. E. (2011). *Discovering Computers.* Boston: Course Technology.

Spearmann, A. (2009). The Probit Model. *SoSe*.

Stevens, S. S. (1946). *Theory of Scale Types.* Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Measurement_scales: Retrieved 16/09/2014

UN. (2013). *e-Government for Development.* UN: https://en.wikipedia.org/wiki/E-Government#UN_e-Government_Readiness_Index. Retrieved 19/06/2013.

Wambugu, L. N. (2008). *Studying the Utilization of Free Primary Education in Urban Areas in Kenya Using Logistic Model.* Nairobi: MSc Research UoN.

Wikipedia. (2014, 05 17). *Principal Component Analysis.* Retrieved from Wikipedia: en.wikipedia.org/wiki/wikipedia

WorldBank. (2011). *Gender, ICT and Entrepreneurship.* World Bank Group, http://go.worldbank.org/8EHJW80CY0: Retrieved 26/04/2013.

Appendix

List of Tables

List of Figures

Questionnaire

Introductory Letter

Dear Respondent,

RE: DATA COLLECTION FOR ACADEMIC RESEARCH ON ICT INTEGRATION

Please consider this letter a sincere appeal to you to assist me, the undersigned, collect data for my academic research. I am currently pursuing a Master of Science degree course at the University of Nairobi, and therefore this research is purely for academic purposes. Data is specifically for research purposes only.

The goal of this research is to determine the current level of ICT integration and to find out the factors which either directly or indirectly influence ICT integration among staff in post school academic institutions in Machakos Town.

Please fill in the following questionnaire by providing information about your personal details. Your name is not a requirement, and therefore your identity remains anonymous.

Thanks in advance.

Yours Sincerely,

Moses M. Owana
P. O. Box 136
Machakos.

Non-ICT Staff Questionnaire - Individual Staff Details

(To be filled by non-ICT staff)

Questionnaire Serial Number: | S | | | |

Please fill in the following questionnaire by providing information about your personal details.
Your name is not a requirement, and therefore your identity remains anonymous.
Data is specifically for academic research purposes only and will be handled in confidence.

Section I: Individual Demographic Information

1. Gender: ☐ Male ☐ Female
2. Year of birth: 1 9 ……. …….
3. Your staff category: ☐ Teaching ☐ Non-teaching
4. Duration in employment: ………………………. years
5. Highest level of education you have attained:
   - ☐ Secondary
   - ☐ Certificate
   - ☐ Diploma
   - ☐ Bachelor degree
   - ☐ Masters/Ph.D. degree
   - ☐ Other (specify) ……………………………………….
6. In which year did you use the computer for the first time? ……………………………..
7. Indicate your hierarchy in management level in your institution:
   - ☐ Top ☐ Middle ☐ Bottom
8. Your profession by training: ………………………………………………………………

Section II: Individual Knowledge and Usage of ICT

9. Indicate the average frequency at which you use the computer:
   - ☐ Daily
   - ☐ None
   - ☐ Occasionally
   - ☐ Other (specify) ……………………………………….
10. Where did you acquire ICT skills for the first time?
    - ☐ Home
    - ☐ School
    - ☐ Other (specify) ………………………………….
    - ☐ College
    - ☐ Workplace
11. Indicate the mode of training through which you acquired ICT skills:
    - ☐ Formal training ☐ Informal training
12. In your opinion, will further training in ICT enable you improve on performance?
    - ☐ Yes ☐ No ☐ Not sure

13. Indicate the type(s) of Internet services in general which you use.

☐ E-mail ☐ Social media (Facebook, Twitter, WhatsApp)
☐ Browsing ☐ Other (specify) ....
☐ Skype

