Comparative Analysis of Usability Between Text and Virtual Reality Based Expert Systems For Cattle Diseases Diagnosis

By

KABO POLOKO NKABITI

Reg. No: 14100047

BSc (Computing) (OU, UK)

Department of Computer Science and Information Systems,

College of Science,

Botswana International University of Science and Technology

E-mail:- kabo.nkabiti@studentmail.biust.ac.bw: Phone Number (+267 719 691 68)

A Research dissertation Submitted to College Science for the Study Leading to a Fulfilment of the Requirements for the Award of the Master of Science in Information Systems of BIUST

Supervisor(s): Dr Galani Malatsi

Department of Computer Science and Information Systems,

College of ICT, BIUST

E-mail Address, Phone Number including the code (+267 4900117)

Signature: _____ Date: 21/04/2017

Declaration and Copyright

DECLARATION AND COPYRIGHT

I, _____, declare that this dissertation/thesis is my own original work and that it has not been presented and will not be presented to any other university for a similar or any other degree award.

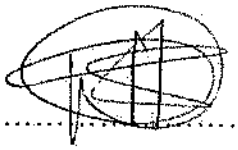
Signature .....

This dissertation/thesis is copyright material protected under the Berne Convention, the Copyright Act of 1999 and other international and national enactments, in that behalf, on intellectual property. It must not be reproduced by any means, in full or in part, except for short extracts in fair dealing; for researcher private study, critical scholarly review or discourse with an acknowledgement, without the written permission of the office of the Provost, on behalf of both the author and the BIUST.

Certification by Supervisor(s)

CERTIFICATION

The undersigned certifies that he/she has read and hereby recommends for acceptance by the College of Science a dissertation/thesis titled: **Comparative Analysis of Usability between Text and Virtual Reality Based Expert Systems for Cattle Diseases Diagnosis**, in fulfilment of the requirements for the degree of Master of Science/Doctor of Philosophy in (*Programme of Study*) of the BIUST.



.....

Dr. Dr Galani Malatsi

(Supervisor)

Date: 21/04/2017

Acknowledgements

I would like to thank everyone who had contributed to the successful completion of this research project. I would like to express my gratitude to my supervisor, Dr. Malatsi Galani, for his priceless advice and guidance, and his huge persistence all the way through the development of the research.

In addition, I would also like to express my gratitude to my loving mother Keneetswe Nkabiti, my girlfriend Tshoganetso Oabile and my daughter Imani Masa Poloko, who had assisted and given me both financial and emotional support. I would also like to thank my sponsor BIUST.

Dedication

I dedicate this thesis to my daughter Imani Masa Poloko.

Abstract

The use of cattle disease diagnosis expert systems has the ability to improve the cattle production as they would help in disease control. Particularly if the system is usable, that is, if it is efficient, effective, satisfying and requires less effort to use.

Cattle disease diagnosis expert systems have been adopted in recent years in the cattle production industry particularly in Asian countries like China. However the usability of these systems still remains a challenge. There is limited or no empirical research that investigates comparison of usability of a text and non-immersive virtual reality-based (multimedia) cattle disease diagnosis expert system.

This study seeks to bridge this knowledge gap by conducting a quasi-experiment research which evaluates the usability of a text and a non-immersive virtual reality-based (VR) cattle disease diagnosis expert system. Software Usability Measurement Inventory (SUMI) was used to gain insight in user satisfaction. Video-recording software (Morae TechSmith) was used to measure task efficiency and effectiveness. A self-reporting ratings using NASA Task Load Index (NASA-TLX scale) was used to gain an insight of mental effort (Cognitive workload) invested by users when performing a task. Sixty-four participants, consisting of aspiring, subsistence and commercial cattle farmers within the Palapye (Botswana) veterinary district, took part in the study. They were divided in two groups of 32, one group using a text-based expert system and other group using a VR-based expert system.

The results indicated that VR-based expert system is better than a text based expert system in terms of effectiveness, efficiency and user satisfaction. Moreover users using a text based expert exerted less mental effort (work load) than those using the VR-based system. The overall differences indicated that VR-based ES was 75% better than the Text based system which only scored 25%. Results acquired from this study imply that software developers should develop VR-based cattle diseases diagnosis expert system that is usable for cattle farmers.

TABLE OF CONTENTS

DECLARATION AND COPYRIGHT II

CERTIFICATION BY SUPERVISOR(S) III

TABLE OF CONTENTS VII

LIST OF TABLES XI

LIST OF FIGURES XII

LIST OF ABBREVIATIONS/ SYMBOLS XIII

CHAPTER 1 INTRODUCTION 14

1.1 INTRODUCTION OF THE RESEARCH 14

1.2 OVERVIEW OF CATTLE FARMING IN BOTSWANA 17

1.3 CATTLE DISEASES MANAGEMENT SYSTEMS 18

1.4 DIAGNOSIS EXPERT SYSTEMS IN AGRICULTURE 19

1.5 PROBLEM STATEMENT 21

1.6 MAIN OBJECTIVE 21

1.6.1 Specific Objectives 21

1.7 RESEARCH QUESTIONS 22

1.8 HYPOTHESIS 22

1.9 EXPECTED OUTCOME 23

1.10 JUSTIFICATION OF THE RESEARCH 23

1.11 RESEARCH DELIMITATION 24

1.12 CONTRIBUTION 24

1.13 THESIS STRUCTURE 25

CHAPTER 2 LITERATURE REVIEW 26

2.1 INTRODUCTION 26

2.2 USABILITY TESTING OF APPLICATIONS 26

2.3 METHODS FOR COMPARING THE SYSTEMS 28

2.3.1 Measures of Effectiveness 29

2.3.2 Measures of Efficiency 29

2.3.2.1 User Temporal Efficiency 30
2.3.3 Measuring user Satisfaction 30
2.4 THE COGNITIVE THEORY OF MULTIMEDIA LEARNING 31
2.4.1 Human information processing 31
2.4.2 Three assumptions of cognitive theory of multimedia learning 33
2.4.2.1 Dual Channel Assumption 33
2.4.2.2 Limited Capacity Assumption 34
2.4.2.3 Active Processing Assumption 34
2.4.3 The Three Cognitive Processes 34
2.4.3.1 Extraneous Cognitive Processing 35
2.4.3.2 Essential Cognitive Processing 35
2.4.3.3 Generative Processing 36
2.4.4 Five forms of representation in theory 36
2.4.5 The five cognitive processes 37
2.4.6 The Modality Principle 37
2.4.7 Cognitive load 40
2.4.7.1 Measuring cognitive load of multimedia learning 42
2.4.7.2 Conclusion of Cognitive load 43
2.4.8 Related Works 44
2.4.8.1 Review Existing Medical Expert systems 44
2.4.8.2 Literature Review Conclusion for existing systems 47
2.4.9 Comparison of Multimedia and other Instructional methods 47
2.4.10 Summary 49

CHAPTER 3 METHODOLOGY 50

3.1 INTRODUCTION 50
3.2 CONTEXT OF STUDY 50
3.3 RESEARCH DESIGN 51
3.4 PARTICIPANTS AND SAMPLING 52
3.5 RESEARCH APPROACH 53
3.6 RESEARCH PURPOSE 53
3.7 METHODS FOR COMPARING THE SYSTEMS 54

3.7.1 Measures of Effectiveness 55
3.7.2 Measures of Efficiency 55
3.7.3 Measures of Satisfaction 56
3.7.4 Measuring cognitive workload 57
3.7.5 The Overall Score 57
3.8 Research Questions and Hypothesis 58
3.9 Data collection and experiments 59
3.9.1 Methods used for collecting data 61
3.9.1.1 Survey Questionnaire 61
3.9.1.2 Video-based Observation 62
3.10 Reliability 63
3.11 Validity 63
3.12 Analysis of data 63
3.13 Summary 64
CHAPTER 4 PRESENTATION OF RESULTS 65
4.1 Data Analysis Plan 65
4.2 Participants' Demographics 66
4.3 Comparing task Efficiency of the systems 68
4.4 Comparing of task effectiveness of the systems 70
4.5 Comparing of user Satisfaction of the systems 72
4.6 Comparing of Cognitive of the systems 76
4.7 Overall Difference 77
4.8 Summary 78
CHAPTER 5 80
RESULTS DISCUSSION CONCLUSIONS, AND RECOMMENDATION 80
5.1 Introduction 80
5.2 Summary of the Findings 80
5.3 Task Efficiency 81
5.4 Task Effectiveness 82
5.5 User Satisfaction 83

5.5.1 Efficiency 83
5.5.2 Affect 83
5.5.3 Helpfulness 84
5.5.4 Control 84
5.5.5 Learnability 84
5.5.6 Cognitive Workload 86
5.5.7 Contribution of Study 87
5.5.8 Implications of the Study 87
5.5.8.1 Usable expert System 87
5.5.8.2 Train users 88
5.5.8.3 Cognitive overload avoidance 89
5.5.9 Limitations of the Study 89
5.5.10 Directions for Future Research 90
5.6 CONCLUSION AND RECOMMENDATIONS 90
5.7 REFERENCES 92
5.8 APPENDICES 98
5.8.1 Appendix 1 Research Permit 98
5.8.2 Appendix 2 Software Usability Measurement Inventory Questionnaire (SUMI) 99
5.8.3 Appendix 3 NASA Task Load Index (NASA-TLX) 102
5.8.4 Appendix 4 Farmers Consent for Research Participation 103
5.8.5 Appendix 5 Journal Paper (Accepted for publication in: International Journal of Information and Electronics Engineering (IJIEE), 2017) 106
IIX REFERENCES 111

LIST OF TABLES

Table 3.3.1: Independent and Dependant 52
Table 3.8.1: Research Questions and Hypothesis 58
Table 4.3.2 Mouse Clicks 69
Table 4.3.3 Key Strokes 69
Table 4.4.1 Errors Made 70
Table 4.4.2 Completed with ease 71
Table 4.4.3 Completed with difficulty 71
Table 4.4.4 Failed Completely 72
Table 4.5.1 :System Efficiency 73
Table 4.5.2 System Affect 74
Table 4.5.3 System Helpfulness 74
Table 4.5.4: System Controllability 75
Table 4.5.5 System Learnability 75
Table 4.5.6 Systems in General 76
Table 4.6.1 Cognitive workload /Mental Effort 76

LIST OF FIGURES

Figure 1.1.1: Expert System 15
Figure 2.2.1 Usability Factors 28
Figure 2.3.1Efficiency 30
Figure 2.4.1Mayer's Cognitive Theory of Multimedia Learning 32
Figure 2.4.2 Cognitive Workload 41
Figure 2.4.3 Measuring Cognitive load 42
Figure 2.4.4 Comparison of Text-based, Standard Multimedia, Multimedia Simulation learning 48
Figure 3.9.1 How data was collected 60
Figure 4.2.1 Participant Gender 66
Figure 4.2.2 Type of farmers 67
Figure 4.2.3 Age Range 68
Figure 4.3.1 Average Time taken to complete Tasks 68
Figure 4.7.1 :Overall Comparison of the System 78

LIST OF ABBREVIATIONS/ SYMBOLS

VR	Virtual Reality
BIUST	Botswana International University of Science and Technology
NASA TLX	National Aeronautics and Space Administration (Task Load Index)
SUMI	Software Usability Measurement
3D	Three Dimensional
MOA	Ministry of Agriculture
BIUST	Botswana Intentional University of Science and Technology
GDP	Gross Domestic Product
Vet	Veterinary
ISO	International Standards Organisation
Music	Measurement of Usability in Context
CTML	Cognitive Theory of Multimedia Learning
CLIPS	C Language Integrated Production System
SUS	System Usability Scale
UI	User Interface
ES	Expert System

α	Alpha
m	Mean
s^2	Variance

CHAPTER 1
Introduction

1.1 Introduction of the Research

An expert system is a knowledge-based information system that uses its knowledge about a specific area to act as an expert consultant to users [1]. An expert system is usually used for advising a non-expert when a human expert is not available, in our case a farmer is advised by an expert system in the absence of a veterinarian. An expert system has three components which include user interface, inference engine and the knowledgebase [2].

User interface: - This is the component which allows end users who are non-expert to communicate with the expert systems by sending questions, the advices are then received through the same user interface. Our study is focused on this component of the expert system, comparing text and VR-based interfaces to check which of the interfaces is useable. Establishing which interface is effective, efficient, satisfying and uses less mental effort. If the expert system is not usable it will defy the purpose of the expert system.

Inference engine: - This is where the questions (queries) that are sent by the non-expert are examined and matched with information provided by the human expert in the knowledgebase as indicated in the Figure1.1.1 below.

Knowledgebase: - Facts and rules derived from the information provided by the human expert are collected here.

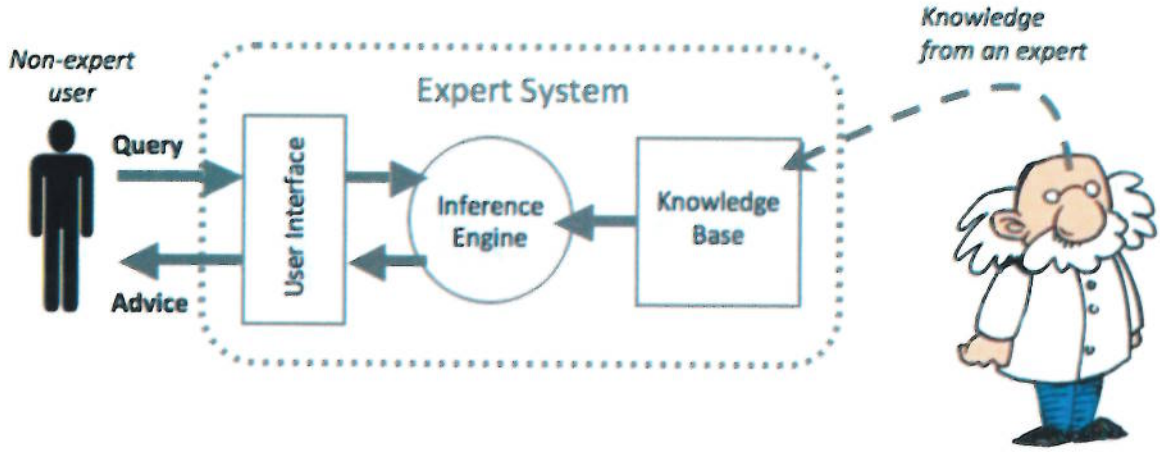


Figure 1.1.1: Expert System [2]

Artificial intelligence technology is growing at a rapid pace and it is being applied to a wide spectrum of practical problems including cattle disease diagnosis [3].

On the other hand, virtual reality incorporates animated 3D models for identifying symptoms and infections. Virtual reality is defined as " an artificial environment that is created with software and presented to the user in such a way that the user suspends belief and accepts it as a real environment" [4]. On a computer, virtual reality is primarily experienced through two of the five senses; sight and sound. When most people think of virtual reality, they often think of full emersion where by people wear head gears that display full visual and special gloves which will allow the user to use almost all the six degrees of input and manipulate the environment. On the contrary there is also another form virtual reality known as non-immersive which is cheaper compared to immersive virtual reality which uses expensive equipment such as head mounted displays, data gloves, motion trackers, domes, and 3D controllers. Non-immersive virtual reality environment is manipulated by using a monitor keyboard and a mouse [5]. This study uses non-immersive virtual reality.

Usability of non- immersive virtual reality based diagnosis expert systems has not been investigated in contrast to text based diagnosis expert systems. Most of the studies conducted focused on comparing text based expert systems and multimedia

based systems that use images and videos. The main focus of the study is on the four aspects of usability which includes:

- Efficiency: This is how the user feels about systems responds when tasks are performed through it which includes aspects such as time taken to complete a task, resources used to complete the task by means of the systems performance [6], [7].
- Effectiveness: is defined as the accuracy and completeness with which users achieve specified goals [7].
- Satisfaction: "User satisfaction refers to the user's comfort and acceptability of a computer application during the consumption of the content and the interaction with the system" [8].

Cognitive Workload: is defined as the total amount of mental effort and mental load that a particular task impose on the human's cognitive system at an instance in time [9] .

In developing a software application, user acceptance of that particular application is very essential hence the application has to be usable. The first thing that needs to be considered is the nature of the interface of the application and its correctness in order to get the expected output. Ignorance of these aspects can lead to an application being of poor quality hence making it meaningless to its users. As a result an application, prior to its release to its intended users, needs to be assessed and evaluated. The methods used to evaluate usability in this study include video based observation using Morae recorder usability evaluation platform developed by TechSmith to evaluate effectiveness and efficiency. Software Usability Measurement Inventory (SUMI) method for evaluating satisfaction and the NASA Task load Index (TLX) scale was used to evaluate the mental effort invested by farmers while using the two expert system [10], [11], [12], [13].

If a usable cattle diseases diagnosis expert system is used, it could help farmers in remote areas to diagnose their cattle diseases on their own. If there is a need to consult a veterinarian (for instance, if the system does not recognize the disorder and there might be a new outbreak) the system will advise the user on what to do before consulting a veterinarian. Farmers can also to learn about common cattle diseases in their areas.

1.2 Overview of Cattle Farming in Botswana

Most Batswana own cattle for economic purposes. Furthermore, the cattle farming represents the third largest income of Botswana's economy after diamonds and tourism contributing significantly to the GDP [14]. In 2013, the beef exported by Botswana was worth US$116.6 million contributing a major share to the economy [15]. As cattle farming is a major contributor to the economy, there are some factors which negatively affect this sector such as diseases. This increased the demand on animal health services such as disease diagnosis, getting advice on different cattle diseases and learning about different cattle diseases [15], [16].

In recent years there has been a major decline in Botswana's beef industry which is due to shortage of veterinary physicians (V.E.T.S). Botswana currently has only 55 V.E.T.S who are responsible for a cattle population of over two million cattle. To service such a large population cattle requires at least 130 veterinarians. This means 75 more veterinary officers are needed [17]. When giving advice to the Parliamentary Special Select , which was responsible for handling the decline in Botswana's beef industry, Dr Mapitse [17] indicated that the number of supporting staff should be reduced from 759 to 416 in order to create vacancies for veterinary officers [17]. Expert systems could be used to overcome the problem mention above. An expert system is "an automated reasoning system that attempts to mimic the performance of the human expert " [18]. This type of expert systems is the one which this research study is focused on.

1.3 Cattle Diseases Management Systems

There are very few cattle diseases management systems in Botswana which are used to disseminate information, for instance Modisar which is a livestock management application developed by Modisar Net [19]. According to the developers of Modisar, the main reason why they developed this system was to tackle the challenges faced by farmers on their relentless effort to feed the nation. These problems included drought which occurs almost every year, outbreaks of parasite and diseases, lack of proper management tools and unavailability of important information on farm management. Modisar farm management has an animal health management system consolidated in it which only gives farmers health advice [19]. Modisar animal health management system cannot be used to diagnose cattle , it just gives a generic advice on some cattle disease, this has resulted in farmers having to rely on veterinary officers, as we have seen that the ratio of veterinary officers to cattle population is very low [15].

In summary, there is evidence that an expert system which could be used on behalf of the veterinary officers is needed. Firstly, it evidently showed that there is a tremendous shortage of veterinary officers in Botswana's beef industry. Therefore, there is a need for an intervention, and one way to mitigate this problem is by introducing an expert system which would mimic the performance of veterinary officers. Secondly, the existing systems like Modisar, offers generic advice on some cattle diseases and do not educate cattle farmers step by step on what to do prior to taking the animal to a veterinary officer. For the reasons mentioned here there is a need to innovatively come up with a usable, effective and efficient cattle diagnosis system.

To address the challenges of limited platforms which Batswana cattle farmers can use to diagnose cattle diseases and advise them on what to do when they encounter those diseases, two expert systems were developed. These two expert systems performed the same tasks which included "Task 1: Starting up the systems, Task 2: Selecting components from menu, Task 3: Choosing the signs and symptoms, Task

4: Getting advice and lastly Task 5: Exiting and saving results". The only differences they had was interfaces which were a text based and non-immersive virtual reality based. A comprehensive comparison of the usability of these two expert systems was conducted so as to determine which one would be suitable within the context of the Botswana environment. According to Carroll [20] the focus of the development process of systems in the twentieth century is user-centred and usability is taken as a primary goal. Baven [21], [22] states that usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" .The study was focused on the evaluation of the three components of usability which includes effectiveness, efficiency and satisfaction and the extra component which is the cognitive work load experienced by farmers when using both systems. The evaluation results collected from the two systems were compared.

1.4 Diagnosis Expert Systems in Agriculture

There has been a number of diagnosis expert systems developed in recent years all with different interfaces both traditional (text) and multimedia based. Even though most of these expert systems were not necessarily for diagnosing cattle diseases they used the same concepts as the two systems being evaluated and they were used in agriculture. The following are some of the agricultural diagnosis expert systems that were reviewed: CUPTEX, Wheat, Pig-Vet and AMRAPALIKA.

Expert systems such as CUPTEX which is text based and developed with Prolog was used for cucumber production management under plastic tunnel [23]. This expert system was developed by the Egyptian Ministry of Agriculture (MOA). Domain experts were given the prototype of the system to test and their results revealed that the system was usable. End users were trained to use the shell based interface.

Wheat is a Web-based Expert System for Diagnosis of Diseases and Pests in Pakistani, the interface of these expert systems is multimedia based in which signs and symptoms of plant diseases are shown with images with a description of what

the plat may be suffering from. The users of this expert system indicated that it is easy to use [24].

Pig-vet is a web-based system developed by an Agricultural University in China. Before rolling the system out, it was first piloted in north of China [25]. The system had about 202 images and graphics for different types of diseases and symptoms which means a multimedia interface was used in the system. It also gave users an option to use a text only interface if they wanted to. This expert system was usable according to what the results show.

AMRAPALIKA just like other expert systems mentioned above, this expert system also use image as their multimedia tool for showing symptoms and infections. This expert system is used for the diagnosis of pests, diseases, disorders in Indian mango [26].