14. Rate your depth of knowledge of e-mail services into the categories below.

☐ None ☐ Low ☐ Average ☐ High

15. In your opinion, what is your level of satisfaction with your current employment?

☐ Very satisfied ☐ Dissatisfied
☐ Satisfied ☐ Very dissatisfied
☐ Moderate

16. Indicate the type of mobile phone which you use:

☐ Ordinary phone ☐ Smartphone ☐ Both

17. Do you use your mobile phone for Internet (e-mail, browsing)? ☐ Yes ☐ No

18. Was computer a compulsory subject during your college training?

☐ Yes ☐ No ☐ Not sure

19. Indicate your level of competency in general computer usage into the categories below:

☐ None ☐ Low ☐ Average ☐ High

20. Indicate your level of competency in installation of computer application software:

☐ None ☐ Low ☐ Average ☐ High

21. Indicate your level of competency in basic computer troubleshooting:

☐ None ☐ Low ☐ Average ☐ High

22. If **incompetent** in 19, 20 or 21, what challenges do you face in using computers well? *

1. ....
2. ....

23. If **low competency**, what type of skills do you need to improve usage of the computer? *

1. ....
2. ....

24. Please indicate if you use the following types of computer application software, and if **yes**, then specify the brand name of the software used. E.g. MS Word, MS Project. *

☐ Word processing ....
☐ Spreadsheet ....
☐ Presentation ....
☐ Desktop publishing ....

- [ ] Database management system ………………………………………………………………………
- [ ] Project management ………………………………………………………………………
- [ ] Accounting package ………………………………………………………………………
- [ ] Statistical package ………………………………………………………………………
- [ ] Computer aided design (CAD) ………………………………………………………………………
- [ ] Other (specify) ………………………………………………………………………

Section III: Institutional Information and ICT Infrastructure

25. Name of your institution: ………………………………………………………………………………………

26. Type of institution in which you work: ☐ Public ☐ Private

27. Status of institution in which you work: ☐ Tertiary ☐ University

28. Was basic ICT skills a mandatory requirement by the employer during your recruitment?

    ☐ Yes ☐ No ☐ Not sure

29. Do you have access to the Internet in your office/room? ☐ Yes ☐ No

30. Have all the software you require/requested for been made available to you?

    ☐ Yes ☐ No ☐ Some ☐ Not sure

31. Is anti-virus software installed in your computer at the place of work?

    ☐ Yes ☐ No ☐ Not sure

32. Indicate which of the following ICT gadgets you have access to use at your place of work:

    ☐ Desktop computer ☐ Computer printer
    ☐ Laptop computer ☐ LCD projector

33. Indicate the two most advanced features you use in the following software types: *

    Word processing: 1. ………………………………………………………………………………………
    2. ………………………………………………………………………………………
    e.g. proofreading, tables, mail merge, equation, table of contents, references, etc.

    Spreadsheet: 1. ………………………………………………………………………………………
    2. ………………………………………………………………………………………
    e.g. formulas, charts, range names, spreadsheet databases, macros, etc.

Section IV: End-user Classification

The respondent to be assisted by the researcher/enumerator if necessary

34. Classify your level of ICT integration as **high**, **low** or **none** based on the criteria on page 4:

    ☐ High ☐ Low ☐ None

Classification of respondents into any of the three levels of ICT integration is based on the following criteria which serve as a guideline:

High:

Usage of the computer is advanced and encompasses performing tasks related to one's responsibilities and duties at the workplace and in his profession in general. This includes but is not limited to:

- Usage of specialised software, e.g. computer aided design (CAD), statistical package, geographical information system (GIS), accounting package, database systems, project management software, video editing software;
- Usage at an advanced level of the general application packages, e.g. MS Word, MS Excel, MS PowerPoint, Lotus 1-2-3, PageMaker,
- All or most of the tasks which can be computerised are performed using the computer;
- Usage of Internet to browse and download information;
- Obtain e-services, fill on-line forms, watch and read digital news;
- Download software from the Internet;
- Install computer software;
- Backup documents on removable media (CDs, DVDs, memory cards, and external HDDs).

Low:

Usage of the computer is limited to very basic computer tasks such as:

- creating and editing simple documents;
- basic data entry, e.g. keying in student marks, student personal details, etc;
- printing of documents;
- sending and receiving e-mail;
- usage of social media e.g. Face Book, Twitter, WhatsApp, etc;
- using mobile phones for banking services.

None:

No usage of the computer as a tool to perform tasks at the place of work in one's profession. No usage of the computer at a personal level in any productive way. ICT usage is limited to very basic mobile phone services such as:

- sending and receiving calls;
- sending and receiving messages;
- using phones for mobile money transfers, such as M-Pesa, Airtel Money, Pesa Mkononi, etc.
- using mobile phones for e-payments, e.g. Pay Bill services.

ICT Staff Questionnaire - Institutional and ICT Staff Details

(To be filled by the Head of ICT Department/Section)

Your name is not a requirement, and therefore your identity remains anonymous. Data is specifically for academic research purposes only. Data provided is handled in confidence.

Questionnaire Serial Number: | T | | | |

Please fill in the following questionnaire by providing information about your institution.

Section I: Staff Demographic Information

1. Name of your institution: ………………………………………………………………………………

2. Year of birth: 1 9 ……. …….

3. Gender: ☐ Male ☐ Female

4. Duration in employment: ……………………… years

5. Your initial professional training: ………………………………………………………………

Section II: Institutional ICT Infrastructure and Capacity

6. Highest level of training offered by your institution:

   ☐ Degree ☐ Diploma ☐ Certificate

7. Highest level of education you have attained: tick as appropriate

   ☐ Secondary ☐ Bachelor degree
   ☐ Certificate ☐ Masters/Ph.D. degree
   ☐ Diploma ☐ Other (specify) ……………………………………………

8. Indicate the roles you play in ICT in your organisation: * tick as appropriate

   ☐ Hardware/software maintenance ☐ Website maintenance
   ☐ Network administration ☐ Database administration
   ☐ Systems administration ☐ Other ………………………………………

9. Does your institution have a local area network (LAN)? ☐ Yes ☐ No

10. If **yes** in 9, then indicate the type(s) of media used: tick as appropriate

    ☐ UTP cable ☐ Optical fibre cable
    ☐ Wireless (Wi-Fi) ☐ Other ………………………………………

11. Does your institution have Internet connectivity? ☐ Yes ☐ No

12. If **yes** in 11, then indicate the bandwidth ……………………… Mbps

13. Does your institution have an MIS installed for its operations? ☐ Yes ☐ No

14. Has your institution subscribed to e-library resources? ☐ Yes ☐ No

15. Number of staff by category in your institution: Teaching staff ………………………
Non-teaching staff ………………………

16. Total number of operational computers/devices by category in your institution:

Desktop computers ………………… Printers ……………………
Laptop computers ………………… Projectors ……………………

17. Number of professionally trained ICT support staff in the institution: ………………………………

18. Who initiates the acquisition/procurement of ICT hardware/software in your institution?

☐ End-user ☐ ICT staff
☐ Administration ☐ Other (specify) ……………………………………

19. Does your institution involve you in drawing ICT budget? ☐ Yes ☐ No

20. Indicate the sponsor of your ICT training.

☐ Self ☐ Both
☐ Employer ☐ Other (specify) ……………………………………

21. ICT maintenance type practised by your institution:

☐ Curative or randomly ☐ Both
☐ Preventive or regularly

22. Maintenance frequency if **preventive** in your institution:

☐ Weekly ☐ Termly/Semesterly
☐ Monthly ☐ Other (specify) ……………………………………

23. Student enrolment (regular/full-time): …………………………… students

24. Have you attended any short or long course in ICT in the last 5 years? *

☐ Yes ☐ No

25. List **three** strengths in order of priority which in your own opinion support/favour ICT adoption and integration in your institution. *

……

……

……

26. List **three** problems in order of priority which in your own opinion hinder/impede ICT adoption and integration in your institution. *

……

……

……