This literature review was conducted to provide a foundation of the research study carried out on usability testing of the two expert systems being compared. The review of the literature is categorized into five sections. Firstly, the literature will look at usability testing of software applications. Secondly, the methods used to compare the systems. Thirdly, the cognitive theory of multimedia learning is discussed which is essential for the effort exerted by users when using a particular system. Fourthly, the literature will look at a usability testing of desktop application particularly agricultural expert systems for diseases diagnosis. Lastly on the fifth section, the literature is looking at an overview of existing diseases diagnosis expert systems both text and virtual reality (multimedia) based. With the literature review some gaps where identified which drove this study.

The reviewed Literature was collected from different sources such as the educational scholar reports, journal publications and web pages.

1.5 Problem Statement

Despite an increased interest in usability testing of software applications, so little empirical researches has been conducted on this area of study. There is lack of empirical research conducted which compares the usability of non- immersive virtual reality and text based diagnosis expert systems for cattle diseases in Botswana [27]. Furthermore, the reviewed literature has indicated that cognitive workload which is often overlooked is a very important aspect of usability [28], [29].

The problem that this study addresses is that usability of non- immersive virtual reality based diagnosis expert systems has not been investigated in contrast to text based diagnosis expert systems [27], [28], [29].

This is a clear sign that there is a need to intervene and help the farmers. So with these deficiencies the researcher found it keen to conduct a study on comparison of the two expert systems.

1.6 Main Objective

The main objective of this study is to compare the usability of non-immersive virtual reality and text based diagnosis expert systems for cattle diseases looking at the effectiveness, efficiency, satisfaction and cognitive workload experienced by farmers when using the system, and then recommends a usable expert system for developers.

1.6.1 Specific Objectives

1 To compare the task **efficiency** of the text-based and non-immersive virtual reality based diagnosis expert systems.

2 To compare the task **effectives** of the text-based and non-immersive virtual reality based diagnosis expert systems.

3 To compare the user **satisfaction** of the text-based and non-immersive virtual reality based diagnosis expert system.

4 To compare the **cognitive workload** exerted by farmers using the text-based and non-immersive virtual reality based diagnosis expert systems

5 To recommend a useable cattle diseases diagnosis expert system for software developers.

1.7 Research Questions

1 What are the differences of **efficiently** completing the tasks between farmers using the text and multimedia based diagnosis expert systems?

2 What are the differences of **effectively** completing tasks between farmers using text and multimedia based diagnosis expert system?

3 What are the differences in user **satisfaction** between farmers using text and multimedia based diagnosis expert system?

4 What are the differences of **cognitive workload** experienced by farmers using text and multimedia based diagnosis expert systems?

1.8 Hypothesis

All the four research questions above have hypotheses that need to be tested. These hypotheses were used to measure the human computer interaction components which were efficiency, effectiveness and satisfaction. The other hypotheses was used to measure different cognitive workload experienced between users when using a traditional text-based cattle diagnostic expert and a multimedia based expert system.

1. H1: The text and non-immersive VR-based expert systems will have a significant difference in task efficiency in terms of time taken to complete the task and mouse clicks.

2. H2: The text and non-immersive VR-based expert systems will have a significant difference in task effectiveness in terms of numbers of errors made, task completed with difficulty, task completed with ease.

3. H3: The text and non-immersive VR-based expert systems will have a significant difference in user satisfaction in terms helpfulness, system control, affect and efficient.

4. H4: The text and non-immersive VR-based expert systems will have a significant difference in user cognitive workload in terms of metal effort invested by the users.

1.9 Expected Outcome

At the end of this research, a usable veterinary medical diagnostic expert system which will assists with self-diagnosis and learning of cattle diseases by farmers prior to official laboratory tests would be identified. Then a recommendation to software developers will be done from either the VR or the text based system.

1.10 Justification of the Research

Currently there are limited studies conducted on usability evaluation of diagnostic tools in Botswana which cattle farmers can use to remotely diagnose their cattle or to advise themselves on what to do as a first response in case of an outbreak. In a meeting held by young farmers association at Ministry of Agriculture on September 20 2014, Mapitse [17] stated that veterinarians charge within a range of P 500 to P 700 per hour which is too high for subsistent farmers, and the government field agents are not enough to serve all the farmer s hence they take a long time to respond.

According to James [30] an interface for diagnosing expert systems integrated with multimedia showing symptoms and infections with images, is more advantageous and much easier to understand than when using textual description, which is the

case with most of the expert systems developed so far. This would be much easier to understand when using virtual reality as the farmers would be able to manipulate the animated 3D avatars and confirm that indeed those are the symptoms they are seeing on their cattle. The virtual reality Animated 3D avatars simulations are also able to mimic the sound made by a sick animal, that is why this research study is integrating virtual reality into the expert system. A high level of presence is achieved by using VR simulation and with that farmers make the right decisions when diagnosing their cattle.

1.11 Research Delimitation

This research is mainly focusing on the comparison of an integrated rule based interactive expert system for diagnosing cattle diseases with virtual reality for better or easier way of identifying signs and symptoms with a traditional text based expert. Data collection will be done in Palapye and surrounding areas including Tuli block due to unavailability of funds. According to Cheon and Grant [31] the design of interfaces cannot be separated from usability testing which although an established field of study lacks universally accepted benchmarks, standards, and best cases. There are no usability laboratory in BIUST which gives us a limitation on the kind of experiments to be carried out thus meaning the one that are complex could not be carried out .This research is a proof of concept which means it is a realization of a certain method or idea to demonstrate its feasibility, or a demonstration in principle, whose purpose is to verify that some concept or theory has the potential of being used.

1.12 Contribution

As of today, there has been very little detailed empirical research conducted which compares the usability of non-immersive virtual reality and text-based cattle diseases diagnosis expert systems. After reviewing the literature it was found that most of the studies conducted were focused on expert systems for diseases diagnosis in agriculture both plants and animals but not comparing their usability. So this study seeks to address this knowledge gap.

1.13 Thesis Structure

Chapter 1: Introduction

This section introduces the problem that this study is addressing and it gives the reader a direction on the focuses of this study. It leads the reader to the main objective, specific objective, research questions and the hypothesis and the contribution of this study.

Chapter 2: The literature review

On this section the reader is given an insight of what the study is addressing through literature review taken from various sources such a journal articles, conference proceeding papers, books and dissertations conducting similar studies on usability evaluation and its measures and cognitive theory of multimedia learning and well modality principle.

Chapter 3: Method

On this chapter the methods used to carry out the experiments of this study are described. It highlight on the tools and the concepts which were used in order to get the necessary results of this study. It also covers the data methods used in this study.

Chapter 4: Results presentation

This chapter presents the acquired findings from the experiments of this study, to answer the research question and the hypothesis.

Chapter 5: Discussion and Conclusion

The results of this study are analysed on this part of the thesis. What has been deduced from chapter 4 is explained here. This part also covers the results implications, what has been learned from the results, do the findings agree with previous research, their strengths and weaknesses and the directions for the future research. Lastly, the chapter summarises by explaining what conclusions can be drawn from the findings.

CHAPTER 2

Literature review

2.1 Introduction

For a better understanding of the issues that surrounds the current study, a thorough review of the literature was conducted. The literature review conducted relates the issue at hand, which compares the usability of non-immersive virtual reality and text based diagnosis expert systems for cattle diseases in the context of effectiveness, efficiency, satisfaction and cognitive workload exerted by farmers when using the system. Furthermore, the gaps identified in the previous literature that motivates this study are alleviated to explore usability aspects of the compared expert systems.

This chapter has three sub sections. The first subsection section 2.2 gives a brief discussion on how the usability of different applications can be tested. The second subsection entails the methods used to compare the usability of the systems, and the measures of effectiveness, efficiency and user satisfaction. The third subsection is a brief discussion of cognitive theory of multimedia learning which provides the basis for understanding cognitive workload or mental effort.

2.2 Usability Testing of Applications

According to [32], [12], to reliably measure overall usability one should assess effectiveness, efficiency and satisfaction with which the representative users will carry out different tasks in different environments. Furthermore, for one to be able to measure usability they need a deep understanding of the context in which the product is going to be used. For instance one need to understand how farmers diagnose their cattle for different diseases when using cattle diseases diagnosis expert systems both virtual reality (multimedia) and text based. With the help of the tools developed by the ESPRIT MUSIC project the research was able to measure usability both in the field and in the small usability lab that was setup using Morae recorder, observer and manager.

According to Bevan and Macleod [32], [33] they consider usability to be a quality of use in a context. Moreover they also take usability as a property of an overall system. In our case quality of use is measured as the outcome of interaction within a context. It also considers the extent to which the intended goal of use of the overall systems is achieved which covers effectiveness. It also considers the resources invested to achieve a specific task such as time and money which covers efficiency. Furthermore, they looked at the degree at which the user finds the overall system good enough or the comfort and acceptability of use which covers satisfaction. Bevan and Macleod also state that equipment, tasks, users both physical and organizational environment that indirectly and directly influence the interaction make up the overall system. The ISO 9241-11 (ISO 1993c) standard define usability as "The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environment" as shown in figure 1. [34], [21], [35].

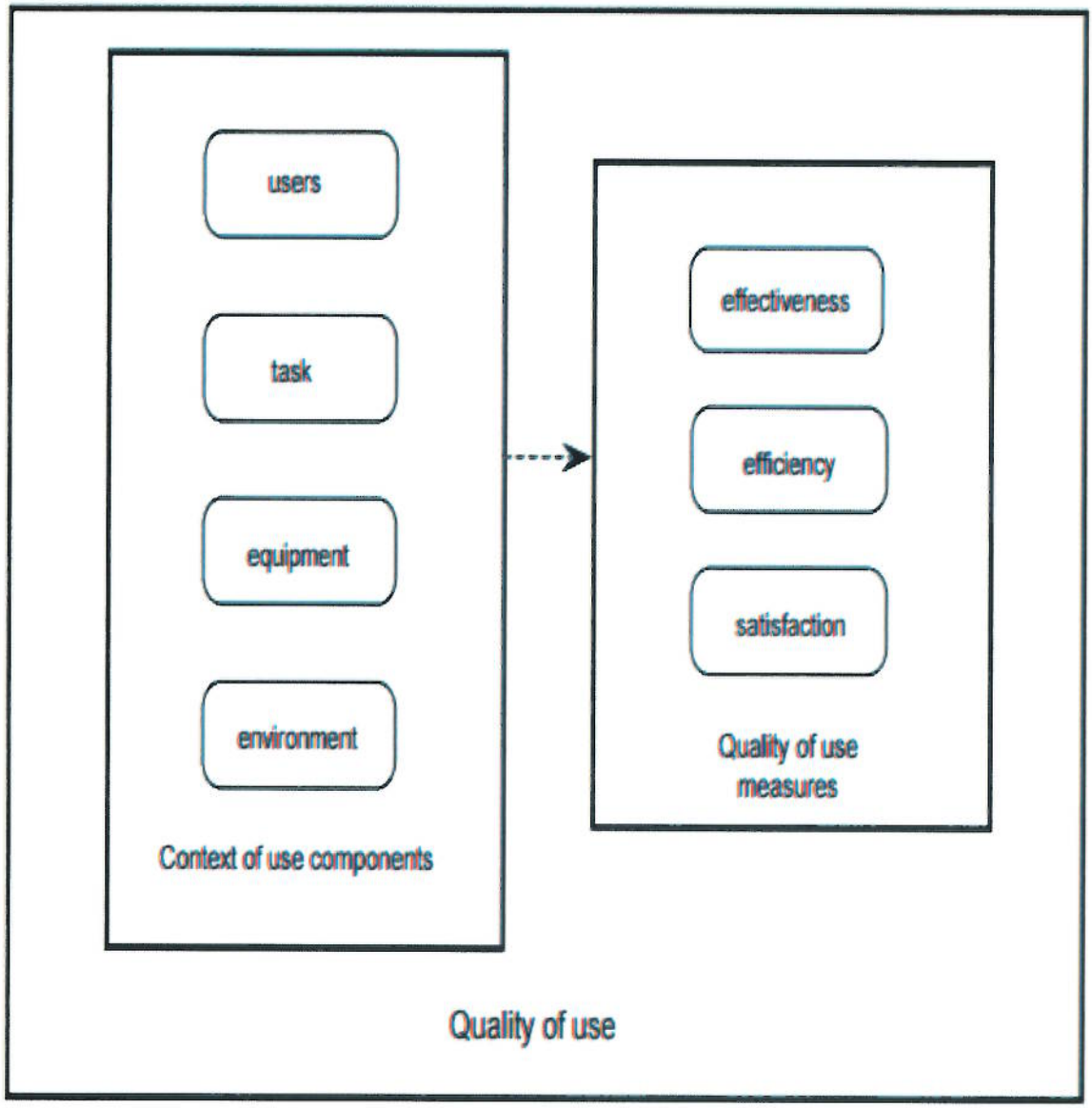


Figure 2.2.1: Usability Factors [34]

2.3 Methods for Comparing the Systems

According to Bevan and Miles [32] for one to say a product is of good quality it should consist of suitable measures of user performance which include effectiveness, efficiency and user satisfaction. Due to the fact that the components of usability usually are dependent on the context of use as well the purposes for describing usability there are no set principles on how the combination of measures should be like. So we need to provide a measure for all the components of quality of use. The measures should be repeated in different context [32].

Using MUSic Performance Measurement Method to measure the usability of a product results of the group of users that has same or related characteristics are usually added together to provide a mean usability measures [32]. Bevan and Miles [32] indicated that even when using 10 users when there is a major difference in the characteristics of users and there will be a lager variance related to the mean values.

2.3.1 Measures of Effectiveness

Measures of effectiveness are described as the goals or sub-goals of using the system to the accuracy and completeness with which these goals can be achieved [32]. The components of effectiveness are measured by the quality of the solution given by the systems to users. The effectiveness of the expert system was measured by the number of sub-tasks completely resolved and the number of errors made by the user.

2.3.2 Measures of Efficiency

Bevan and Miles [32] argues that measures of efficiency are usually linked to the level of achievements accomplished in effectiveness, particularly looking at the expenses of the resources used. The resources in our case referring to mental effort that could be used to provide the measures of an individual's efficiency, time that can be used to give the measure of temporal efficiency [32].

Looking at the users of the two expert systems text-based and multimedia based, the time they take to complete a specific task is measured, and the effort and the resources needed for completing a specific task is also measured. The resources create two different definitions of efficiency as shown in figure 3 below.

From a user's perspective, the time he or she spends carrying out the task, or the effort required to complete the task are the resources he or she consumes. These two types of resources produce two different definitions of efficiency:

$$\text{Temporal Efficiency} = \frac{\text{Effectiveness}}{\text{Task Time}}$$

$$\text{Human Efficiency} = \frac{\text{Effectiveness}}{\text{Effort}}$$

Figure 2.2: Efficiency [32]

2.3.2.1 User Temporal Efficiency

The performance measurement method, of the temporal efficiency of a user according to Bevan and Miles [32] is defined as:

$$\text{User Efficiency} = \frac{\text{Effectiveness}}{\text{Task time}}$$

Figure 2.3: User Temporal Efficiency [32]

Efficiency is said to be measured in different context, the values compared were of a great meaning, in our case two cattle diagnosing expert systems, also used as a learning tool were compared. These expert systems were used by the same user group which are farmers in the same environment.

2.3.3 Measuring User Satisfaction

According to Bevan and Miles [32] measures of satisfaction describe the perceived usability of the overall system by its users and the acceptability to the system to the people who use it and to other people affected by its use Measures of satisfaction could be directed to some specific modules in the systems or the entire system.

User satisfaction of the systems was measured using the attitude scale known as Software Usability Measurement Inventory (SUMI) which is a questionnaire with 50 questions which are answered through either "agree", "disagree" or "Undecided" as shown in (Appendix 3). Measures of satisfaction provided very useful indication of the user's perception towards the usability of the system. SUMI can be used in to evaluate any system as long it has user inputs through the keyboard, pointing device

display screen and secondary memory [36], [37]. SUMI gives a global usability figure along with five subscales which includes efficiency, affect, helpfulness, control and learnability.

2.4 The Cognitive Theory of Multimedia Learning

On this sub-section the researcher reviewed the constructs of the cognitive theory of multimedia learning so as to clarify and show its significant contribution on this thesis. The research is based on a comparison of two interfaces one with multimedia components (virtual reality) and the other one is just a traditional text based interface. So it is very important to lay out a foundation of cognitive theory of multimedia learning and its theoretical context in which users (farmers) can be investigated in the amount of mental effort they invest when using the expert systems. According to Mayer [38], [39], [40] this theory is addressing the issue of how to structure multimedia instructional practice and use effectively the cognitive strategies in order to assist people to learn more effectively. The first part of this sub-section will cover an overview of cognitive workload or mental effort. Moreover, this sub-section continues to articulate the research question used in collection and the analysis of data. Lastly, this sub-section discusses how data for this research is analysed using cognitive theory of multimedia learning as an analytical construct.

2.4.1 Human Information Processing

There are quite a number of studies that has been conducted on cognitive structures which forms part of human information processing [40]. The cognitive structures which these studies were focused in included the sensory memory, working memory and long-term memory and for the cognitive theory of multimedia learning which the study is focusing on has the third structure added which is multimedia presentation. The sensory memory is responsible for holding a copy of whatever is being presented for less than 25 seconds, and also used in preserving incoming information which could be either visual or auditory information. The working memory is used for selecting the information for being processed and integrated from the sensory memory [40], [9]. The long term memory is the one that stores the whole

knowledge and skill of a person for unspecified period of time in a hierarchical manner and its volume or size is unlimited as shown in the figure below.

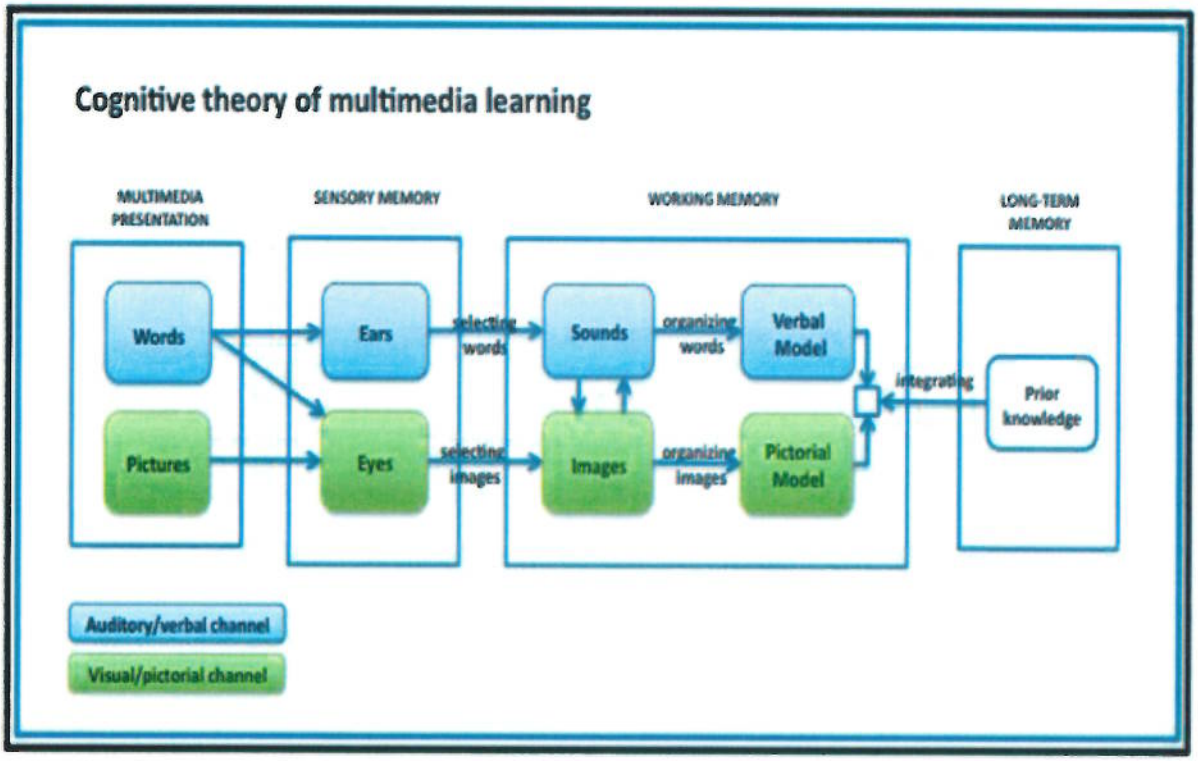


Figure 2.4: Mayer's Cognitive Theory of Multimedia Learning [40]

Mayer states that the cognitive theory of multimedia learning (CTML) is based upon three science principles of learning which includes the human information processing system consists of dual channel, limited capability assumption and active processing [40].

Mayer [40] argues that people learn deeply from words and pictures than from words alone, but he insists that adding pictures to words is not a guarantee that there will be an improvement because multimedia presentations are not equally effective, for instance images (static graphics) cannot be more effective than animations (dynamic graphics) because animations display more features such as movement and facial expression [40].

According to Mayer [38] learning can be measured by the test of retention such as being able to remember what you have learned, and it can also be measured by the ability to use new information to solve new emerging problems. The focus of this study is on transfer because the investigation is how using virtual reality and words instead of words only can promote understanding [39]. Mayer continues to argue that transfer test can be used to give us an insight of how well people understand what they have learnt [41]. We are also looking at the cognitive process which the farmers use to build meaningful learning outcomes from text and animated characters other than when using text only.

The cognitive theory of multimedia learning consists of the following elements that are as follows; firstly the three assumptions of cognitive theory of multimedia learning, secondly the three memory stores, thirdly the five cognitive process and the forth element is the five forms of representation in theory.

2.4.2 Three Assumptions of Cognitive Theory of Multimedia Learning

Mayer explores three assumptions which are essential to the cognitive theory of multimedia learning which includes dual channel assumption, limited capacity assumption and active processing assumption.

2.4.2.1 Dual Channel Assumption

This assumption is that human beings process information using two different channels which are audio (verbal) or virtual (pictorial) represented materials [41]. The differences between these channels are conceptualized by two ways which are based on presentation modes which mainly focuses on whether the stimulus is verbal (words being spoken or written) or non-verbal (animations pictures or video) and in contrast sensory modalities focuses on whether the presented materials is processed by ears (spoken words, sounds on the background) or through the eyes (animations, pictures or presented words). Novice farmers are able to devote adequate cognitive resources to the task being performed, it is possible for information originally

presented to one channel to also be presented in the other channel hence increasing a chance of user to understand the presented materials much easier [40].

2.4.2.2 Limited Capacity Assumption

This assumption argues that people have a limitation in the amount of information they can process in a single channel at a time. Most researches according to Mayer has shown human processing capacity is very limited. Meta cognitive strategies which are the techniques for allocating, monitoring, coordinating and adjusting the limited resources, helps in understanding human intelligence [42]. These techniques could help governing the novice users not to exceed their limited memory capacity.

2.4.2.3 Active Processing Assumption

With this assumption the humans are engaged in active learning by giving attention to the incoming information which is relevant, then by organizing that incoming information into a coherent mental representation then that mental representations will be integrated with other knowledge [42]. Active learning involves three main processes which are selecting the relevant material, then organizing the selected materials and lastly getting the selected materials and integrating them with existing materials [42]. Hence this would allow the farmers to actively learn about different diseases of their choice in a proper and more effective way without actually exceeding their cognitive capacity.

2.4.3 The Three Cognitive Processes

Mayer [39], [43] has also indicated that when learning and considering cognitive theory of Multimedia learning there are three types of cognitive processes which includes extraneous, generative and essential. The processes are discussed in detail in the section below.

2.4.3.1 Extraneous Cognitive Processing

According to Ginns [43] instructional messages are usually presented by multimedia by different educators. These educators habitually neglect some cognitive structures and processes when designing the multimedia presentations hence these leads the instructional messages not clearly understood and thus results in learners learning a little or nothing from that multimedia presentation [43].

When the leaning materials do not consider the maximum limit or the capacity of the working memory then extraneous cognitive leaning process occurs, this is usually cause by a poor design of the multimedia presentation [43]. For instance in our case when the animated avatar tries to show lumps (lumpy skin disease) on the skin of a cattle and it is not clear, there is unnecessary music on the background and some extra materials that has no direct contribution to the massage being displayed then extraneous cognitive processing would occur. If one forces the learner to process extraneous information that will demand a lot of cognitive load then the thinking capacity will not engage in proper leaning processes that includes selecting, organising and incorporating the leaning materials. This could also be caused by learners being confronted with a new problem, whereby they will have to look for solutions because they cannot find the solutions to that particular new problem.

2.4.3.2 Essential Cognitive Processing

The main aim of essential cognitive processing is to allow the learner to represent mentally the essential learning materials in the working memory [43]. If a learner selects then builds a mental representation with the learning materials which are essential then the essential cognitive processing will occur. The representation of the information is usually determined by the intricacy of the materials used to learn.

Mayer [43] confidently declared that for the management of cognitive processing one needs to use either one of the two techniques which are segmenting and modality in multimedia designs. Segmenting is whereby instead of presenting multimedia leaning material in one presentation that is continuous, it will be broken down into smaller

chunks which are easier to understand and manage. Whereas with modality leaning materials are presented using animations or picture with words at the same time instead of using words and pictures only. Modality is applicable in our VR-based expert system for instance when cattle are in pain from foot and mouth the sound it makes will be heard and the limping movements will be shown. Modality is further discussed in depth in the section below because it is adopted on this research project as our main focus.

2.4.3.3 Generative Processing

This processing is whereby the learners make a comprehensible understanding of the essential material, it entails putting words and images into a more clear and consistent structure, it then merges the word structure and the image structure together and any other knowledge [43]. Generative processing also facilitates the knowledge structures. If the multimedia learning material is done properly it will make a substitute of the prior knowledge and thus making it easier for learners to create a more effective schema. Generative processing is promoted by multimedia presentations which are using words and pictures rather than those using words only.

2.4.4 Five Forms of Representation in Theory

Mayer [41], [39] states that words and pictures can be represented in five forms when being processed by the memory. These forms are represented in each three stages of processing in the memory stores model of multimedia learning [39]. These five forms include firstly the form represents the element of multimedia itself which are words and pictures. Secondly the form is the acoustic (auditory) representation and the iconic (images) representation in the sensory image [39]. Thirdly the form represents sound and image in working memory. Fourthly there is the representation found in the working memory of pictorial and verbal models. Then lastly the fifth representation is prior knowledge, commonly known as schemas and is found on the long term memory.

2.4.5 The Five Cognitive Processes

For a farmer to have meaningful learning through words and picture tool Mayer contends that the learner should engage the five cognitive processes which are as follows [40]:

1) Selecting the relevant words for processing in the verbal working memory
2) Selecting relevant images for processing in visual working memory
3) Organizing selected words into a verbal coherent model in the working memory
4) Organizing images into a pictorial coherent model in the working memory
5) Integrating pictorial and verbal representations with each other and with prior (existing) knowledge

2.4.6 The Modality Principle

Throughout different research which has been conducted on multimedia presentation it is evident that when using words at the same time with pictures and animations, it is more effective and has helpful effects, this is formally known as modularity effect [44]. For a while now research on cognitive theory has supported this modularity effect which reduced the cognitive processes. Modularity effect suggests that people learn more effectively through a combination of word and pictures or animations rather than through text only [44], [45], [43].

Farmers are taught different cattle diseases using different methods by veterinary services. The commonly used methods include using text-based study guide and books which gives them detailed instructions on what to do if they come across a particular disease. The extension workers on different regions usually conduct some workshops to educate farmer s using posters with explained pictures [46]. Some farmers may also use online disease's diagnostic tools to learn about different and diagnose their cattle.

The main problem with the learning materials is that they are presented in textual instructions supported with either pictures or images. The methods mentioned above are usually done without the consideration cognitive load theory hence do not consider maximizing cognitive processes [44]. There have been quite a number of studies conducted on modality principles in different domains of learning [44], [47]. It has shown an improvement and good quality learning when presenting learning materials using images or animations and spoken words [44]. Nevertheless there has not been a study conducted to investigate thoroughly to see effectiveness of using spoken words together with non-immersive virtual reality (animations) showing instructions stepwise to learn about different cattle diseases and diagnose them using an expert system.

With multimedia the instructional messages are usually presented in two forms the audio mode and the visual mode which helps in maximising the cognitive processes and thus making learning more meaningful [47]. In modality effect it is clearly stated that if one presents words as a narration instead of using printed text only at the same time with pictures and images learning will be improved tremendously.

Ginns [44] carried out a meta-analysis in which he reviewed 43 independent studies and on that number 39 of them between-group experiments were then analysed. When conducting this meta-analysis major hypotheses for instructional benefits when a doing a presentation transversely on different modalities were supported and they includes pacing of presentation in between different studies and fields, the interaction of elements and performance.

The overall results throughout all the studies has shown that the modality effects is moderated at 0.72 which is its average effect size and thus bringing it closer to large. The results also indicates that for studies that used self-paced learning which is described as presenting the learning materials in smaller chunks or segments and the learner is given liberty to stop, pause or even skip a certain segment and the ability go back to the previous content if they want. Leaning materials can also be

presented in a system paced format which is whereby materials are presented in a continuous manner. It has evidently shown that system-paced materials have a larger effect size (0.93) than self-paced effects size (0.14).

The study showed that when comparing the presentation modality of immersive VR and computer screens and audio tapes the effect size of virtual reality head mounted displays (immersive virtual reality) which is recorded as 1.34 is greater or larger than that of computer screens (0.55) and audio-tapes (0.64).

The meta-analysis results indicated that graphic used with narration at the same time has larger modality effect as compared to the modality effects of learning materials which used printed words and text only. There were quite a number of domains reviewed in the meta-analysis which included engineering and science just to mention but a few. However it evidently showed that studies on examination of using multimedia particularly virtual reality to present learning materials using animated avatars or pictures/images at the same time with spoken words to teach farmer s about different cattle diseases presented as an expert system which could also be used for diagnosing those diseases.

Other scholars Dunsworth partnered with Atkinson [48] to conduct a research on evaluation of three effects which included the image effects, the modality effects, narration in caparison to text on screen, and embodied agent effect al used to present a cardiovascular system for humans. The outcome of this study was evaluated though retention, far transfer and near transfer questions for the evaluation of the learning out comes. There were 51 under-graduate students who participated in the study. The students were divided into three (by computer based learning conditions) groups of 17 members. The results indicated that the students who were reviewing narration effects and pedagogical agents scored better than those who were using on screen text with a large effect size of 0.80. The students who were also using narration and pedagogical agents performed better that those using narration only scoring a medium effect size of 0.66. This study concludes that

narration plus pedagogical had a higher score on the post-test than when using narration only. Nonetheless there a no statistical evidence that a great significance between this three conditions.

To conclude we can say that despite the fact that modality effect was re-used in the study curried by Dunsworth in partnership with Atkinson [48], there a quite a number of studies carried out which supported the use of modality effects in different disciples and different domains [48]. Since most of the research carried out on modality effect has indicated to be more effective in comparison to using text only for representation of learning content, this research was using multimedia in a form of virtual reality to present the learning materials to farmer s in two different ways. With the use of modality principle the learner (farmer) is able to diagnose and learn about cattle diseases by listening to interactions and seeing the signs and symptoms.

2.4.7 Cognitive Load

Cognitive load has been explored extensively by Sweller [49] . Cognitive load is defined by Sweller [49] as, "a total amount of mental effort being used in the working memory", whereas Paas and Merrionboer *et al* [50] define cognitive load as the multidimensional construct representing the mental load and mental effort that a certain task imposed on a trainee's cognitive systems at a particular time. According to De Jong [42] [9] learning can be stalled when cognitive overload occurs and when working memory capacity is exceeded.

The causal factors and assessment factors of cognitive load are differentiated, this distinction is done by looking at the factors that are affecting cognitive load and those that could be used to measure the cognitive load [51], [52]. Cognitive load can be measured using three different constructs that which includes mental load, mental effort, and performance. When assessing cognitive load through metal load the assessment is task centred which means it is based on the characteristics of the task [51]. Mental effort is known to be human-centred which means it considers the amount of capacity or resources which a learner could allocate in order to satisfy the

demands of the task [51]. Lastly this cognitive load assessment grounded on mental effort is said to be showing the quantity of engagement of the learner in a controlled processing [51]. As a result it is presumed that the reflection of the interaction between the learner and the characteristics of the task being learning. Cognitive load can also be accessed by the level of performance. According to Paas et al "similar learners, faster task performance with less effort can be considered to indicate a lower cognitive load than slower task performance with more errors" [51]. The diagram below (Figure 2.5) shows the factors that influence the level of cognitive load.

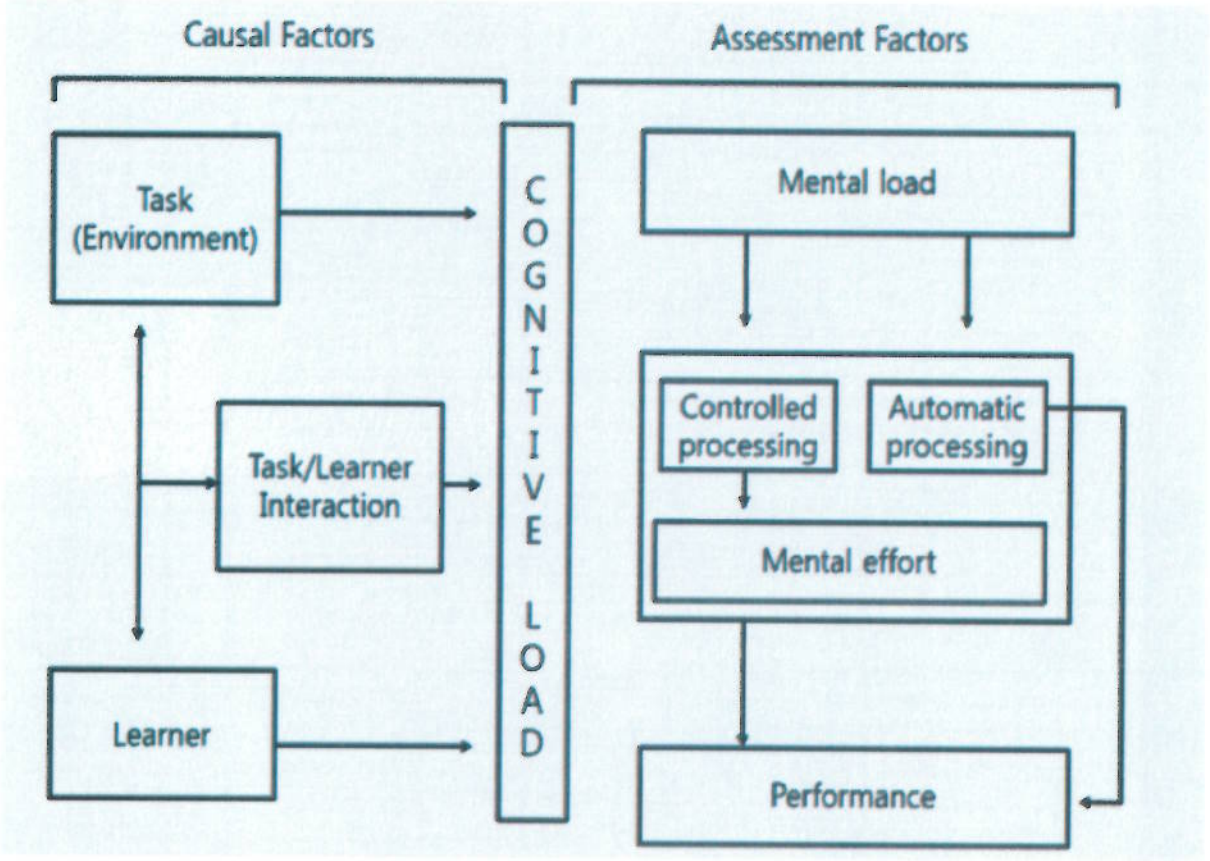


Figure 2.5: Cognitive Workload [50]

In this study the researcher described cognitive load as a mental effort and the demands which are placed on working memory resources. These mental effort demands includes firstly the task that is being done, secondly other tasks performed at the same time, thirdly disturbances caused by the environment when using the

system. Cognitive load is a very important aspect when it comes to learning and capturing specific information and materials to perform tasks diligently. So if the learning task surpasses the working memory capacity then it will lead to the cognitive being overloaded.

Cognitive load needs to be monitored and well a managed to avoid exceeding the working memory capacity which could lead to cognitive overload and by thus meaningful learning would be disturbed [50].

2.4.7.1 Measuring Cognitive Load of Multimedia Learning

Many researches has been conducted on measuring cognitive load of multimedia learning [50]. Paas, Tuoviane et al [50] indicated support for the claims that performing a particular task imposes specific cognitive load and to demonstrate that unpremeditated cognitive load is due to the interaction between the trainers characteristics and the task in hand [50]. This measures are done to ensure that user's memories are not overload or even under-loaded.

Classification of Methods for Measuring Cognitive Load Based on Objectivity and Causal Relationship

	Causal Relationship	
Objectivity	*Indirect*	*Direct*
Subjective	Self-reported invested mental effort	Self-reported stress level
		Self-reported difficulty of materials
Objective	Physiological measures	Brain activity measures (e.g., fMRI)
	Behavioral measures	Dual-task performance
	Learning outcome measures	

Figure 2.6: Measuring Cognitive Load [53]

As indicated in figure 2.4 the subjective measure method is whereby the cognitive load is indicated directly by self-reported stress level and indirectly by self-reported invested mental effort. There are quite a number of approaches that are used to measure subjective self-reported cognitive and they include; the NASA-TLX rating scale which was based on perceived task difficulty then later improved by Hart and Steveland [54], [13] to measure different aspects of cognitive load. Later on Deleeuw and Mayer [50] suggested a more improved version that indicated that the load indicators react differently in terms of sensitivity when different types of load are introduced to them [50].

The NASA-TLX indicated that participants are able to measure and rate their mental effort levels on their own [50]. Other scholars alongside Paas repeated the study and concluded that this subjective measure is more reliable and valid.

The other type of the indicator is the objective indicator as in indicated figure 5. In this type of cognitive load measure participants are not the ones that manipulate the cognitive load measure like in subjective measures. The measure could either be indirect which covers physiological measures (eye movements' behaviours, heart activity brain activity) these could be invoked by involuntary changes in human physiology [55]. There are is also behavioural measures or performance (e.g. completion times, error rates, test scores, mouse clicks, capturing mouse speed and pressure, key board pressing) which are commonly used in human computer interaction for observation of interactive behaviour [23], [56]. The other indirect measure is the learning outcome measures. There are also the direct measures which include Brain activity measure and dual-task measures [57].

2.4.7.2 Conclusion of Cognitive Load

In conclusion this sub section covered an overview of cognitive load which was one the objectives of this study. The literature has shown different ways in which ineffective cognitive load can be avoided in order to improve the learning process. The study focused on comparison cognitive workload (mental effort) experienced by

farmers when using the text based and virtual reality based expert systems for cattle disease diagnosis which covered by objective 4 research question 4 and hypothesis 4.

The participants on this study were visited at their farms (natural setting of a user) and some were randomly picked when visiting the Botswana government veterinary offices this was done so as to address the issue of external validity of results of cognitive load theory as indicated by different studies conducted on the premises of educational institutions conducive for learning. We could learn and understand with the help of these cognitive factors, which system between the two (virtual reality and text based) could be more effective and serve the purpose intended more productively.

Self-reported subjective measures of cognitive load were applied and investigated and compared to participants on their natural environment (cattle-posts and ranches at Tuli block) in this study. The next sub-sections review the related works.

2.4.8 Related Works

This chapter reviews existing literature on diagnostic expert systems with the view of identifying gaps in literature and addressing those gaps.

2.4.8.1 Review Existing Medical Expert Systems

Daoliang Li et al. [58] states that in China they developed a web-based expert system for fish disease diagnosis known as Fish-Expert to diagnose 126 types of diseases affecting 9 types of fresh water fishes. It has an image-base which contains images showing symptoms and infections. The farmers are said to have welcomed symptom identification method of using multimedia interface, and they found it more accurate to identity symptoms and infections by using images because of low level of literacy [6]. Similarly, Pig vet: a web-based expert system for pig disease diagnosis was also developed in china for diagnosing pig diseases using images for identification of symptoms because farmers were having some difficulties recognizing

symptoms and infections described textually [25]. Both of these systems collect the feedback from the users about problems faced by farmers, voicing their opinions on the usability of the system and suggesting how to make the system better.

Prasad et al. [26] states that AMRAPALIKA is an expert system for diagnosing common diseases for mangos in India. This expert system contains a database of pictures with infections and symptoms. The expert system is also comprised of knowledge base of diseases and their remedies. AMRAPALIKA was built using the ESTA platform, which is said to be more advantageous than CLIPS as it is easy to use because one does not need to have programming experience [26]. ESTA is suitable to use in many problem domains and structuring of knowledge. It has an explanation facility for questions asked and the given advice [13]. Similarly, DIARES-IPM which is a diagnostic advisory rule-based expert system for integrated pest management in Solanaceous crop systems in Greece has an integrated knowledge base [12]. This system also uses images as their multimedia tool for identifying symptoms and infections; DIARES-IPM is a standalone system.

Maize AgriDarksh [9] is ontology based advisory expert system developed in India used for diagnosing maize diseases and for identifying pests and then suggesting the prevention and control measures. The system also uses images as its multimedia tool for identification of symptoms and infections. This is a web-based expert system that allows the user to directly ask an expert some questions. Likewise, CUPTEX an expert system for diagnosing diseases affecting cucumbers was developed in Egypt [59]. Furthermore, Dr Wheat a web-based expert system for diagnosis of diseases and pests in Pakistan has the same features such using text to identify infections and systems [24]. The only difference it has is that it was developed using e2gLite™ expert system shell which uses java and with a problem domain of Pakistan [23]. The setback on this system is that it is text based it does not use any multimedia tool to display symptoms and infections.

An integrated expert system called Doctor Verdes was developed in Turkey [60]. This system is also a decision support system for diagnosing two vertebral column diseases (disk hernia or spondylolisthesis). It gives physicians a platform to discuss their patients and share ideas. It also uses different techniques such as machine learning and pattern recognition for diagnosis. The system has a share case page which physicians can share everything about their cases, such as history, examination, Pre-treatment images, diagnosis, selected treatment, or post-operative images including videos. It was developed using Visual Studio 2010 developer tool and SQL server 2010. It has five main components which are Diagnoses, Patient Operations, Share, Forum, and Help. It is multimedia based which means it uses picture and words.

Patient modelling Expert System for medical training simulations based on virtual reality known as UVIMO was developed in Spain [61]. This system uses virtual reality to model different patient behaviour. Artificial Intelligence (AI) has allowed the development of more believable virtual living beings which could be used to show symptoms and infections in a manner that it looks more real. Expert Systems have been used to continuously provide information about patient behaviour. In this way, a complex behaviour is achieved based on rules, which correlate with those of the expert. The patient simulations were modelled using CLIPS because it easily integrates with C++. High degrees of presence was attained on this system and trainers managed to acquire necessary knowledge when using the system [61].

A veterinary diagnosis expert system for remote use was developed in South Africa [30]. This expert system is used to diagnose illnesses in cattle and is accessible through cell phone web browsers. The system was build using drools rule engine and a web-based interface was developed for it. The system developed was a proof of concept. The system does not have any multimedia tool for identification of symptoms and infections. The system was specifically developed to address a problem domain of South African farmers [30].

2.4.8.2 Literature Review Conclusion for Existing Systems

The nine systems reviewed in this chapter address the problem domains of European and Asian countries and some of them use the native language of the host country (e.g. Pig-vet and Fish-Expert are using Chinese language) [58], [25]. The study also reveals that no attention has been given to the integration of expect systems and virtual reality for disease diagnosis. Furthermore most of these systems were developed to address one particular disease. Seven of the reviewed expert systems use images as a multimedia tool. According to Rafea et al [62] users are said to identify symptoms and infections much easier when they are represented by images than textual description. But the systems developed and studies curried out on this systems did not indicate whether they were effective, efficient, satisfying and the cognitive load or the mental effort invested by users were overload or under load. The next sub-sections review the Comparison of Multimedia and other Instructional methods.

2.4.9 Comparison of Multimedia and other Instructional Methods

Several studies on multimedia for educational and instructional methods have been conducted by different scholars [63]. In 1994 Petit [63] conducted a study comparing a text-based instruction, standard multimedia and multimedia simulating a setting of a mock [63]. Analysis of variance (ANOVA) was conducted on the scores recorded as indicated in figure 6 below. The results obtained from ANOVA indicated that there was a significant effect in the method instruction. The standard multimedia showed a much better original significant leaning in comparison to both text and multimedia simulating. As for text-based instruction and multimedia simulating there was no significant difference.

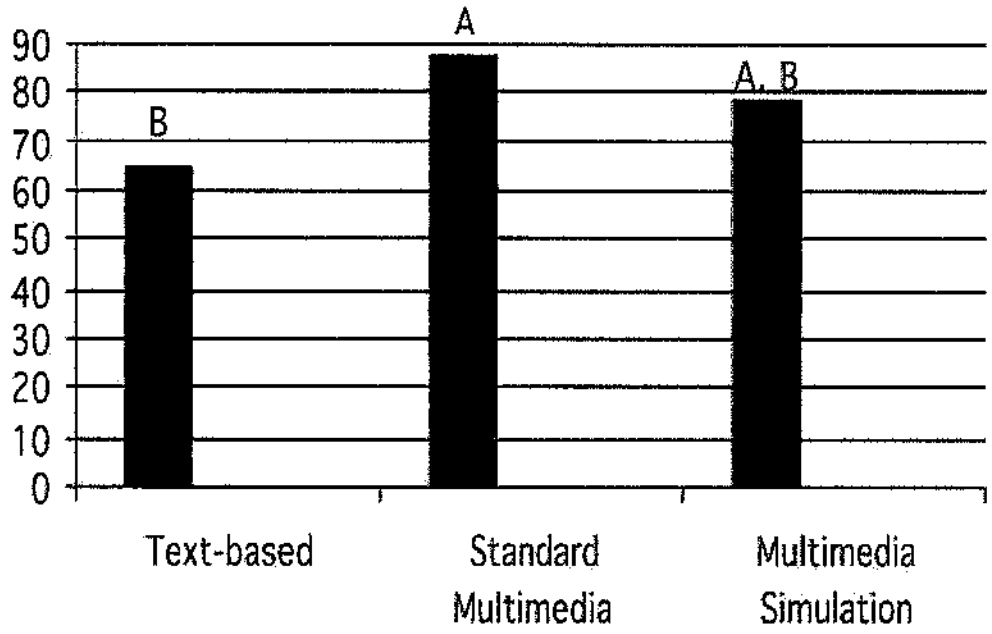


Figure 2.7: Comparison of Text-based, Standard Multimedia, Multimedia Simulation Learning [63]

According to Ramsey [63] there are some problems which he identified after conducting this study, for instance he indicated that subjects/learners felt a compulsion to concentrate on the story line presented by the simulation hence devoting less cognitive attention to the relevant instructional material relation to the time study and its analysis. Pettit indicates that to improve the simulation study the learner should be provided with a more challenging problem.

Another similar study was conducted by Au [63] he compared interactive multimedia (electronic book) with plain text version. The result obtained from the study indicated that the interactive multimedia systems have a significantly higher scored in comparison to a test-only based system. With the results obtained Au indicated that using new media such as interactive multimedia for delivering knowledge can be used to reduce performance gabs between student [63].

2.4.10 Summary

In both studies reviewed above comparing multimedia learning with different instructional learning such as text–only, the results for both studies indicated that the more multimedia learning systems involve the learners as active participants, the learner would be able to keep hold of the information presented [63].

There are very few studies that compare virtual reality (interactive multimedia component) with text based expert system for cattle diseases in Botswana. Therefore this study draws upon task effectiveness, task efficiency, task satisfaction and cognitive workload (metal effort) when using both virtual reality and text based systems. The next chapter presents the methodology used to address the research questions and the gaps identified in the literature chapter.

CHAPTER 3

Methodology

3.1 Introduction

In chapter 2 from the reviewed literature a number of flaws were identified that need to be studied that addresses. These issues could lead to a better understanding of strategies that could be laid down to come build expert systems for cattle diseases diagnosis in Botswana with good usability. The expert system has to be efficient, effective, satisfying and uses less metal effort (not overloaded or under-loaded) when used by farmers.

This chapter is aimed at giving a detailed foreground about the general research methodology used in this research, and the ways of a achieving the specific objectives by answering the research questions. The chapter covers the research design, participants and sampling, reliability and validity, analysis of data and data presentation.

3.2 Context of Study

This study was conducted between 2015 and 2016 in Palapye village in Botswana. Cattle farmers in and around Tuli Block ranches located 100 kilometres East of Palapye were used as participants in this study. Some cattle farmers using Palapye veterinary services were used as participants. This study was awarded a research permit by the Ministry of Agriculture to conduct the research (Appendix 2). The participants were using a laptop to test both the text based and virtual reality based expert systems and they were recorded with Morae recorder. Thirty two participants were recorded using the text-based expert system and the other Thirty two recorded using a virtual reality based expert system. The veterinary officers gave us permission to work in their premises and offered us assistance by referring the interested cattle farmers to come and test the systems. The farmers from Tuli block also gave us permission to give there interested employees to tests the system.

3.3 Research Design

The research design adopted for this study is a quasi-experimental research. According to Gray [64] quasi-experimental research design is usually used when the researcher instead of taking from the random samples they take an existing group. The main reason for choosing this research design is because it takes participants who are experienced in their own natural settings rather than just to manipulate an independent variable done in experimental research design. The behaviour of a similar group which has not experienced the event will then be compared with a group which has not experienced the phenomena; in our case comparing participants who have used text based systems with those who have used a virtual reality based expert system [64]. The other reason why this research design was adopted is because it supports the use of hypotheses which could either be supported or rejected. There was only one dependent variable for this study which is the usability of cattle diseases diagnosis expert system. Usability was defined by four constructs effectiveness, effectiveness, satisfaction and cognitive workload (metal effort). This study has one independent variables expert systems and it is categorised into two (text-based expert system vs. VR-based expert system) and five independent variables which were also categorised by different metrics as indicated in table 3.1 below.

Table 3.3.1: Independent and Dependant Variables

Independent Variables	Dependent variable	Hypothesis
Text-based expert system vs. VR-based expert system	Effectiveness (Number Errors made, Completed with ease, Completed with difficulty, Failed Completely)	H1
	Efficiency (Average time for task completion, Mouse clicks)	H2
	Satisfaction (Systems Efficiency, System Affect, ,System Helpfulness, System Controllability, System Learnability)	H3
	Cognitive Workload (Mental effort)	H4

3.4 Participants and Sampling

The sampling method adopted on this study is the probability sampling technique that gives the population of the study a chance of being sampled. The main reason for choosing this sampling technique was because as stated by Bhattacherjee [65], it is the simplest of probability sampling methods and it is not biased as everyone is given a chance to be chosen as participant .This study had 64 cattle farmers from Palapye and surrounding areas more especially Tuli block ranches who participated by using and testing the usability of text based and virtual reality based cattle diseases expert system. The participants were divided into two groups of 32 members per group. The participants were given 5 similar tasks to perform on the system they were given, which could be either text or virtual reality based expert system. It is common for usability studies to use a small sample size [66]. If one is comparing means as in our case, they use the data such as task time or rating scales which is generally

continuous (not binary), and it uses the two sample t-test. Two sample t-test been shown to be accurate for small sample sizes [66]. The other reason for choosing to use only 64 participants was because of time as the researcher could spend 30-40 minutes on one participant. Most of the studies on usability testing that one came across had smaller sample size representing the entire community being studied [66], [67], [68]. The participants were informed that participation in this study was voluntary and there will be no payments. A written consent was given to the participants. They were given a form to fill to show if they agree or disagree to participate in the study (Appendix 3). The participants did not use their names anywhere and their files are kept safely on a locked cabinet.

3.5 Research Approach

The research approach adopted is quantitative, which involves studies that enumerate their results through the use of statistical summaries [65]. This approach involves the use of requirements gathering methods that have formal measurements such as questionnaires and surveys. This approach also endeavours to get an exhaustive view from the participants. The main reason for choosing this approach is because there is a need to understand in detail if the virtual reality based cattle diseases expert system identifies symptoms and infection much easier and better than using traditional text-based expert system. This approach is highly suitable because one will be directly talking to users unlike basing on what one expects to find.

3.6 Research Purpose

This research is inferential in nature because it reasons and draws a conclusion by making a logical judgment on the basis of circumstantial evidence and prior conclusions rather than on the basis of direct observation [65]. It uses tests hypothesis using the t-test statistical method to check there is significant difference between the systems. With this method the hypothesis are either rejected or accepted. A null hypothesis is devised in opposition to an alternative hypothesis than empirical evidence is used to reject null hypothesis to show indirect, probabilistic (p-

value) support of the alternative hypothesis. The significant level of our probability is at 5%, this means that if p>0.5 there is no enough evidence to reject the null hypothesis or accept the alternative hypothesis.

3.7 Methods for Comparing the Systems

Due to the fact that the components of usability usually are dependent on the context of use as well the purposes for describing usability there are no set principles on how the combination of measures should be like. So one needs to provide a measure for all the components of quality of use, it is appropriate to for measures to be repeated in different context [32].

A statistical method independent *t*-test also known as sample *t*-test was used to compare text and VR-based expert systems. According to Gabriel et al [69] "Independent *t*-test involves examination of the significant differences on one factor or dimension (dependent variable) between means of two independent groups (e.g., text vs. VR-based expert systems).

To answer the first research question and hypothesis 1, the research used an independent–sample *t*-test to check whether there is any statistical significant difference between the two systems. The metrics used to measure task efficiency were task completion time, mouse clicks and mouse movements. If there is statistical significant difference the effect size computation used to assess the degree of the difference. The second research question and hypothesis 2 also used an independent –sample *t*-test, to check for statistical significant difference between the two groups, as the data is also continuous. The metrics used to measure task effectiveness were errors made, tasks completed with ease, tasks completed with difficulty and completely failed tasks. The affect size was also computed to check the margin of the difference. The third research questions and hypothesis 3 because its data is continuous; the independent–sample *t*-test was also used to check for the statistical significant difference between the two groups. The SUMI means, medians and the standard deviating for the five components global, affect, efficiency,

helpfulness, learnability, and controllability, were computed by the software to measure user satisfaction. After the finding the statistical difference the effect of the size computation to assess the degree of the difference was done. Lastly in the fourth research question which measured the mental effort also used the independent–sample *t*-test to test for statistical difference because the data was continuous. To confirm which system is better than the other the average scores for all the measure of the four objectives was used.

3.7.1 Measures of Effectiveness

Measures of effectiveness are described as the goals or sub-goals of using the system to the accuracy and completeness with which these goals can be achieved [4]. The components of effectiveness are measured by the quality of the solution given by the systems to users. The effectiveness of the expert systems was measured by the number of sub-tasks completely resolved and the number of errors made by the user.

A usability testing tool Morae recorder developed by TechSmith was used to measure task effectiveness [11]. The recording were then examined to determine number errors made, tasks completed with ease, tasks completed with difficulty, tasks failed completely while used both text and VR-based expert systems. Then a *t*-test statistical method was used to check if there is significant difference between the systems using of the metrics mentioned above. The means of the metrics were used a determiner of the performance of the compared systems.

3.7.2 Measures of Efficiency

Bevan and Miles [4] argued that measures of efficiency are usually linked to the level of achievements accomplished in effectiveness, particularly looking at the expenses of the resources used. The resources in our case referring to mental or physical effort that could be used to provide the measures of an individual's efficiency, time that can be used to give the measure of temporal efficiency, and lastly there is monetary cost, which could be used to give a measure of the economic efficiency [4].

Morae recorder was also used to measure task efficiency; the metrics used were Average time taken to complete a task and Mouse clicks. Sample *t*-test statistical method was used to check if there is a significant difference between the systems using of the metrics mentioned above. The means of both systems in terms of the metrics of efficiency were used to determine which of the systems performed better.

3.7.3 Measures of Satisfaction

According to Bevan and Miles [70] measures of satisfaction describe the perceived usability of the overall system by its users and the acceptability to the system to the people who use it and to other people affected by its use. Measures of satisfaction could be directed to some specific modules in the systems or the entire system.

User satisfaction of the system was measured using an attitude scale known as SUMI. This scale consists of a questionnaire with 50 questions that approximately take 10min to answer which are answered through either "Agree", "Disagree" or "Undecided" as shown in (Appendix 2). Measures of satisfaction provided very useful indication of the user's perception towards the usability of the system. SUMI can be used in to evaluate any system as long it has user inputs through the keyboard, pointing device display screen and secondary memory [4]. SUMI gives a global usability figure along with five subscales which includes efficiency, affect, helpfulness, control and learnability.

Participants, for both text and VR-based expert systems were required to fill in the SUMI questionnaire immediately after they finish using the systems. Then a *t*-test statistical method was used to check if there is a significant difference between the systems using of the metrics mentioned above. The means of the metrics were used a determiner of the performance of the compared systems. Four types of measures are achieved through SUMI includes overall assessment, usability profile consensual analysis and user acceptance. More emphasizes would be given to user acceptance.

3.7.4 Measuring Cognitive Workload

The distinct feature or element of cognitive workload includes mental effort which users need to expand excessively so as to gain acceptable performance [53]. The measures of cognitive workload can be either objective (independent to personal judgment and not directly under conscious control of the subject) or subjective (obtained through questionnaires asking users how difficult the task was) [53].

According to G. Hart [13] "NASA-TLX is a multi-dimensional scale designed to obtain workload estimates from one or more operators while they are performing a task or immediately afterwards". This technique has been used in different domains. NASA-TLX is also a subjective cognitive workload measure tool. This tool is made of scales that represent different variables which include mental, physical, and temporal demands, frustration, effort, and performance. All the scales have a description and they are rated out of 100 points with each step having 5 points as shown in (Appendix 3). After scaling all the scales then the ratings will to be put together to get a task load index. The two sample *t*-test statistical method was used to check if there is significant difference between the systems using of the metrics mentioned above. The means of both systems in terms of the metrics of efficiency was used to determine which of the systems performed better. The *t*-test statistical method was used to check if there is significant difference between the systems using of the metrics mentioned above. The means of both systems in terms of the metrics of efficiency was used to determine which of the systems performed better.

3.7.5 The Overall Score

To achieve the firth objective that was to recommend a useable cattle diseases diagnosis expert system for software developers. After comparing all the four components of usability efficiency, effectiveness, satisfaction and cognitive workload, an average score was calculated. The system which performed better than the other in terms of the measures was given 1 hence the competitor was given 0. This was only done when the alternative hypothesis was accepted. If an alternative

hypotheses was rejected a score of 0 was awarded to that measure because there will be no significant difference.

3.8 Research Questions and Hypothesis

To answer all the research questions derived from the objectives the researcher used table 3.7.1 below. The hypothesis for human computer interaction and cognitive workload were also tested with the methods provided in the table below.

Table 3.8.1: Research Questions and Hypothesis

The research questions and their Hypothesis	
	Methods
Question 1: What are the differences of efficiently completing the tasks between farmers using the text and VR-based diagnosis expert systems? Hypothesis 1: The type of expert system interface, text-only or non-immersive virtual reality will have a significant difference on tasks efficiency in terms of time taken to complete the task and mouse clicks	Video-based Observation using Morae recorder and manager
Question 2: What are the differences of effectively completing tasks between farmers using text and VR-based diagnosis expert system? Hypothesis 2: The type of expert system interface, text-only or non-immersive virtual reality will have a significant difference on tasks effectiveness in terms of numbers of errors made, task completed with difficulty, task completed	Video-based Observation using Morae recorder and manager

with ease.	
Question 3: What are the differences in user satisfaction between farmers using text and VR-based diagnosis expert system? Hypothesis 3: The type of expert system interface, text-only or non-immersive virtual reality will have a significant difference on user satisfaction in terms helpfulness, system control, affect and efficient	Software Usability Measurement Inventory (SUMI)
Question 4: What are the differences of cognitive workload experienced by farmers using text and VR-based diagnosis expert systems? Hypothesis 4: The type of expert system interface, text-only or non-immersive virtual reality will have a significant difference user cognitive workload in terms of metal effort invested	Cognitive load self-report rating scale using the NASA TLX scale

3.9 Data Collection and Experiments

The two groups of farmers (virtual reality and text based) were given five tasks which are performed on daily basis on both systems. The tasks performed included Task 1: Starting up the systems, Task 2: Selecting components from menu, Task 3: Choosing the signs and symptoms, Task 4: Getting advice and lastly Task 5: Exiting and saving results.

While the participants were performing the tasks mentioned above they were video recorded using the Morae recorder software by TechSmith. This was done so as to monitor and observe how the farmers performed the given tasks and when they

perform them and to reveal what exactly they do. With this study the mental effort invested by farmers when performing a given task was captured. Each session for recording took a minimum of 30 minutes and a maximum of 40 minutes. After completing the tasks, the software automatically gave the participants the SUS questionnaire survey to answer but the feedback was not used on this research. They were given the NASA TLX questionnaire survey to answer immediately after answering the SUS. Lastly the participants were given the SUMI questionnaire to answer. The efficiency, effectiveness and satisfaction were also captured with Morae recorder software.

The figure below (figure 3.1) shows the structure of data collection process and the experiments.

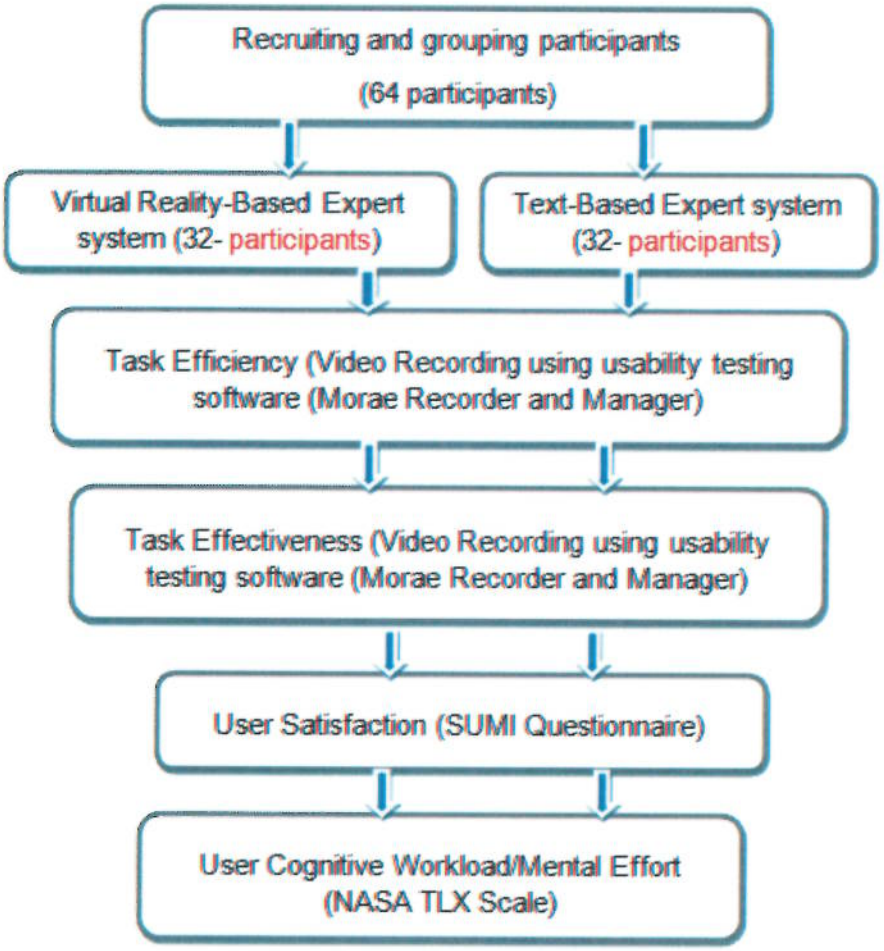


Figure 3. 1: How Data was Collected

3.9.1 Methods Used for Collecting Data

There were two methods used to collect data and they included the survey questionnaires and Video-based Observation.

3.9.1.1 Survey Questionnaire

A questionnaire is regarded as a tool used for gathering information that is self-reported by people responding to the questions, usually with regards to their feeling, attitude, belief, knowledge and their perception [71]. This method of data collection came as a package of the usability testing software Morae by TechSmith, but the data was not used in this study because of the limitations of SUS compared to SUMI. SUS introduced in 1986 was used to measure user satisfaction on this study (Appendix A) [72]. This questionnaire survey was said to be reliable and it was the most cost effective usability scale. This scale had ten items which gave a global view of subjective assessments of usability [72]. The respondents were given a scale to indicate whether they agree or disagree, and the scale is usually on 5 to 7 points.

The SUMI questionnaire has 50 questions which the user has to reply to by either showing if they agree, disagree or do not know and it is also used to measure user satisfaction as mentioned in the ISO 9241 standard [36]. The tool can also be used to measure perceived quality of use of the software. It is more detailed as compared to SUS as it scales other items such as efficiency, affect, helpfulness, control, learnability and global usability (Appendix 3). Affect is used for measuring the general emotion reactions to the software being studied it also known as likeability. The second item is efficiency and it is used for measuring the degree at which participants feel being assisted by the software when using it, similar to transparency. Helpfulness measures the degree at which the application or software does not need any explanations and is well documentations to use it [36]. Then control measures the degree at which the user feels there are in control of the system rather than being controlled by the system when performing a task [36]. Then lastly there is learnability which measures the degree at which the user has a feeling of being master of the system and easily learns newly added features [36]. SUMI is a commercial

application, meaning the user has to buy the license in order to use it but it also offers a free academic license which the research was using. SUMI was started in the late 1990's and its first ever questionnaire was started in 1993.

There is also the NASA TLX scale which was used to measure the mental effort invested by the participants when using the system [13]. The reason for adopting this method of gathering data was because it is quick and efficient (Appendix 2) [13]. All the three questionnaires used on this study are self-administered as mentioned in sub-section (3.13).

3.9.1.2 Video-based Observation

It has become easier to study how people interact with the software's and web-sites by using video-based observation software's such as Morae recorder. With this data collection method we were able to capture the video and audio. The videos were used for micro analysis of the interactions the participants were having with the system, for instance errors made by the user while interacting with system could be captured, the mouse clicks, keystrokes and cursor movement [73]. The reason why this method was chosen was because it provides records which could be referenced or used later unlike in a real-time where records are made taking notes on a diary or even journal or by self-reported surveys. The other reason for choosing this method is that it capture a detailed and rich data in a natural environment [73], [74]. The researcher is able to manipulate the video for analysis which means they are able to edit, crop and annotate the video.

The sub-section that follows describes reliability and validity of the data that was collected. According to Gray [64] "reliability and validity are tools of an essentially positivist epistemology." This is study is a positivism study which means the researcher agrees that "inquiry should be based upon scientific observation and there should be empirical" [64].

3.10 Reliability

According to Weiner [75] reliability is described as "the degree to which a measurement technique can be depended upon to secure consistent results upon repeated application". For reliability to be maintained in this study the researcher gave the participants the same questionnaires (SUS, SUMI, NASA TLX) which have been used and tested in many different studies for so many years. They proved to be successful in those studies in which they were used. With this the researcher concluded that the respondents will interpret the questions in the same way. All the questionnaires were in English.

3.11 Validity

According to Weiner [75] validity is described as "The degree to which any measurement approach or instrument succeeds in describing or quantifying what it is designed to measure". Some content expert we given content validity forms (Appendix 4) to confirm if the diagnosis items met the diseases being diagnosed. For instance if a participant is diagnosing a cattle for foot and mouth diseases, the content expert then confirmed whether the signs and symptoms chosen are for foot and mouth diseases. The experts confirmed and agreed that indeed all the signs and systems given for each disease matched that particular disease and the advice given by the system is appropriate. The sub-section that follows describes how data was processed after being collected.

3.12 Analysis of Data

Satisfaction was also measured using the SUMI questionnaire the results were then compiled and sent back to SUMI for analysis. As for the cognitive workload survey (mental effort) the NASA TLX version 2.1.2 by David Sharek of playgraph LCC was used for data analysis and generated the mean, median and standard deviation of the participants. A descriptive data analysis method two sample *t*-test was used for data analysis and an assumption that there is a statistical significance when the value of p was less than 0.05. As for the performance measures recorded by Morae

software (video based observation) the statistical analysis was generated by the same software.

3.13 Summary

This chapter highlighted and described the instruments used for collecting data and analyzing data. It has shown that these methods and instruments can be adopted in a wide range of scientific studies. The researcher found it fit to discuss the roles of participants and their protection on this study. Validity and reliability of the data being used on this was also discussed.

CHAPTER 4

Presentation of Results

4.1 Data Analysis Plan

This study adapted the inferential analysis method which tests the hypothesis. Bhattacherjee [65] has shown that" the probability also known as a p-value is compared with the alpha (α) value to determine the significant level". In our study if the p-value is less than α=0.05 then the null hypothesis was rejected and the alternative hypothesis is accepted hence if the p-value is more that α=0.05 we accept the alternative hypothesis and reject the null hypothesis. The mean is (*m*) used to determine which system is better than the other one.

The research questions the study includes the following:

1 What are the differences of **efficiently** completing the tasks between farmers using the text and multimedia based diagnosis expert systems?

2 What are the differences of **effectively** completing tasks between farmers using text and multimedia based diagnosis expert system?

3 What are the differences in user **satisfaction** between farmers using text and multimedia based diagnosis expert system?

4 What are the differences of **cognitive workload** experienced by farmers using text and multimedia based diagnosis expert systems?

The Hypotheses are as follows:

1. H1: The type of expert system interface, text-only or non-immersive virtual reality will have a significant difference on tasks efficiency in terms of time taken to complete the task and mouse clicks.

2. H2: The type of expert system interface, text-only or non-immersive virtual reality will have a significant difference on tasks effectiveness in terms of numbers of errors made, task completed with difficulty, task completed with ease.

3. H3: The type of expert system interface, text-only or non-immersive virtual reality will have a significant difference on user satisfaction in terms helpfulness, system control, affect and efficient.

4. H4: The type of expert system interface, text-only or non-immersive virtual reality will have a significant difference user cognitive workload in terms of metal effort invested

4.2 Participants' Demographics

There were 64 SUMI questionnaires analysed from all participants, the researcher also watched and analysed data from the 64 recorded videos, lastly there were 64 NASA TLX scale questionnaire analysed by the researcher. As indicated by the Figure 4.1 33% of the participants were females and 67% were males.

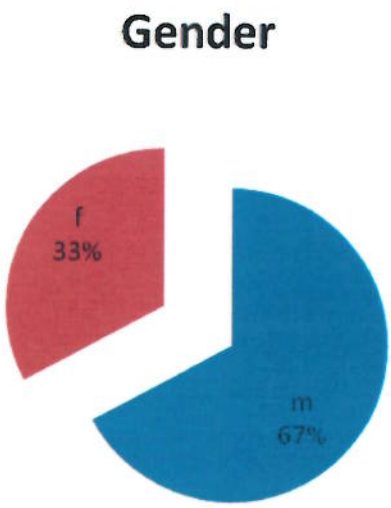


Figure 4.2.1: Participant Gender

Figure 4.2 below shows the type of farmers who participated in our study, the majority of the farmers were subsistance farmer s who made 55% population of our

study. The commercial farmers made 37% of our population of our study. And lastly we had aspiring farmers who aspires to become commercial farmers and due to some changes they face are still working towards commertial farming and they only made 8% or the poupulation of the study.

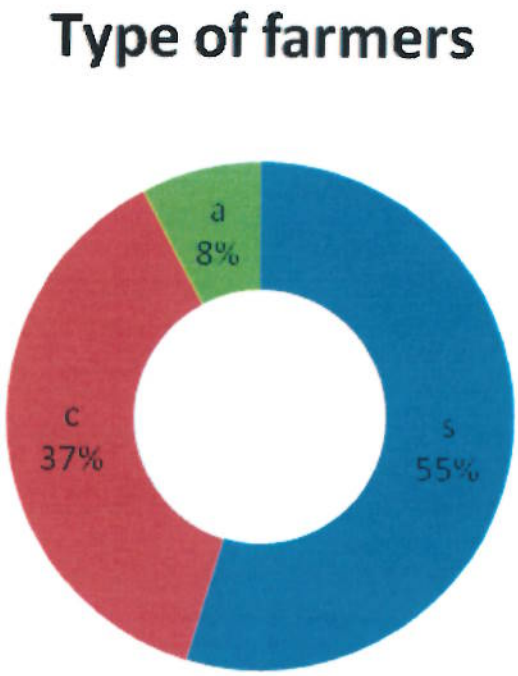


Figure 4.1.2: Type of Farmers

Most of the participants of our study were between the ages of 41 to 60 who made 50% of the population of our study. The age range of 31 to 40 is the second largest number of participants who made up to 31% of our participants. Young people with the age range of 18 to 30 made only 13% of our participant's population. The least of all age range we got were the elderly people of the ages between 61 and 80 which only made 6% of our participant's population.

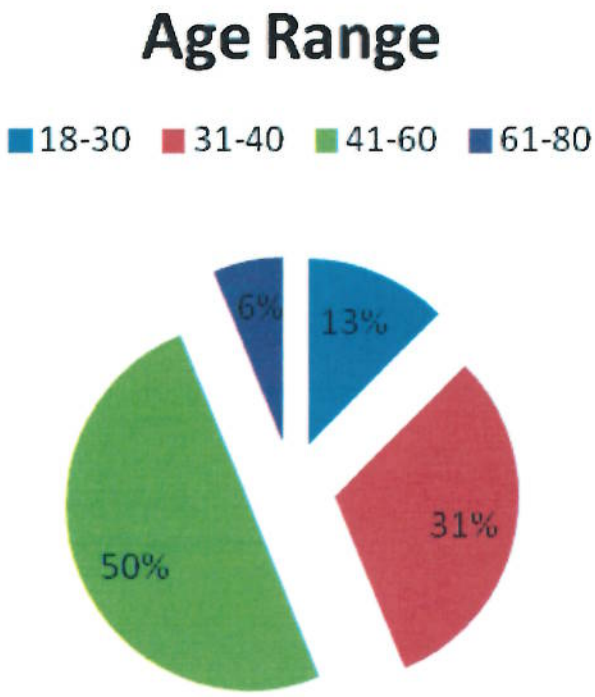


Figure 4.2.2: Age Range

4.3 Comparing Task Efficiency of the Systems

Tasks efficiency was measured through sample *t*- test using the following measures

Table 4.3.1 Average Time taken to Complete Tasks

Means, standard Deviations, and *t*-test Results for Average time for task completion (Table 1)

Average time for task completion	Text Based System (n=32)		VR-based System (n=32)		P-value (α=0.05)	Hypothesis 1 (Reject or Accept)
	Mean (*m*)	SD	Mean(*m*)	SD		
	106.77	89.58	70.45	89.70	0.0275	Accept

The means, standard deviations and the p-value were overviewed for average time for task completion (Efficiency) using the *t*-test two samples assuming unequal variances between text and VR expert system as shown in Table 1. The means for

average time for task completion are similar with only a very minor difference text being slightly higher. The "p-value" obtained from this test is 0.0275 is less than 0.05 that means that there is a statistical significant difference between the two groups. Because the VR expert system has a slightly lesser mean it means that it is better than the Text based system in terms of task completion time thus it take less time.

Table 4.3.1 Mouse Clicks

Means, standard Deviations, and *t*-test Results for Mouse clicks (Table 2)

Mouse clicks	Text Based System (n=32)		VR-based System (n=32)		P-value (α=0.05)	Hypothesis 1 (Reject or Accept)
	Mean(*m*)	SD	Mean(*m*)	SD		
	39.5	28.8623	52.0625	22.0116	0.0172	Accept

Regarding the mouse clicks made while performing the tasks results presented in table 4.2.1 indicates that the participants using the VR-based expert system (52.0625) had a higher mean as compared to those who were using the text based systems (39.5). The p-value for mouse clicks is at 0.0172 which is less than the threshold of 0.05 which means there is a statistical significant difference between the two groups. We can conclude that Text based system is more efficient in terms of Mouse clicks because it has a lesser mean of 39.5.

Table 4.3.2 Key Strokes

Means, standard deviations, and *t*-test results for mouse clicks (Table 2)

Key Strokes	Text Based System (n=32)		VR-based System (n=32)		P-value (α=0.05)	Hypothesis 1 (Reject or Accept)
	Mean(*m*)	SD	Mean(*m*)	SD		
	58.02	29.8882	22.60	12.0232	0.0127	Accept

With regard to single depression of a key on a keyboard (key strokes) the results presented in table 4.3.3 indicates that there is a significant difference between the two systems as the p value (0.0127) is less than the alpha value of 0.05. The text based expert system has a higher mean of 58.02 which mean more key strokes were pressed the VR-based expert system has a lesser mean. This means it is better that the VR-based system as it uses less keystrokes hence making more efficient Text based system.

4.4 Comparing of Task Effectiveness of the Systems

Table 4.4.1 Errors Made

Means, standard Deviations, and *t*-test Results for Number Errors made (Table 9)

Number Errors made	Text Based System (n=32)		VR-based System (n=32)		P-value (α=0.05)	Hypothesis 3 (Reject or Accept)
	Mean(*m*)	SD	Mean(*m*)	SD		
	1.75	1.077	0.8125	0.859	0.0001	Accept

Table 9 show the *t-test* independent –sample results for errors made, and the results acquired indicates that there is a significant difference between the two system because the p-value 0.0001 is way less than less than 0.05 threshold as shown in table 9. There is also a huge difference of means between the two systems with the VR expert system having a lesser mean of 0.8125 which means there were less errors made when using it than when using the text based system which had a mean of 1.75.

Table 4.4.2 Completed with Ease

Means, standard deviations, and *t*-test results for task success rate/ completed with ease (Table 10)

Completed with ease	Text Based System (n=32)		VR-based System (n=32)		P-value (α=0.05)	Hypothesis 3 (Reject or Accept)
	Mean(*m*)	SD	Mean(*m*)	SD		
	65.1620	25.7466	1628.2200	3457.4023	0.1846	Reject

For the research question three, task effectiveness the results for the success rate of the tasks completed with ease indicate that there is no significant difference because the p-value 0.1856 is greater than 0.05 thresholds as indicated in table 10. The mean difference for both systems is very high with the VR-based expert system has the highest mean of 1628.2200 which is due to the fact that the percentage of the tasks completed with ease was very high, the mean of the text based expert system is small because the percentage rate at which participants completed the tasks with ease was low. This means that the VR-based expert system's tasks were completed much easier than those of text based system.

Table 4.4.3 Completed with Difficulty

Means, standard deviations, and *t*-test results for task success rate/ completed with difficulty (Table 11)

Completed with difficulty	Text Based System (n=32)		VR-based System (n=32)		P-value (α=0.05)	Hypothesis 3 (Reject or Accept)
	Mean(*m*)	SD	Mean(*m*)	SD		
	31.008	26.8155	15.5280	8.0193	0.1356	Rejected

As for the, task Effectiveness the results for the success rate of the tasks completed with difficulty indicated that there is no significant difference between the systems because the p-value is more than 0.005 as indicated in table 11 above. There is huge difference between the means, with the VR mean being smaller which means fewer tasks for the VR systems were completed with difficulty as compared to those of text based expert system.

Table 4.4.4 Failed Completely

Means, standard Deviations, and *t*-test Results for Task Success Rate/ Failed Completely (Table 12)

Failed Completely	Text Based System (n=32)		VR-based System (n=32)		P-value (α=0.05)	Hypothesis 3 (Reject or Accept)
	Mean(*m*)	SD	Mean(*m*)	SD		
	3.8720	4.2064	1.8460	1.6858	0.1817	Rejected

Table 11 presents the results of task which failed to complete. They were acquired through an independent-samples *t*-test to examine whether there is any significant statistical difference. The results show that there is no significant statistical difference between the two systems because the p-value is more than 0.0.5 threshold. The results indicates there are fewer instances of complete failure of tasks when using the VR system as compared to when using the text based system because the VR system has a smaller mean of 1.8460.

4.5 Comparing of User Satisfaction of the Systems

User satisfaction is one of the usability measures, it is perceived usability of the overall system, and how its users accept it and the people affected by the system. Satisfaction may be measured on specific aspects of the system [32].

The SUMI was used to test the usability of both text and VR-based system. The data collected through the SUMI questionnaire is analysed by an application known as

SUMISCO. After the data has been coded it will be transformed to Global scale which depicts the systems user satisfaction in general. SUMI also has some sub-scales which include Efficiency, Affect, Helpfulness, Controllability and Learnability. According to Dolendo [76] "the z-score transformation is used to make the scales have an expected (population) mean of 50, and a standard deviation of 10".

This sub-section presents the data collected through SUMI to test user satisfaction of a text based expert system for cattle diagnosis (prototype). The results obtained from this sub-section were summarized, and compared using statistical tool (*t*- test) so as to prove the hypotheses as indicated in chapter 5 Discussion of Results.

Table 4.5.1: System Efficiency

		Text	VR	P-value (α=0.05)	Comment
Efficiency					Hypothesis 3
System		1	2		Accept
	Mean	38.0937	43.1875		
1	38.0937	0	12.0484	0.0010	System 2 (VR) much more efficient
2	43.1875	12.0484	0		

Because the sample size for both systems was the same there was an assumption for equal variances. The results of an independent-samples *t*-test for Table 4 (Efficiency) confirm that there is a significant difference between Text and VR expert systems because p-value is 0.0010 which is less than 0.05. The mean difference shows that VR-based expert system has a lesser mean than text based expert system which makes VR expert system more efficient than the text based expert system.

Table 4.5.2: System Affect

		Text	VR	P-value (α=0.05)	Comment
Affect					Hypothesis 3
System		1	2		Reject
	Mean	44.4687	46.5625		
1	44.4687	0	2.0356	0.1594	Not much difference between systems in terms of Affect
2	46.5625	2.0356	0		

With regard to affect the results on Table 4 indicates that there is no statistical significant difference between the two group's affect scores because the p-value is 0.1594 which is more than the p-value threshold of 0.05. The results were also acquired through independent-samples *t*-test therefore there was no effect size computed for the total affect to determine which is better because was no statistical significant difference.

Table 4.5.3: System Helpfulness

		Text	VR	P-value (α=0.05)	Comment
Helpfulness					Hypothesis 3
System		1	2		Accept
	Mean	41.25	49.6562		
1	41.25	0	32.8141	0.0001	VR system enormously more 'helpful'
2	49.6562	32.8141	0		

As for the systems Helpfulness the independent-samples *t*-test was used to examine whether there is any statistical difference between the groups. There results indicated that there is an enormous statistical difference between the groups as the p-value (0.0001) was way less than 0.05 thresholds as shown in table 5. The differences of the means were also huge with VR expert system having a smaller mean making it more helpful as compared to a text based expert system.

Table 4.5.4: System Controllability

		Text	VR	P-value (α=0.05)	Comment
Control System		1	2		Hypothesis 3 Accept
	Mean	43.28125	46.3437		
1	43.2812	0	4.3552	0.0416	VR system slightly more controllable, statistically significant anyway
2	46.3437	4.3552	0		

With regard to control as indicated in table 5 above an independent-samples *t-test was* also conducted to examine if there is any statistical significant difference. The results indicate that there is slight statistical significant difference between the groups as the p-value (0.0416) is less than 0.05. VR-based expert system is slightly more controllable because its mean (43.2812) is less than that of text based expert system (46.3437).

Table 4.5.5: System Learnability

		Text	VR	P-value (α=0.05)	Comment
Learnability System		1	2		Hypothesis 2
	Mean	41.8437	42.7812		
1	41.8437	0	0.40813	0.5256	Not much difference in Learnability
2	42.7812	0.408131	0		

There was no statistical significant difference between the groups in learnability as indicated by table 7 because the p-value is more than 0.05. The results obtained were computed through the independent-samples *t* test. So because there was no statistical significant difference between the groups the effect size was not computed for learnability means.

Table 4.5.6: Systems in General

		1 is text	2 is VR	P-value (α=0.05)	Comment
General System		1	2		Hypothesis 3 Accept
	Mean	41.4375	46.8125		
1	41.4375	0	13.4157	0.0005	Big difference in reactions to the two systems in general
2	46.8125	13.41572	0		

Lastly the entire system (Global/General) was tested for user satisfaction through independent-samples *t-test* and the results indicated in table 8, there was a big difference in reactions to the two systems in general because the p-value 0.0005 is way less than the 0.05 threshold which means statistically there was is a significant difference between the two systems. Nevertheless the VR expert system performed way much better than the text based system because it had a smaller mean number of 41.4375. This means that VR expert system satisfies the users much better than system the text based expert system.

4.6 Comparing of Cognitive of the Systems

Table 4.6.1: Cognitive Workload /Mental Effort

Means, standard Deviations, and *t*-test for self-imposed mental effort Results

Cognitive workload /Mental Effort	Text Based System (n=32)		VR-based System (n=32)		P-value (α=0.05)	Hypothesis 4 (Reject or Accept)
	Mean	SD	Mean	SD		
	45.52	2.90	20.80	2.75	0.0001	Accept

With regard to cognitive workload, the participants took a self-imposed test using the NASA-TLX scale and the results for both systems were compared using the independent *t*-test to examine whether there is a statistical significant difference

between the two systems. The results indicated that there is a statistical significant difference between the two systems. The mean difference between the systems was huge with the VR-based expert system having a lesser mean of 20.80 as indicated in table 13, which means participants using the VR systems exerted less mental effort as compared to those using the text based expert system.

4.7 Overall Difference

After comparing all the four components of usability efficiency, effectiveness, satisfaction and cognitive workload, an average score was calculated. The overall score was calculated using the measure/metrics of all the usability components as indicated in Table 4.7.1.

Table 4.7.1: Systems Overall Scores

Measures	VR-ES	TEXT-ES	Performed better
Time taken to complete the task	1	0	VR-ES
Mouse clicks	0	1	TEXT-ES
Key strokes	1	0	VR-ES
Errors made	1	0	VR-ES
Completed with ease	0	0	NO DIFFERENCE
Completed with difficulty	0	0	NO DIFFERENCE
Failed completely	0	0	NO DIFFERENCE
System Efficiency (For Satisfaction)	1	0	VR-ES
System Affect	0	0	NO DIFFERENCE
System helpfulness	1	0	VR-ES
System Control	1	0	VR-ES
System Learnability	0	0	NO DIFFERENCE
System General	1	0	VR-ES
Mental Effort	0	1	TEXT-ES
Total	7	2	
Total(No difference)	5		

The system which performed better than the other one in terms of the measures was given 1 hence the competitor was given 0, this was only done when the alternative hypothesis was accepted. If an alternative hypotheses was rejected a score of 0 was awarded to that measure as there will be no significant difference.

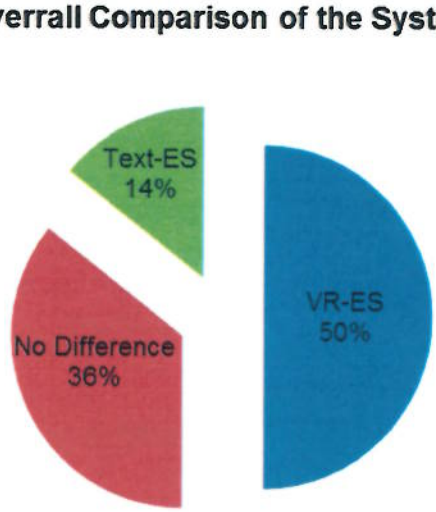


Figure 4.7.1: Overall Comparison of the System

The figure 4.7.1 above indicates that the VR-based expert systems took 50% of the measures and the text based was awarded 14 % where the measures which did not have any difference between the systems were 36%. This means that VR-based expert systems was way much better than the text based system hence we recommend that the software developer develop VR-based expert systems for cattle diseases diagnosis instead of text-based expert systems.

4.8 Summary

In summary, the overall results indicated that there is a statistical significant difference between the two systems in terms of task efficiency as its subset the average time for task completion recorded a p-value of 0.0275, and the VR-based expert system has a smaller mean making it more efficient than the text based system. In terms of user satisfaction the global scale which is the overall combination of all the four components recorded a p-value of 0.0005 which means there is a

statistical significant difference between the two systems, and the mean differences indicates that the VR-based system satisfies the users much better. As for the effectiveness of both systems the VR-based expert was more effective because it had less error rates, and as for its task success rate there were no statistical significant differences between the systems in terms of tasks completed with ease, difficulty and tasks that completely failed to complete. But in overall we can conclude that VR system was more effective because it had smaller means for the subsets of task success rate (completed with difficulty and completely failed). There is a significant difference between the two groups in terms of cognitive workload experience by the users because the p-value of 0.0001 less than the 0.05. This results indicates that while using the VR-based system, the participants experience less mental effort as compared to those using the text based system.

CHAPTER 5

Results Discussion Conclusions, and Recommendation

5.1 Introduction

The main objective of this research was to compare the usability of non-immersive virtual reality and text based diagnosis expert systems for cattle diseases. The context of the study was focused task effectiveness, task efficiency, user satisfaction and cognitive workload (mental effort) experienced by farmers when using these systems, to check if there is a statistical difference between the two systems. This study was focused on investigating and comparing the usability of these two systems in Palapye (Botswana) veterinary district area using an experimental method to collect quantitative data. The main and specific objectives were met by conducting video observations (through Morae recorder, observer, and manager), answering the SUMI questionnaire, and by taking a self-imposed cognitive workload test through NASA TLX scale. This chapter, is the summary of the findings, the contribution of study, implications of the study, allocation of resources, the limitations of the study and feature research are presented. At the end of the chapter, there is a conclusion with the limitations encountered during the study and the recommendations.

5.2 Summary of the Findings

For a deep understanding of this research topic an academic literature review addressing the comparison of text-based and multimedia based expert systems for cattle diseases was conducted, so as to put together the research questions and the hypothesis. This subsection looks at previous similar studies which have been conducted and compares their results and findings with those acquired in this study.

The reviewed literature has indicated that even though there is an increased interest in usability testing of software applications, it is surprising that so little empirical researches has been conducted on this area of study, especially in comparing the usability components effectiveness, efficiency, satisfaction and cognitive workload of

non- immersive virtual reality and text based diagnosis expert systems for cattle diseases in Botswana [27], [77]. It is also evident from the literature that there are limited ICT platforms for diagnosis of cattle diseases [15], [19], [30]. There are also limited ICT platforms for advising farmers on cattle diseases in Botswana [15], [19], [30]. All this is due to lack of sufficient knowledge on usability of these platforms. As indicated by literature there is a need to develop systems with good usability which is often a very difficult task for developers [70]. This is a clear sign that there is a need to intervene and help the farmers. So with these deficiencies the researcher found it keen to conduct a study on comparison of the two expert systems with different text and VR-based interfaces in order to distinguish which one is ideal for the users.

Software developers are able to design and develop effective, usable software applications (expert systems) which could help cattle farmers, diagnose and learn about different cattle diseases.

The research focused on usability testing which is described as "technique for ensuring that the intended users of a system can carry out the intended tasks efficiently, effectively and satisfactorily" [78]. The research supported the MUSiC Performance Measurement Method [70]. Another most important aspect of usability which its literature was reviewed and showed that is that is often over looked is cognitive workload [28], [29].

The study compared the quantitative usability testing data acquired from the following aspects of aspect of usability.

5.3 Task Efficiency

The research question 1 looked into the task efficiency of both the text and VR-based expert systems for cattle diseases diagnosis systems. We measured Task efficiency by the resources used and completely achieving a specific goal, the metrics used for measuring this construct were average time for task completion, a number of keystrokes per second and mouse clicks. The participants using the VR-based expert

system as compared to those using text-based took a lesser time to complete the tasks as compared the given tasks, the VR-based expert system also had a lesser number of keystrokes per second whereas the text based expert system had number of keystrokes per second because the commands are typed in, lastly the VR expert system had more mouse clicks reasons being it uses the mouse to manipulate everything in the system.

In summary looking at the facts above we conclude that the VR-based expert system is more efficient as compared the text based system. Our results which indicated there was a significant difference be text and VR-based expert system in terms of efficiency contradicted those acquired from a study conducted by Cheon and Grant who compared learning efficiency of a text-based interface, a graphical interface and a metaphorical interface for a web-based instruction based on cognitive load theory, which pointed out that there no significant difference between the groups. But Cheon's multimedia graphical interface (images) was not as interactive as the non-immersive virtual reality (animated 3D models) used in our system[79], [80]

5.4 Task Effectiveness

In research question 3 task effectiveness was examined, this was achieved by users accurately completing a given task of a particular goal. Effectiveness was indicated by a number of errors made by users when performing a given task and task success rate (tasks completed with ease and those completed with difficulty). The results pointed out that there was a significant difference on numbers of errors made between the two systems, the participants using the VR-based expert system making fewer errors as compared to those of a text based system. As for task success rate the results indicates that there is no significant difference between the two systems in case of tasks completed with ease and task completed difficulty. We conclude that the VR-based expert system is more effective as it has fewer errors.

The findings above are consistent with the results of a comparative study conducted by Granados [81] on two teaching methods for paediatric medication administration:

multimedia and text-based modules. The text-based group made more errors for mathematical functions as compared to the multimedia based group; this means that in terms of error made the multimedia based group was better. However the multimedia components used in this study (video clips) were different from the one we used in our study, none the less the results still shows that multimedia based systems are superior in terms of effectiveness as compared to text only based systems.

5.5 User Satisfaction

Research question two explored user satisfaction which a according to Beven [82] could be measuring using attitude rating scales such as SUMI. The following aspects of satisfaction were measured with SUMI:

5.5.1 Efficiency

The results indicated that there is a significant difference in terms of efficiency between the two systems as the p-value is at 0.0010 which is way less than the standard p-value of 0.05 so the VR-based system was more efficient. The results acquired for this aspect are consistent with results from other similar studies such as those from the study conducted by Arh et al [83] in which they evaluated the usability if an educational portal which had graphics (multimedia) and some text. The SUMI database states that the average value of an aspect can be 50 which means if it is more than 50 the user would be highly satisfied [84]. The score for efficiency for this study was 55 which means users were satisfied as the users took a shorter time in comparison to the SUMI database with the data of similar products. The VR-based system also took a shorter time in comparison to the text based system.

5.5.2 Affect

This aspect is about how the emotions of the user are captured by the system [84]. The results acquired in our study have indicated that there is no statistical significant difference between the systems. The findings above suggested that both systems were the same in terms of affect. The score of affect for both systems were above 50

which means users were feeling mentally stimulated and pleased when using both systems. The above findings were consistent with the results of a similar study conducted by Arh *et al.* [83] who was testing the usability of an educational portal which has both text and multimedia features, and it its score for affect was at 66.

5.5.3 Helpfulness

As for helpfulness the results acquired in our study indicated that there was an enormously large statistical significance between the two systems with a p-value of 0.001. The VR-based expert system emerged to be more helpful than the text based system. This means that the users of VR-based systems perceived it to be communicating with them in a helpful manner and by thus assisting them to resolve operational problems [83]. The results for this study are similar to those obtained by Chen et al who was comparing text-based and graphic User Interfaces for novice and expert Users. There was a difference in terms of helpfulness for novice users, graphic user Interfaces was more helpful to novice users than the Text-based [85].

5.5.4 Control

With regard to control the results we got for our study indicated that the VR-based system was slightly more controllable than text based system, which means there was a statistically significant difference between the two systems. This means that the users using the VR-based expert system felt that they, not the system, were setting a pace. The findings of this study in terms of the usability aspect control are consistent with the results acquired from a similar study conducted by Cheon and Grant who compared learning efficiency of different interface types based on cognitive load theory. The results which used both linear and user-controlled indicated that the graphical interface had a higher score as compared to a text based interface [31].

5.5.5 Learnability

Modality principle emphasizes that learners learn much better from images with spoken words instead of images with printed text [38]. The modality principle

instructional design technique which considers the limitation of the working memory was applied to the VR reality based expert system [55]. The signs and symptoms were shown on the 3D animated models labelled with printed words accompanied by an audio narration describing them. Most of the multimedia instructional design do not take into consideration the limit of the working memory hence they cognitive load becomes higher and thus impedes learning. Mayer [43] states that learning materials are presented in two modes, audio and visual in multimedia which takes into consideration the limit of cognitive process and the way the mind process information and so doing reducing the cognitive processes which promotes learning.

The results acquired for this aspect indicated there was no statistical significant difference between the groups with the means for text based system being 41.8437 and for VR-based system being 42.7812 and the p-value 0.5256. This means that both systems are easy to learn, there is no difference on learning how to use both systems. The users perceived both systems as easy to learn therefore they were able to perform tasks on these systems with ease. To the researcher's understanding this study compares virtual reality and text based expert systems for cattle diseases diagnosis in Botswana so it is difficult to get closely related studies to compare their results with our study, however some studies used the same method we used (SUMI) to evaluate usability.

The findings above were consistent with those acquired by Aizawa *et al* [86] who conducted a study on comparative Study of the Routine Daily Usability of FoodLog: A Smartphone based Food Recording Tool Assisted by Image Retrieval. This is a multimedia based tool which records food and permits users to photograph their meals and the produce a text based food record. In this study an evaluation of daily use was conducted with (multimedia based) and without (Text) image assistance. The results acquired from this study indicated that the participants using an image based system had a slight higher score in terms of ease of use as compared to a text only system. But the quantitative evaluation indicated that the records for both

systems did not differ significantly. This means that the multimedia and text only systems were easy to use just as indicated in the results our study acquired.

5.5.6 Cognitive Workload

The findings from this research study have suggested that there is a significant difference between text-based and virtual reality (Interactive Multimedia) cattle diseases diagnosis expert systems with VR-based expert systems being better than a text-based one in terms of task efficiency, task effectiveness, and task satisfaction. We hoped that the results acquired shows that the VR-based expert system is superior to a text based one. And if the VR-based expert system was observed to be prominent, to decide how to develop and design user interfaces which best brings information to the cattle farmer. In comparison to the results of a study conducted by Petitt [87] [63] in which he compared the text-based instructions with standard multimedia and a multimedia system that simulates an industrial setting. As opposed to the results acquired from our study which indicated that a significant difference exists between the text and virtual reality based expert systems, the results obtained from Pettit's study indicated that there was no significant difference between multimedia simulations and text-based instructions but there was a significantly better and original learning when using a standard multimedia instruction than when using a text-based one.

According to Pettit [87], [63] simulated multimedia may be more effective than text-based systems depending on the context it was used on, for instance, animations and videos with audio could easily illustrate concepts and thus assist in conveying ideas, as we have seen in our findings. Pettit has indicted that about 30 percent of the participants from his study could not visualize concepts presented in a textual format and these findings are consistent with our findings which indicated that virtual reality based system (Multimedia) is more effective, efficient and satisfying than a text-based system.

5.5.7 Contribution of Study

An effective diseases diagnosis expert system should be easy to use and understandable. As of today, there has been very little detailed empirical research conducted which compares the usability of non-immersive virtual reality and text-based cattle diseases diagnosis expert systems. After reviewing the literature it was found that most of the studies conducted were focused on expert systems for diseases diagnosis in agriculture for both plants and animals but did not dwell much on evaluating their usability. So this study seeks to address this knowledge gap.

One more contribution which this study provided the was the relative knowledge about task efficiency, task effectiveness, user satisfaction and the cognitive workload exerted by users while using a text or virtual reality based expert systems within a context of Botswana. This study provided in detail an understanding of the problems faced by farmers using text or virtual reality based expert systems for cattle diseases diagnosis.

5.5.8 Implications of the Study

There are quite a number of practical implications for system developers and farmers not only in Palapye veterinary district but anywhere with a need to introduce the cattle diseases diagnosis expert system. The implications generated on this study would be of interest to software designers and developers, cattle farmers and the veterinary physicians. There were three themes derived from the findings of this research which includes improved usableness, production of satisfied users and Cognitive overload avoidance

5.5.8.1 Usable Expert System

It is of utmost importance for any system to be usable. That is, there should be an effortless usage of the system. The findings have indicated that in terms of usage the virtual reality based expert system was way better than the text based system. For instance there was limited scrolling on the VR-based expert system. Even though there was less call to action with few clicks when using the text based expert system

it still took lesser time to accomplish a given task when using the VR-based system. Therefore the recommendation is to use a VR-based expert system to diagnose cattle diseases as it turned out to be more usable in comparison a text-based expert system.

Furthermore it is suggested that farmers use the VR-based expert system as the findings indicated that they had less error and also completed tasks with ease hence making it more preferable than the text based expert system. It seemed easier to identify signs and symptoms when using the VR-based expert system as it illustrates and shows how the signs and symptoms in real life other than when being described textually.

With an easy to use cattle diseases diagnosis expert system, this could relieve veterinary physicians and extension workers the burden of overworking as it has been indicated that we have less veterinary physicians and extension workers to take care of a large number of cattle in one region. This could also help in containing the outbreaks of certain cattle diseases. Faster response to particular diseases could help in containing that disease before it spreads or before it weakens the animal hence preventing father spread of the disease.

5.5.8.2 Train Users

Cattle farmers have different educational levels and backgrounds hence there is a large variance in computer use. The results have indicated that farmers using text based expert system committed many errors and also completed most tasks with difficulty and even though those using VR-based made less errors and also completed less tasks they need to be thoroughly trained because if they are not trained they may be frustrated by the system and hence perceive it to be difficult to use and time consuming. This may result in some farmers totally rejecting the system and sticking to the older ways of contacting a veterinary physician hence defeating the purpose of introducing the cattle diseases diagnosis expert system.

Therefore it is essential for those who would be training the users to have a deep understanding of how cognitive load which could be either overload or under load could have an effect on knowledge acquisition of new knowledge. Hence it is suggested that the trainer has to use instructional strategies which could handle learning without metal overload.

5.5.8.3 Cognitive Overload avoidance

Due factors that influence mental overload such as divided attention and duplication of data the users of the VR expert systems experienced higher cognitive over-load as compared to those using a text based expert system as it was simple and strait forward. It is suggested that as the VR expert systems has already indicated that it has an upper hand in terms of other usability factors efficiency, effectiveness and satisfaction. There should be instructional strategies deployed which could facilitate the use of the VR expert systems without metal overload. Therefore there should be only one screen (with a 3D character) showing signs and systems with fading descriptions on the symptoms, other than having a description below the 3D character.

5.5.9 Limitations of the Study

There was a smaller sample size population of 64 participants used in the study hence generalization to all cattle farmers in the country was limited. The study was also conducted in the Palapye veterinary district hence it was only limited to this area. The study was focused only on cattle farmers and does not include other farmers such as small stocks farmers (goat and sheep). The study also generalised the type farmers, it did not categorise them as commercial, subsistent or aspiring, and this could be relevant in trying to understand which type of farmers use which type of expert system more effectively and efficiently than the other.

Another limitation was that the study was only comparing two types of expert systems interfaces which were the non-immersive VR-based interface and the Text-based interfaces. There are other interfaces which could be compared such as full

immersive VR–based and semi-immersive VR-based interface expert systems. This could give a clue of which interface is more usable than other. This could be a possible future research area.

5.5.10 Directions for Future Research

There a numerous research studies conducted in this research field as it has shown that it is constantly growing. As it has been indicated in this research, a virtual reality based cattle diagnosis expert system is more efficient, satisfying and effective that the text based expert system, but there so many opportunities for expanding this research study. The limitations identified in section 5.5.9 pointed out several interesting directions which could be studied and explored in the future:

Other types of interfaces full and semi-immersive virtual reality could be compared with the one used in this study the text and non-immersive virtual reality this could lead to a deep understanding of the expert systems which could be efficiently, effectively, satisfactorily used by farmers. The farmers should also use less mental effort when using the identified expert system

Furthermore the study should categorise its users according to the type of farming they are involved in such as the commercial, subsistent, and aspiring farmers. This could help in identifying if there are distinctions in different users when using different expert systems and add features in the expert system which support different users according to their needs.

5.6 Conclusion and Recommendations

This study's main focus was to compare usability between text and virtual reality based cattle diseases diagnosis expert system, the focal point being which of this systems is better than the other in terms of effectiveness, efficiency, satisfaction and cognitive workload (mental effort). The findings indicated that the VR-based was better than text based one in terms of three factors which included effectiveness, efficiency, and satisfaction but in terms of the mental effort exerted when using the

systems the participants used more effort when using the text based system than when using the VR-based system, This could be improved by avoiding an interface with divided attention and also by avoiding repetition of data (e.g. having a 3D model describing a particular diseases by showing signs and symptoms and also having a textual description of that diseases in the same page). So above all the virtual based cattle diseases diagnosis expert system is way much better and usable as compared to the text based expert system.

The virtual reality based cattle diseases diagnosis expert system is recommended as a viable expert system given its superiority in terms of its effectiveness, efficiency and user satisfaction as compared to text base expert system. This study has only investigated the usability of the text and VR-based expert system it has not yet looked at or assessed the market demand of the this system. We recommend that these factors be assed.

REFERENCES

[1] A. Ruth, A. Babalola, S. Adedayo, and M. A. Hameed, "Web based expert system for diagnosis and management of kidney diseases," vol. 3, no. 2, pp. 9–19, 2015.

[2] C. F. Tan, L. S. Wahidin, S. N. Khalil, N. Tamaldin, J. Hu, and G. W. M. Rauterberg, "THE APPLICATION OF EXPERT SYSTEM : A REVIEW OF RESEARCH," vol. 11, no. 4, pp. 2448–2453, 2016.

[3] R. S. Michalski ;A. B. Baskin, "Integrating Multiple Knowledge Representations and Learning Capabilities in an Expert System: The ADVISE System," vol. 1, no. 1, p. 4, 1983.

[4] C. Vögeli, "Sentences Virtual Reality Training System for Patients with Dementia," 2009.

[5] H. Mclellan, "Virtual realities," pp. 461–498, 2001.

[6] E. Frekjmr, "Measuring Usability : Are Effectiveness , Efficiency , and Satisfaction Really Correlated ?," vol. 2, no. 1, pp. 345–352.

[7] H. Petrie and N. Bevan, "The evaluation of accessibility , usability and user experience," 2009.

[8] J. Eriksson, "The Relationship between Customer Satisfaction and Service Quality : a study of three Service sectors in Umeå .," 2011.

[9] R. Online, P. Chandler, and J. Sweller, "Cognitive Load Theory and the Format of Instruction," *Cogn. Instr.*, vol. 8, no. 4, pp. 293–332, 1991.

[10] TechSmith, "Software Usability Testing with Morae," Michigan, 2009.

[11] TechSmith, "Morae Software in the Usability Testing Lab," Michigan, 2013.

[12] N. Bevan and I. Curson, "Methods for Measuring Usability," *Esprit*, pp. 3–4, 1997.

[13] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Adv. Psychol.*, vol. 52, no. C, pp. 139–183, 1988.

[14] K. Jefferis, "How Trade Liberalisation can Contribute to Resolving the Crisis in the Beef and Cattle Sector," no. May, 2005.

[15] P. Lo Moro, "Botswana CDE and ITC partnership project on value chains Beef value chain findings , strategy and proposed interventions October 2014 PEER

REVIEW DRAFT," 2014.

[16] J. (Houses of P. Edgar, "Livestock Disease Surveillance .," *Parlimentary Off. Sci. Technol.*, no. 392, pp. 1–4, 2011.

[17] T. Shabani, "Shortage of veterinary officers cost beef industry," *Online News Paper*, 2013. [Online]. Available: http://www.dailynews.gov.bw/news-details.php?nid=1335. [Accessed: 19-Apr-2016].

[18] D. Xu, M. Wu, and J. An, "Design of An Expert System Based," *Network*, no. October, pp. 903–908, 2003.

[19] T. W. Timane, "Livestock management made simpler," 2016. [Online]. Available: https://www.modisar.com/. [Accessed: 20-Apr-2016].

[20] J. M. Carroll, "HUMAN-COMPUTER INTERACTION : Psychology as a Science of Design," pp. 61–83, 1997.

[21] N. Bevan, "ISO and Industry Standards for User Centred Design," *Ppt*, no. October, 2000.

[22] N. Bevan, "Human-computer interaction standards," *Adv. Hum. Factors/Ergonomics*, vol. 20, no. B, pp. 885–890, 1995.

[23] A. A. Rafea, S. El-azhari, I. Ibrahim, S. Edres, and M. Mahmoud, "Experience with the Development and Deployment of Expert Systems in Agriculture."

[24] F. S. Khan, S. Razzaq, K. Irfan, F. Maqbool, A. Farid, and I. Illahi, "Dr . Wheat : A Web-based Expert System for Diagnosis of Diseases and Pests in Pakistani Wheat," vol. I, pp. 2–7, 2008.

[25] F. Zetian, X. Feng, Z. Yun, and Z. Xiaoshuan, "Pig-vet : a web-based expert system for pig disease diagnosis," vol. 29, pp. 93–103, 2005.

[26] R. Prasad, K. Rajeev, and A. K. Sinha, "AMRAPALIKA : An expert system for the diagnosis of pests , diseases , and disorders in Indian mango," vol. 19, pp. 9–21, 2006.

[27] M. Macleod, "Usability : Practical Methods for Testing and Improvement," pp. 1–4, 1994.

[28] R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications : literature review and rationale for a new usability model," pp. 1–16, 2013.

[29] K. Arning, "Training novice adults to use a Pda in an active learning environment ask and You will receive :," vol. 2, no. March, pp. 21–47, 2010.

[30] S. Case, "A Veterinary Expert System for Remote Use," no. May, pp. 1–21, 2013.

[31] J. Cheon and M. Grant, "Do flashy interfaces help ? Comparing learning efficiency of different interface types based on cognitive load theory," pp. 2003–2007, 2003.

[32] N. Bevan and M. Macleod, "Usability measurement in context," *Behav. Inf. Technol.*, vol. 13, no. 1–2, pp. 132–145, 1994.

[33] A. Seffah and H. K. Padda, "Usability Measurement : A Roadmap for a Consolidated Model," *Quality*, pp. 1–24, 2010.

[34] N. Bevan, J. Carter, and S. Harker, "ISO 9241-11 revised : What have we learnt about usability since 1998 ?," pp. 143–151, 2015.

[35] N. Bevan, "International Standards for HCI," *Encycl. Hum. Comput. Interact.*, vol. 55, no. May, pp. 1–15, 2006.

[36] B. Journal and O. F. Educational, "SUMI : the Software Usability Measurement Inventory," no. OCTOBER 2006, 2016.

[37] Z. Mansor, Z. M. Kasirun, S. Yahya, and N. H. Arshad, "The Evaluation of WebCost Using Software Usability Measurement Inventory (SUMI)," vol. 2, no. 2, pp. 197–201, 2012.

[38] R. E. Mayer and R. Moreno, "A Cognitive Theory of Multimedia Learning: Implications for Design Principles," *Annu. Meet. ACM SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 1–11, 1998.

[39] Mayer. Richard. E, *Multimedia learning second edition*, Second edd. New York: Press, Cambridge University, 2009.

[40] S. D. Sorden, "The cognitive theory of multimedia learning," *Handb. Educ. Theor.*, pp. 1–31, 2012.

[41] "Cognitive Theory of Multimedia Learning (Mayer)," p. 2014, 2014.

[42] T. de Jong, "Cognitive load theory, educational research, and instructional design: Some food for thought," *Instr. Sci.*, vol. 38, no. 2, pp. 105–134, 2010.

[43] R. E. Mayer, "Research--Based Principles for Designing Multimedia Instruction Overview of Multimedia Instruction," *Acknowledgement and Dedication*, pp. 1–12, 2013.

[44] R. Low and J. Sweller, "The Modality Principle in Multimedia Learning," *Cambridge Handb. Multimed. Learn.*, no. 2007, pp. 147–158, 2005.

[45] E. G. Harskamp, R. E. Mayer, and C. Suhre, "Does the modality principle for multimedia learning apply to science classrooms?," *Learn. Instr.*, vol. 17, no. 5, pp. 465–477, 2007.

[46] J. P. T. Dalsgaard, M. Thai Thi, G. Vo Ngan, and J. C. Riise, "Introducing a farmers' livestock school training approach into the national extension system in Vietnam," *Netw. Pap. - Agric. Res. Ext. Netw.*, no. 144, 2005.

[47] K. Butcher, "The multimedia principle in multimedia learning," 2014.

[48] X. Li, S. F. Fong, B. Muniandy, A. Hanafi, and M. I. Rozhan, "Computer-based Human Agent Image and Modality Effects – Does it Facilitate Deeper Information Processing ? Overview of Present Study," *Educ. Technol.*, vol. 10, no. 2, pp. 11–18, 2010.

[49] J. Swezller and J. Sweller, "Cognitive Load Theory , Learning Difficulty , and Instructional Design," *Learn. Instr.*, vol. 4, pp. 295–312, 1994.

[50] H.-H. Choi, J. J. G. van Merriënboer, and F. Paas, "Effects of the Physical Environment on Cognitive Load and Learning: Towards a New Model of Cognitive Load," *Educ. Psychol. Rev.*, vol. 26, no. 2, pp. 225–244, 2014.

[51] J. Paas, Fred; Renkl, Alexander; Sweller, "Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture," vol. 32, no. 1–2, pp. 1–8, 2004.

[52] F. Paas, A. Renkl, and J. Sweller, "Cognitive Load Theory and Instructional Design: Recent Developments," *Educ. Psychol.*, vol. 38, no. 1, pp. 1–4, 2003.

[53] Eunbae, "Direct measurement of cognitive load in multimedia learning," *Brunken, Plass & Leutner*, 2010. [Online]. Available: https://dixieching.wordpress.com/2010/02/21/direct-measurement-of-cognitive-load-in-multimedia-learning-brunken-plass-leutner/. [Accessed: 09-Jun-2016].

[54] S. G. Hart, "NASA-TASK LOAD INDEX (NASA-TLX); 20 YEARS LATER."

[55] R. Brünken, J. L. Plass, and D. Leutner, "Direct Measurement of Cognitive Load in Multimedia Learning," *Educ. Psychol.*, vol. 38, no. 1, pp. 53–61, 2003.

[56] N. Nourbakhsh, Y. Wang, F. Chen, and R. a. Calvo, "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," *Proc. 24th Conf. Aust. Comput. Interact. OzCHI '12*, pp. 420–423, 2012.

[57] Y. Kim and C. Author, "Dual-Task Performance as a Measure of Mental Effort in Searching a Library System and the Web."

[58] D. Li, Z. Fu, and Y. Duan, "Fish-Expert : a web-based expert system for fish disease diagnosis," vol. 23, pp. 311–320, 2002.

[59] I. Ibrahim, A. A. Rafea, and S. El-azhari, "Experience with the Development and eployment Agriculture of Expert Systems in," 1995.

[60] C. Sciences, "Expert Doctor Verdis : Integrated medical expert system," pp. 1032–1044, 2014.

[61] C. De Teatinos, "Patient modelling using Expert Systems for medical training simulations based on Virtual Reality."

[62] J. Melorose, R. Perroy, and S. Careas, "Intergrating Diagnostic Expert with image prosessing via loosely coopled technique," *Statew. Agric. L. Use Baseline 2015*, vol. 1, 2015.

[63] T. D. Ramsey, "The Effects of Multimedia Interface Design on Original Learning and Retention," no. December, p. 107, 1996.

[64] E. Gray , David, *Doing research in the real word*, 3rd ed. Lodon,Los Ageles,New Delhi,Singapore,Washington DC: Sage Publications India Pvt. Ltd, 2014.

[65] A. Bhattacherjee, *Social Science Research: Principles, Methods, and Practices*. 2012.

[66] M. Galani, "Factors affecting nursing staff use of nursing information systems in residential aged care homes," 2015.

[67] V. Z. Ogozalek, "A Comparison of the Use of Text and Multimedia Interfaces to Provide Information to the Elderly," pp. 65–71, 1994.

[68] J. A. Ogolla, "Usability Evaluation : Tasks Susceptible to Concurrent Think-Aloud Protocol MASTER THESIS-TQDV30."

[69] H. Gabriel, Schlomer.Lynne M. Borden.Stacy , Ann, "t -T ESTS A H OW -T O G UIDE," 2014.

[70] M. Macleod, R. Bowden, and N. Bevan, "The MUSiC Performance Measurement Method," 1993.

[71] C. Guidelines, "Research & Consultation Guidelines," *Kirklees Gov. uk*, p. 6, 2006.

[72] J. Brooke, "SUS - A quick and dirty usability scale."

[73] C. Jewitt, "An introduction to using video for research An Introduction to Using Video for Research," *Natl. Cent. Res. Methods Work. Pap.*, pp. 1–22, 2012.

[74] Esomar, "Passive Data Collection, Observation and Recording," pp. 1–20, 2009.

[75] J. Weiner, "Measurement : Reliability and Validity Measures," *Johns Hopkins Univ.*, pp. 1–27, 2007.

[76] M. E. Dolendo, "USABILITY MEASUREMENT OF LEARNING MANAGEMENT SYSTEMS : A RESPONSE TO EDUCATIONAL TECHNOLOGY INFLUX," vol. 2016, no. March, pp. 146–157, 2016.

[77] M. Saeed, "Usability Evaluation of a Health Web Portal," no. June, 2009.

[78] G. Gaffney, "Usability Testing What is Usability Testing ?," p. 1999, 1999.

[79] L. V. G. Marc Pollefeys, "From Images to 3D Models How computers can automatically build realistic 3D models from images acquired with a hand-held camera," pp. 1–7, 2006.

[80] Y. Tisserand and N. Magnenat-thalmann, "Image-based 3D Avatar for Virtual Try-on Applications," pp. 269–279, 2015.

[81] R. Granados, "A Comparison of Two Teaching Methods for Pediatric Medication Administration : Multimedia and Text-Based Modules," 2013.

[82] N. Bevan, "Measuring usability as quality of use," vol. 150, 1995.

[83] T. Arh and B. J. B. Č, "A Case Study of Usability Testing – the SUMI Evaluation Approach of the EducaNext Portal," vol. 5, no. 2, 2008.

[84] D. Erik and P. W. M. V. V. Cisa, "Questionnaire based usability testing," no. November, pp. 1–9, 1998.

[85] J. Chen and J. Zhang, "Comparing Text-based and Graphic User Interfaces for Novice and Expert Users," pp. 125–129, 2007.

[86] K. Aizawa, K. Maeda, M. Ogawa, Y. Sato, M. Kasamatsu, and K. Waki, "Comparative Study of the Routine Daily Usability of FoodLog : A Smartphone-based Food Recording Tool Assisted by Image Retrieval," pp. 1–6, 2014.

[87] C. S. Petitt, "Simulating user experiences in computer-based multimedia instruction," Virginia Polytechnic Institute and State University, 1994.

APPENDICES

Appendix 1 Research Permit

Telephone: 3689504
Fax No: 3903744
Internet: *www.moa.gov.bw*
Email:
REF: **DVS 7/4/1 VI (51)**

Republic of Botswana

Ministry of Agriculture
Department of Veterinary Services
Private Bag 0032
Gaborone
Botswana

31 August 2016

Botswana International University
Of Science & Technology
Private Bag 16
Palapye
Attention Dr Keletso Letsholo

Dear Sir/Madam

RESEARCH PROJECT – MR KABO POLOKO NKABITI

We refer to yours dated 12 July 2016 on the above subject matter.

Be informed that we do not have any objection on study and will cooperate as necessary where we can.

Thank you.

Yours faithfully

Dr Letlhogile Modisa
Director of Veterinary Services

OUR VISION: Attain national food security and global competitiveness in agricultural products
OUR MISSION: To improve agricultural productivity through technology development and transfer, diversification and commercialization, in order to promote food security in partnership with our stakeholders

BOTSWANA
Our pride, your destination

Appendix 2 Software Usability Measurement Inventory Questionnaire (SUMI)

This questionnaire has 50 statements. Please answer them all. After each statement there are three boxes.

- Check the first box if you generally AGREE with the statement.
- Check the middle box if you are UNDECIDED, or if the statement has no relevance to your software or to your situation.
- Check the right box if you generally DISAGREE with the statement.

In checking the left or right box you are not necessarily indicating strong agreement or disagreement but just your general feeling most of the time.

There are also five general questions at the end.

Statements 1 - 10 of 50.	**Agree**	**Undecided**	**Disagree**
This software responds too slowly to inputs.	○	○	○
I would recommend this software to my colleagues.	○	○	○
The instructions and prompts are helpful.	○	○	○
This software has at some time stopped unexpectedly.	○	○	○
Learning to operate this software initially is full of problems.	○	○	○
I sometimes don't know what to do next with this software.	○	○	○
I enjoy the time I spend using this software.	○	○	○
I find that the help information given by this software is not very useful.	○	○	○
If this software stops it is not easy to restart it.	○	○	○
It takes too long to learn the software functions.	○	○	○

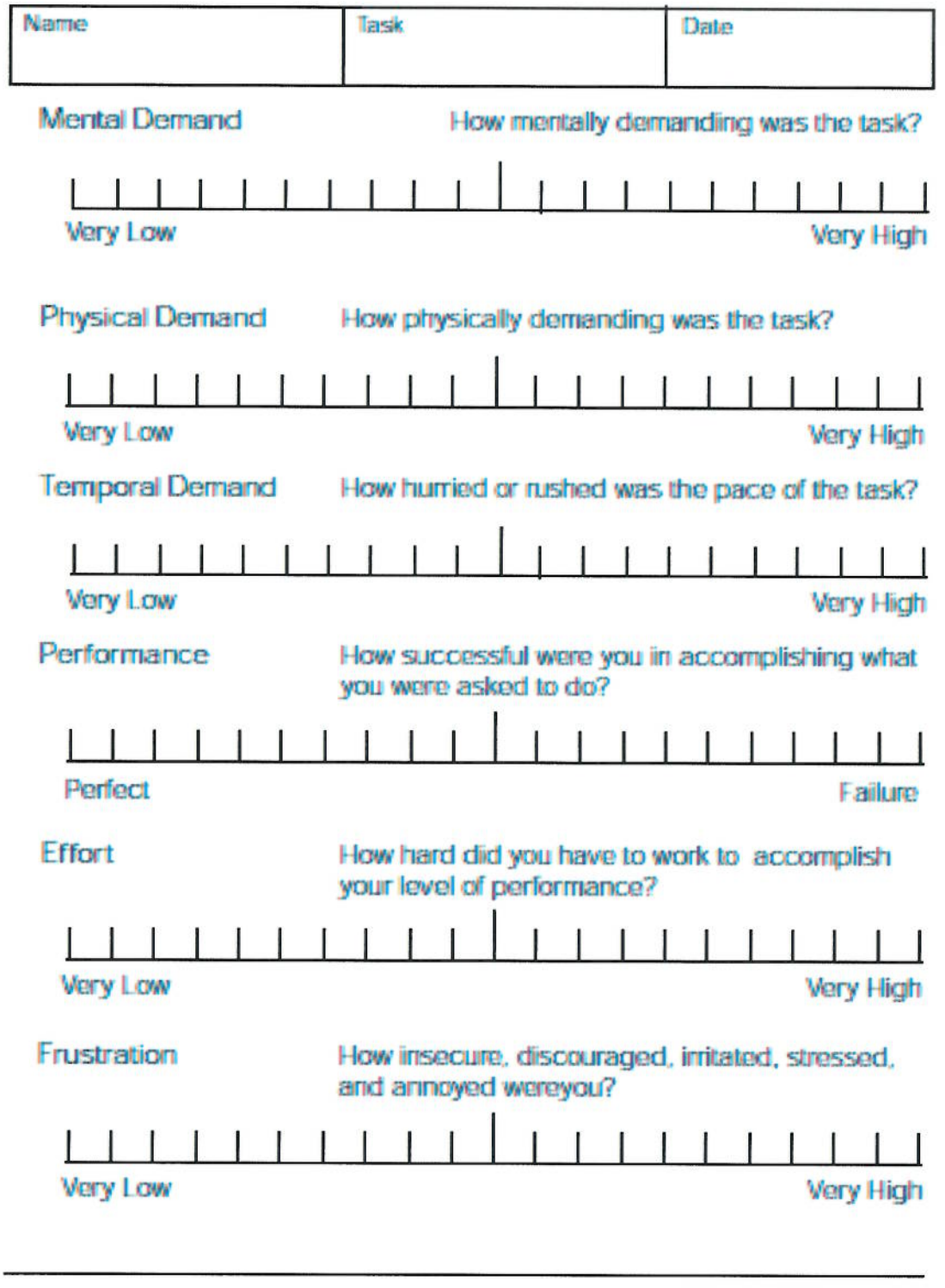
Statements 11 - 20 of 50.	Agree	Undecided	Disagree
I sometimes wonder if I am using the right function.	○	○	○
Working with this software is satisfying.	○	○	○
The way that system information is presented is clear and understandable.	○	○	○
I feel safer if I use only a few familiar functions.	○	○	○
The software documentation is very informative.	○	○	○
This software seems to disrupt the way I normally like to arrange my work.	○	○	○
Working with this software is mentally stimulating.	○	○	○
There is never enough information on the screen when it's needed.	○	○	○
I feel in command of this software when I am using it.	○	○	○
I prefer to stick to the functions that I know best.	○	○	○

Statements 21 - 30 of 50.	Agree	Undecided	Disagree
I think this software is inconsistent.	○	○	○
I would not like to use this software every day.	○	○	○
I can understand and act on the information provided by this software.	○	○	○
This software is awkward when I want to do something which is not standard.	○	○	○
There is too much to read before you can use the software.	○	○	○
Tasks can be performed in a straight forward manner using this software.	○	○	○
Using this software is frustrating.	○	○	○
The software has helped me overcome any problems I have had in using it.	○	○	○
The speed of this software is fast enough.	○	○	○
I keep having to go back to look at the guides.	○	○	○

Statements 31 - 40 of 50.	Agree	Undecided	Disagree
It is obvious that user needs have been fully taken into consideration.	○	○	○
There have been times in using this software when I have felt quite tense.	○	○	○
The organisation of the menus seems quite logical.	○	○	○
The software allows the user to be economic of keystrokes.	○	○	○
Learning how to use new functions is difficult.	○	○	○
There are too many steps required to get something to work.	○	○	○
I think this software has sometimes given me a headache.	○	○	○
Error messages are not adequate.	○	○	○
It is easy to make the software do exactly what you want.	○	○	○
I will never learn to use all that is offered in this software.	○	○	○

Statements 41 - 50 of 50.	Agree	Undecided	Disagree
The software hasn't always done what I was expecting.	○	○	○
The software presents itself in a very attractive way.	○	○	○
Either the amount or quality of the help information varies across the system.	○	○	○
It is relatively easy to move from one part of a task to another.	○	○	○
It is easy to forget how to do things with this software.	○	○	○
This software occasionally behaves in a way which can't be understood.	○	○	○
This software is really very awkward.	○	○	○
It is easy to see at a glance what the options are at each stage.	○	○	○
Getting data files in and out of the system is not easy.	○	○	○
I have to look for assistance most times when I use this software.	○	○	○

Appendix 3 NASA Task Load Index (NASA-TLX)



Appendix 4 Farmers Consent for Research Participation

Farmers Consent for Research Participation

Purpose of the research study

The use of cattle disease diagnosis expert system has the ability to improve the services for cattle production in terms of disease control. Particularly if the system is usable that is if it is efficient, effective, satisfying and users invest less effort when using it.

The main objective of this study is to compare the usability of non-immersive virtual reality and text based diagnosis expert systems for cattle diseases in the context of effectiveness, efficiency, satisfaction and cognitive workload experienced by farmers when using the system.

You are requested to take part on this study as a participant because you are a farmer who currently needs to use a usable cattle diseases diagnosis expert system. You will be asked to use either a Text or Virtual-reality based cattle diseases diagnosis expert system. After using the system you will be asked to answer the SUMI questionnaire and fill the NASA TLX scale

Procedures

When you agree to participate on this study, you be asked to do the following;

- The main aim of carrying out the study will be explained to you.
- You will be given a consent letter.
- You will be asked to use the system.
- You will be asked to give feedback about the system through answer the SUMI questionnaire and the NASA TLX scale.
- You will be asked to commit about 75 min of your time

Discomfort or Risks

There are no known psychological discomforts or risks which could face participants.

Reimbursement or Cost

Participants will not be compensated for taking part in the study, nor will they incur any costs.

Confidentiality

The names of the participants will be kept confidential and they will not be mention anywhere in the study.

Comments and Questions

If you have any questions or comments about the study you are free to contact the research on his mobile phone (+267 71969168) or on his email. If you want to contact the university about the research you reach them on +267 49000177.

Consent

Keep the copy of the consent given to you.

The participant has the liberty to accept or decline to participate on the study as participation is voluntary. The participant can also withdraw his or her participation at any time.

Signature_____ Date-

_____

Participant

Signature_____ Date-

_____

Researcher

Thank you for your participation

Kabo Poloko Nkabiti

Masters Student, BIUST

Consent form

_____I will participate in this study.

_____ I will not participate in this study.

Appendix 5 Journal Paper (Published in: International Journal of Information and Electronics Engineering (IJIEE), 2017)

Comparison of User Satisfaction between Text and Virtual Reality Based Expert Systems for Diagnosing Cattle Diseases

[1]Kabo Nkabiti, [2]Malatsi Galani, [3]Adamu Murtala Zungeru

[1,2]*Department of Computer Science and Information Systems*
Botswana International University of Science and Technology
Private Bag 16, Palapye, Botswana

[3]*Department of Electrical, Computer and Telecommunication Engineering*
Botswana International University of Science and Technology
Private Bag 16, Palapye, Botswana

kabo.nkabiti@studentmail.biust.ac.bw, galanim@biust.ac.bw, zungerum@biust.ac.bw

***Abstract*— Even though usability testing of software applications has been studied for decades, little empirical researches have been conducted in Agriculture domain, especially in comparing the usability of none- immersive virtual reality and text based diagnosis expert systems for diagnosing cattle diseases. A quasi -experimental research was conducted in which 62 cattle farmers both commercial and subsistent used either a non-virtual reality based (System A) or a text-based (System B) expert systems to diagnose cattle diseases. An attitude rating scale Software Usability Measurement Inventory (SUMI) was used to collect data and the reactions towards the expert-based systems. The measures were efficiency, the affect, helpfulness of the system, controllability system, and system Learnability. The results obtained from the research indicated that the group of farmers using System A was more satisfied as compared to those using system B. System A is also more efficient than System B. There is not much difference between systems in terms of Affect. System A is enormously more 'helpful than B, there is a statistical significance difference between System A and system B in terms of system control with systems A being more controllable. There is also not much difference in the learnability of both systems. There was a significant difference in reactions to the two systems in general with of System A being better than System B. In overall summary we can conclude that System A satisfies users much better than system B.**

Keywords: usability, human computer interaction, non-immersive virtual reality, expert system, cattle diseases, user satisfaction.

I. INTRODUCTION

To set out an atmosphere for this research paper we start by giving a brief overview of usability. Usability is a key issue in human-computer interaction (HCI) since it is the aspect that commonly refers to the quality of the user interface [30]. The International Standards Organization (ISO) defines usability as [16]: *The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context.* The main key subject of human-computer interaction is usability because it usually deals with the quality of the user interface [1]. Usability is defined by International Standards Organization 9241-11 as [2] "*The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context.*" This study is focused on user satisfaction which is defined as the perceived usability of the overall system, and how its users accept it and the people affected by the system.. Satisfaction may be measure on specific aspects of the system or may be used to measure the overall system [3]. It is essential for software, websites, mobile application, or any user-operated technology developers and designers to create systems which satisfy the needs of the users [4]. Despite an increased interest in usability testing of software applications, it is surprising that so little empirical

researches have been conducted in this area of study, especially in comparing the usability of non-immersive virtual reality and text based diagnosis expert systems for cattle diseases. In this study, the focus is to compare the user satisfaction of the two diagnosis expert systems for cattle diseases. Different expert systems for diagnosing diseases have been developed mostly using CLIPS shell [1] [2] [3] which are text based. Some of these systems uses software such as ESTA [4] [3]which is an "expert system shell developed by Prolog that allows the use of using pictures in the title, parameter, in advice and in action list"[9].

We experimented with two expert systems that use different user interfaces to determine user's satisfaction. The two systems compared were text-based expert systems (System B) developed using CLIPS, expert system development shell, and the virtual reality based expert systems developed using 3ds max 2016. We added interactivity to the 3D models and animation using Unity software. This helped to show disease symptoms and signs more clearly.

In our investigation, we aimed to address the following research question;

- Is there a significant difference in user satisfaction between farmers using text and non-immersive virtual reality based cattle diseases diagnosis expert system?

The paper is arranged and divided into the following sections, firstly we have the introduction which introduces the problem and gives an insight into user satisfaction. The second section is the related work which looks at what has already been done which is similar to our research. Under section three we also discuss the methods used to achieve the main of the study, the participants, and the procedure. The fourth section is the result presentation. The fifth section is results discussion which includes its limitations, implications and comparison to the existing results then the last section is the conclusions.

II RELATED WORK

This sub-section describes similar works and discusses the results obtained from those studies.

Ogogzake[10] carried an experimental research by comparing the use text and multimedia interface for the provision of information about drugs prescription to old people with an average of 71years. The participants (62 divided into two groups) compared the paper information-based system with a computerized information system. The results indicated that the group of users using multimedia-based presentations got much better results, the users preferred to use multimedia interfaces. The electronic system performed better that the text based. In terms of gender, men preferred both the multimedia and text-based interfaces whereas women found the multimedia system more pleasant than the text-based interface.

In another study, two teaching methods for presentation of learning materials to pediatric nursing students were compared by Granados [9] at the University of San Francisco. Text-based modules and multimedia based modules were compared. The independent variable of the research was "knowledge acquisition of mathematical calculation skills for medical administration"[11]. The research defined calculation skills as the ability of the student to calculate (1) weight–based safe dose ranges, (2) intravenous flow rates for primary and secondary medical infusion, (3) Converting pounds to kilograms, and lastly (4) the maintenance of fluid [5]. The results of the research indicated that there was no statistical significant between the two modules in respect of the pass rate, as well the four items (1),(2),(3) and (4) mentioned above. There was also an indication of the results that the modules were not as effective as anticipated for teaching and delivering the pediatric medical administration materials. In addition to the findings, it was also evident that the preponderance of errors was similar for both groups with an exception of only a few errors with a minor statistical significant.

Also, Jung-Wei Chen et al [12] conducted a research on a comparison of graphical user interface (commonly considered to be better) with text-based interfaces for a dental records system. Two groups of users were evaluated in this study, the novice, and the expert users to check the time and steps taken to complete the tasks. Different usability evaluation and analysis techniques such as task analysis, GOMS (Goals, Operations, Methods and Selection rules), HTA (Hierarchical task analysis) and Distributed representation analysis were used to compare the effectiveness of the systems. To assess whether experience affects the task performance of the systems the within-subject design evaluation method was used. The results indicated that for experts' user the graphical user interface (GUI) was not better that the text-based interface (TBI). The novice users using the systems with GUI performed better than those using the TBI. The research concluded by indicating that the user-friendliness of the user interface of the system depends on how the task was mapped which means that GUI can be better than TBI or may not be better it only depends on the task being carried out.

Furthermore in 1996 Ramsey [13] conducted a research on a comparison of three instructional learning methods which includes a text-based instructional system and two multimedia systems. The two systems using multimedia instructions used different interfaces the topic-oriented and problem-solving interface. Intellectual skills and verbal information were the domains of learning being examined in this experimental research design. The two designs used for examination of original learning as well as the retention components of learning [13]. The research measured the original learning first then followed by instructional treatment. Then after a fortnight

the retention will be measured. There were 30 participants participating in each session. Analysis of variance was used to analyze the scores. The results obtained indicated that there was no statistical significance difference between the text-based and the two multimedia systems during both domains original learning and retention sessions [13] .

Lastly, Cheon and Grant [14] compared the learning efficiency of different types of interfaces which were based on the cognitive load theory. The interfaces being compare included the text-based, graphical interface as well as the metaphorical interfaces so as to establish which interface has a lower cognitive workload. Fifty undergraduates metal effort levels and their performance score we compared. The results obtained indicated that there was no statistical significance difference between the groups in terms learning efficiency.

From the reviewed related work it evidently showed that studies on a comparison of user satisfaction between text-based and non-immersive virtual reality based expert systems are very limited. Non-immersive virtual reality application on animal health expert systems is a new concept so there is limited research made on the field [15] [16] . With the factors identified above the researcher found it fit to do research on this title and hopefully this paper will contribute knowledge in this area of study.

There are two factors which distinct our research from the related works. Firstly the comparison that we conduct is based upon text based and virtual reality (interactive multimedia) based expert system, unlike the most related works which compared the text with normal multimedia (images) [10] [5] [12] [13]. Secondly, we adopted the software usability measurement inventory questionnaire (SUMI) to assess the participant's satisfaction with both systems and evaluate their preference. The ISO 9241 standard regards SUMI as a method used to test the satisfaction of a user [2][17]. This common usability tool is made up of validated 50-questions in which participants score each question on a scale of three points which are agree, undecided or disagree. The participants were allowed to ask the researcher questions during task performance. The questionnaire (SUMI) was answered by the participants immediately after using the systems. The data collected from the questionnaire was then sent to Dr. Jurek Kirakowski (SUMI employee) for analysis. The researcher managed to acquire the academic SUMI license which is free [18][19][20].

III MEASURES OF USER SATISFACTION

Different studies have indicated that user satisfaction has five measures a listed below;

• Efficiency: This is how the user feels about systems responds when tasks are performed through it which includes aspects such as time taken to complete a task, resources used to complete the task by means of the systems performance [11].

• Affect: This is referred to as emotional feeling in psychology. In the context of our case, it means when the participant is feeling mentally stimulated in an unpleasant way or in a pleasant way after using the systems [18].

• Helpfulness: This whereby the user perceives the system to be communicating in a helpful manner, and there is a proper assistance to resolve the problems encountered while operating the system [18].

• Control: This is the degree at which the user feels that he/she has power over the system and he/she, not the system is pace setter [18].

• Learnability: the ability of the user to learn new features of the systems at ease and get easily gets started.

IV THE HYPOTHESES

The hypotheses that were tested in the experiment were as follows;

Hypothesis 1: There is no significant difference in the type of cattle disease diagnosis expert system text or virtual reality based on tasks efficiency.

Hypothesis 2: There is no significant difference in the type of cattle disease diagnosis expert system text or virtual reality based on tasks affect.

Hypothesis 3: There is no significant difference in the type of cattle disease diagnosis expert system text or virtual reality based on tasks helpfulness.

Hypothesis 4: There is no significant difference in the type of cattle disease diagnosis expert system text or virtual reality based on tasks controls.

Hypothesis 5: There is no significant difference in the type of cattle disease diagnosis expert system text or virtual reality based on tasks learnability.

Hypothesis 6: There is no significant difference in the type of cattle disease diagnosis expert system text or virtual reality based on tasks in general.

V METHOD

This section provides information for judging the validity of this study. It is a clear description of how the experiment for this study was done [27].

Quasi-Experimental Design

The independent variables tested were the usability measures task efficiency, task affect, task control, task learnability and

task helpfulness tested on both text and virtual reality based expert systems. The between-subject design was used to compare the text and the virtual reality based expert systems. The 62 participants were dived into half and the half used text based system whereas the other half used the virtual reality based expert systems. Both interfaces were desktop based. Usability testing software called Morae from TechSmith the use of the systems by the participants [21] [22]

Participants

We tested the two systems with 64 individuals and they were divided into a group of 32 participants each. The groups included both commercial and subsistent farmer young and old mostly those who could and read and were registered with the department of veterinary services which is where they were recruited. The age range of the participants was approximately between 18 and 65. The participants were assured that the information they shared with us would be kept private and confidential and it would be anonymous [23] [24].
Procedure

When starting up the experiment the participants were giving a short lesson on how the systems works and they familiarized themselves with its purpose in general. They were informed that they have limited time to complete different tasks. For the situation to be realistic the participants were requested to imagine diagnosing their cattle in real life as opposed to just using their knowledge to as or even their opinion when performing the set tasks.
As the participants performed the given tasks they were left alone in a room and Morae recorder recorded a video of everything they were doing. After the participants completed all the tasks they signaled the experimenter and he gave them a SUMI questionnaire to complete for user satisfaction assessment which determined their preferred system.

The procedures for both systems were the same. Even some participants had difficulty using the systems the experimenter gave them more time to familiarize themselves with the systems, but otherwise, most participants did not have any difficulties understanding how the systems work.

VI SUMI Results

Figure 1 below shows the SUMI scale profiles for the means and standard deviations for the VR-based expert system. According to Erik and Veenendaal [25] "The SUMI database states that 50 is the average value for the global (General) score in a normal distribution and the standard deviation is 10". When a value is greater than 50 then user satisfaction is considered to be more that average. The data from figures 1, figures 2, figures 3 and figures 4 were used for statistical assessment for comparing the two systems.

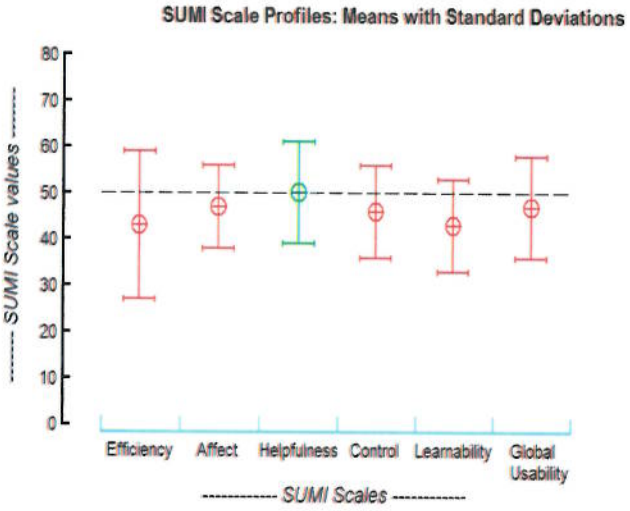


Figure 1: VR Means and SD

	Mean	St Dev	Median
Global	46.81	11.12	43.5
Efficiency	43.19	15.58	39.0
Affect	46.56	9.23	45.0
Helpfulness	49.66	11.32	47.5
Controlability	46.34	10.48	44.0
Learnability	42.78	10.47	41.0

Figure 2: Virtual reality-based System Means and SD and Median

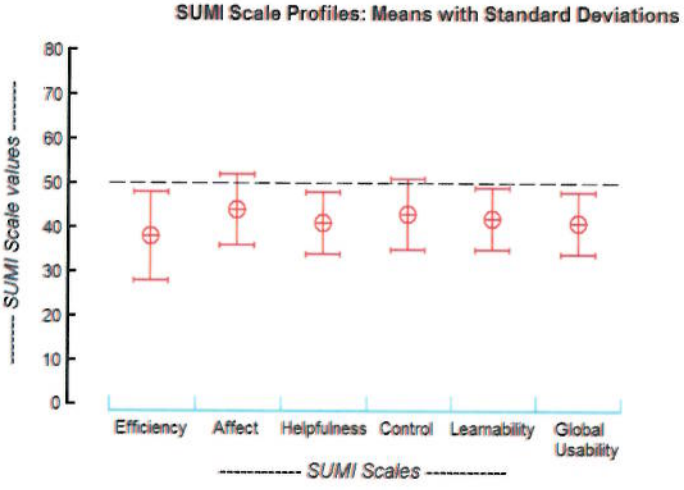


Figure 3: Text-based System Means and SD

	Mean	St Dev	Median
Global	41.44	7.37	40.0
Efficiency	38.09	9.69	37.0
Affect	44.47	7.84	44.5
Helpfulness	41.25	7.26	42.5
Controlability	43.28	7.58	43.0
Learnability	41.84	7.34	40.0

Figure 4: Text-based System Means and SD and Median Results Presentations

An overview of means, p-values, and independent-samples t-test for the total scores of user satisfaction (efficiency, affect, helpfulness, controls, learnability and systems in general) for both virtual reality and text based expert systems are presented in tables below;

Table 1: Efficiency

Compared Systems	Means	P-value(0.05)	Hypothesis 1
System A (VR)	38.09375	0.0010	Rejected
System B (Text)	43.1875		

Because the sample size for both systems was the same there was an assumption for equal variances. The results of an independent-samples t-test for Table 1 (Efficiency) confirm that there is a significant difference between Systems A and System B because p-value is 0.0010 which is less than 0.05. The mean difference shows that System A (VR) has a lesser mean than System B (Text) which makes System A (VR) more efficient than System B (Text).

Table 2: Affect

Compared Systems	Means	P-value(0.05)	Hypothesis 2
System A (VR)	44.4688	0.1594	Accepted
System B (TEXT)	46.5625		

With regard to affect the results in Table 2 indicates that there is no significant statistical difference between the two group's affect scores. The results were also acquired through independent-samples t-test, therefore, there was no effect size computed for the total affect to determine which is better.

Table 3: Helpfulness

Compared Systems	Means	P-value(0.05)	Hypothesis 3
System A (VR)	41.25	0.0001	Rejected
System B (TEXT)	49.6562		

As for the systems, Helpfulness the independent-samples t-test was used to examine whether there is any statistical difference between the groups. The results indicated that there is an enormous statistical difference between the groups as the p-value (0.0001) was way less than 0.05. The difference of the means was also huge with System A (VR) have a smaller mean making it more helpful as compared to System B (Text).

Table 4: Control

Compared Systems	Means	P-value(0.05)	Hypothesis 4
System A (VR)	43.28125	0.0416	Rejected
System B (TEXT)	46.34375		

With regard to control, an independent-samples t-test was also conducted to examine if there is any significant statistical difference. The results indicate that there is a slight statistical significant difference between the groups as the p-value (0.0416) is less than 0.05. System A (VR) is slightly more controllable because its mean is less than that of system B (Text).

Table 5: Learnability

Compared Systems	Means	P-value(0.05)	Hypothesis 5
System A (VR)	41.8437	0.5256	Accepted
System B (TEXT)	42.7812		

There was no significant statistical difference between the groups in learnability as indicated by table 5. The results obtained were computed through the independent-samples t-test. For that reason, the effect size was not computed for learnability means.

Table 6: General

Compared Systems	Means	P-value(0.05)	Hypothesis 6
System A (VR)	41.4375		Rejected

System B (TEXT)	46.8125	0.0006	

Lastly, the entire system (Global/General) was tested through independent-samples t-test and the results indicated that there was a big difference in reactions to the two systems in general. Statistically, there was a significant difference between the systems. But System A (VR) performed better than System B (Text) because it had a smaller mean number. This means that System A (VR) is satisfied the users much better than the system.

VII DISCUSSION OF RESULTS

On this section, a summary of findings from this study is presented. The contribution of this research and its implication on the future work is also presented. Lastly, we conclude with the limitations and recommendations for feather research.

Results implications

The results acquired in this study implies that non-immersive virtual reality expert systems used for cattle disease diagnosis are more efficient, easy to control, easy to learn, helpful and more affect when compared to a text-based expert system. The results uncovered that the users preferred using an interactive multimedia-based system over a text only based system. Interactive multimedia-based system seemed to provide characteristics of a good interface such as clarity, concise, responsive (even though not faster than a text-based one), consistency and attractive.

Existing results comparison with our work

Even though a number of studies have been conducted comparing text and multimedia based systems, most of them were focused on images, only a few covered interactive multimedia particularly virtual reality or augmented reality. It has evidently shown that there is little to no published scholarly articles that compare of user satisfaction between text and virtual reality based expert systems for diagnosing cattle diseases.

Results limitations

Due to limited time and resources a convenient smaller sample size of 32 in each group was used [26]. There this could have could lead to a poor interpretation of results more especially the confidence level and the p-value. None the less small sample size enabled us to address the research questions in a shorter period of time. The other problem we faced was our target ordinance some of our participants did not take the study seriously. The other problem we encountered was that we were looking at the entire cattle farmers' regardless of whether one is a commercial, subsistent or aspiring and the challenge was that they were faced with different problems.

VIII CONCLUSIONS

The overall analysis as indicated by the general or global aspects showed that system A performed much better than system B even thought two aspects affect and learnability did not have any significant statistical difference. Most of the participants were very impressed with non-immersive virtual reality technology being used in agriculture for cattle diseases diagnosis. They thought the adoption of such a system could help them manage their farmers well and educate them on different cattle diseases which in return can save money and time if utilizing such a system. A few of the participants, however, showed displeasure in the perception that computers and Smartphone's are for people residing in urban areas and that other farmers might have inconveniences getting used to the new innovation.

The study also has shown that it is not easy to make conclusions about cattle farmers more especially the subsistent farmer because it consists of the population that different and heterogeneous. Nevertheless, we hope that this study will have a contribution to the body of knowledge of interface designs for expert systems in agriculture more especially cattle farming. We intend to expand our study in the future and compare three interfaces for cattle diseases diagnostic expert systems which consist of two multimedia based interfaces (non-immersive virtual reality and augmented reality) with Text using mobile platforms. We would like to examine the preference of the three interfaces in terms of gender and age.

IX ACKNOWLEDGEMENT

The study was completed as one of the objectives of an MSc thesis at Botswana International university of Science and technology who sponsored in all the endeavors of this research study. Sincere gratitude also goes to the Department of Veterinary Services (DVS) who allowed me to use their clients as my participants. Dr. Galani Malatsi also helped in everything he could in this study. Dr. Jurek Kirakowsk and the SUMI team also helped me a lot with the results. Lastly, I would like to thank my colleagues for the support they gave me.

IIX REFERENCES

[1] S. Parlangeli, O., Marchingiani, E. & Bagnara, "Multimedia in Distance Education: Effects of Usability on Learning, Interacting with Computers," vol. 1, no. 12, pp. 37–49, 1999.

[2] N. Bevan, "International Standards for HCI," *Encycl. Hum. Comput. Interact.*, vol. 55, no. May, pp. 1–15, 2006.

[3] N. Bevan and M. Macleod, "Usability measurement in context," *Behav. Inf. Technol.*, vol. 13, no. 1–2, pp. 132–145, 1994.

[4] L. Hasan, "Evaluating the Usability of Educational Websites Based on Students' Preferences of Design

Characteristics," vol. 3, no. 3, pp. 179–193, 2014.

[5] M. Hazman, A. Eldin, and A. Yassin, "Animal Knowledge Based Systems in Egypt," vol. 71, no. 21, pp. 33–38, 2013.

[6] S. S. A. Naser, A. Z. A. Ola, and I. Technology, "AN EXPERT SYSTEM FOR DIAGNOSING EYE DISEASES," pp. 923–930, 2008.

[7] P. M. N. Rani, T. Rajesh, and R. Saravanan, "Expert Systems in Agriculture : A Review," vol. 3, no. 1, pp. 59–71, 2011.

[8] S. K. Jha, "Development of knowledge Base Expert System for Natural treatment of Diabetes disease," vol. 3, no. 3, pp. 44–47, 2012.

[9] R. Prasad, K. Rajeev, and A. K. Sinha, "AMRAPALIKA : An expert system for the diagnosis of pests , diseases , and disorders in Indian mango," vol. 19, pp. 9–21, 2006.

[10] V. Z. Ogozalek, "A Comparison of the Use of Text and Multimedia Interfaces to Provide Information to the Elderly," pp. 65–71, 1994.

[11] R. Granados, "A Comparison of Two Teaching Methods for Pediatric Medication Administration : Multimedia and Text-Based Modules," 2013.

[12] J. Chen and J. Zhang, "Comparing Text-based and Graphic User Interfaces for Novice and Expert Users," pp. 125–129, 2007.

[13] T. D. Ramsey, "The Effects of Multimedia Interface Design on Original Learning and Retention," no. December, p. 107, 1996.

[14] J. Cheon and M. Grant, "Do flashy interfaces help ? Comparing learning efficiency of different interface types based on cognitive load theory," pp. 2003–2007, 2003.

[15] M. Haubner, C. Krapichler, L. Andreas, K. Englmeier, and W. Van Eimeren, "Virtual Reality in Medicine – Computer Graphics and Interaction Techniques," vol. 1, no. 1, pp. 61–72, 1997.

[16] R. Albert, Rizzo.JoAnn, Difede. Barbara , Rothbaum. J. .Martin, Daughtry. Greg, "VIRTUAL REALITY AS A TOOL FOR DELIVERING PTSD EXPOSURE THERAPY," pp. 1–22, 2005.

[17] N. Bevan, "Measuring usability as quality of use," vol. 150, 1995.

[18] T. Arh and B. J. B. Č, "A Case Study of Usability Testing – the SUMI Evaluation Approach of the EducaNext Portal," vol. 5, no. 2, 2008.

[19] Z. Mansor, Z. M. Kasirun, S. Yahya, and N. H. Arshad, "The Evaluation of WebCost Using Software Usability Measurement Inventory (SUMI)," vol. 2, no. 2, pp. 197–201, 2012.

[20] J. A. Ogolla, "Usability Evaluation : Tasks Susceptible to Concurrent Think-Aloud Protocol MASTER THESIS-TQDV30."

[21] TechSmith, "Morae Software in the Usability Testing Lab," Michigan, 2013.

[22] TechSmith, "Software Usability Testing with Morae," Michigan, 2009.

[23] G. Crow and R. Wiles, "Managing anonymity and confidentiality in social research : the case of visual data in Community research."

[24] L. Ted, Palys John, "Protecting Research Confidentiality : Towards a," vol. 21, no. 1, 2006.

[25] D. Erik and P. W. M. V. V. Cisa, "Questionnaire based usability testing," no. November, pp. 1–9, 1998.

[26] S. Goes, "Sample Size Matters," 2012.

[27] R. H. Kallet and R. R. T. Faarc, "How to Write the Methods Section of a Research Paper," no. 2, pp. 1229–1232, 2004.