Towards a methodology to conduct comparative inter-organisational usability evaluations

by

Rendani Maarten Kruger

10101145

Submitted in fulfilment of the requirements for the degree

M IT (Information Systems)

in the

FACULTY OF ENGINEERING, BUILT ENVIRONMENT & IT

at the

UNIVERSITY OF PRETORIA

Study leader:

Prof. H Gelderblom

Date of submission

2017-10-31

Declaration regarding Plagiarism

The Department of Informatics emphasises integrity and ethical behaviour with regard to the preparation of all written assignments.

Although the lecturer will provide you with information regarding reference techniques, as well as ways to avoid plagiarism, you also have a responsibility to fulfil in this regard. Should you at any time feel unsure about the requirements, you must consult the lecturer concerned before submitting an assignment.

You are guilty of plagiarism when you extract information from a book, article, web page or any other information source without acknowledging the source and pretend that it is your own work. This doesn't only apply to cases where you quote verbatim, but also when you present someone else's work in a somewhat amended (paraphrased) format or when you use someone else's arguments or ideas without the necessary acknowledgement. You are also guilty of plagiarism if you copy and paste information directly from an electronic source (e.g., a web site, e-mail message, electronic journal article, or CD ROM), even if you acknowledge the source.

You are not allowed to submit another student's previous work as your own. You are furthermore not allowed to let anyone copy or use your work with the intention of presenting it as his/her own.

Students who are guilty of plagiarism will forfeit all credits for the work concerned. In addition, the matter will be referred to the Committee for Discipline (Students) for a ruling. Plagiarism is considered a serious violation of the University's regulations and may lead to your suspension from the University. The University's policy regarding plagiarism is available on the Internet at *http://upetd.up.ac.za/authors/create/plagiarism/students.htm*.

I (full names & surname):	Rendani Maarten Kruger
Student number:	10101145

Declare the following:

1. I understand what plagiarism entails and am aware of the University's policy in this regard.
2. I declare that this assignment is my own, original work. Where someone else's work was used (whether from a printed source, the Internet or any other source) due acknowledgement was given and reference was made according to departmental requirements.
3. I did not copy and paste any information directly from an electronic source (e.g., a web page, electronic journal article or CD ROM) into this document.
4. I did not make use of another student's previous work and submitted it as my own.
5. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his/her own work.

Signature

Date

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Prof. Helene Gelderblom, for her continuous guidance and support throughout the challenging and yet often delightful journey to complete this dissertation. I would further like to thank my fiancé, Charné Janse van Rensburg, for her motivation and handholding that got me through the tough times associated with this research.

The debates, continues feedback and challenging thoughts from my friend, Jacques Brosens, are also acknowledged to have contributed to the successful conclusion of this dissertation.

This dissertation was accomplished in affiliation with the Department of Informatics, University of Pretoria as well as CAIR, CSIR Meraka Centre for Artificial Intelligence Research.

ABSTRACT

The competitive analysis of the usability of software products has been shown to be a useful tool to encourage the adoption of the user centred design philosophy within organisations. Furthermore, the enhanced usability of a customer facing software interface can be a source of competitive advantage for organisations. The usability of systems has also been shown to encourage the effective and efficient completion of tasks, which may have a significant influence on the operations of an organisation. With all this in mind, a methodology to compare the usability of software products between organisations may be useful. However, after a literature survey, no such methodology was found. This research study develops a methodology to conduct comparative inter-organisational usability evaluations, which is done with the use of design science research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

This dissertation presents the development and evaluation of a methodology to conduct Comparative Inter-organisational Usability Evaluations (CIUEs). This study is within the subject area of Informatics (information systems) and studies phenomena within Human-Computer Interaction (HCI), particularly usability evaluations. Usability is defined by the international standards organisation as “the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments” (Bevan, 2009a, p. 14).

User Centred Design (UCD) is a broad term that describes the design process where the end user of a product influences its design (Abras, Maloney-Krichmar, & Preece, 2004). Even though the application of UCD and HCI practices in information systems has been advocated by the research community for a long time (Donald & Draper, 1986), the adoption of the UCD philosophy and other HCI practices has been limited within organisations in South Africa (Pretorius, Hobbs, & Fenn, 2015) and various other countries world-wide (Lee, Kim, & Choi, 2016; Wilkinson & De Angeli, 2014).

This adoption could be encouraged if management ensures that UCD is part of business strategy within organisations and supported at a higher level (Venturi, Troost, & Jokela, 2006). Venturi et al. (2006) found that competitive analysis of the usability of software products and the effective communication of the outcomes of UCD phases within organisations will help with the UCD adoption process.

During the literature survey, no formal methodology was found that could be used to do competitive usability analysis on software products by comparing them to similar products at competing organisations. This study works towards a methodology for this.

1.2 THE RESEARCH PROBLEM

For most enterprises, business competition has become only a click away. Where in the past businesses used a physical store front to attract customers, since the dot-com bubble, this is often done through websites or other technologies (Buenstorf & Fornahl, 2009).

In physical stores, attractive store fronts are presented to encourage customers to walk into their doors and not refrain from entering the store due to the more welcoming nature of a competing business. Businesses may have optimised the way the store looked by making it more alluring than that of their competition. To do this, they compared how their store looks as opposed to their competition (Swinyard, 1993).

With the rise of technology, it has become easier for a customer to opt for a competing organisation. If a customer has trouble using a computer interface or simply likes a competing organisation’s computer interface more, then that customer may simply opt to use a competing organisation’s information system. Should it then not be of concern for an organisation to ensure that the computer interface presented to its customers is better than that of its competition?

Furthermore, a large range of tasks that people at organisations conduct have become technologically dependent (Stein, Jensen, & Hekkala, 2015). It may be beneficial for an organisation if these tasks happen in an effective and efficient manner as it may save time and money and may lead to greater profit margins. The usability of the interfaces of the technology that people use to complete their tasks at organisations have an impact on the effective and efficient completion of those tasks (Stein et al., 2015). A methodology to conduct a Comparative Inter-organisational Usability Evaluation (CIUE) may be a useful tool in determining whether the usability of such technologies is optimal.

It has long been shown that the adoption of UCD and HCI practices improves the user experience and actual use of information systems (Donald & Draper, 1986). Although this is the case, the adoption of UCD and HCI practices is limited (Lee et al., 2016; Wilkinson & De Angeli, 2014). The comparison of information system usability may encourage the adoption of these practices (Venturi et al., 2006).

1.3 RESEARCH QUESTIONS AND OBJECTIVES

The main objective of this study is to present a methodology to conduct CIUEs. It is also a goal of this study that the proposed methodology could be used to convince organisations to invest in the usability of their information systems. Furthermore, the study aimed to identify comparable indicators of usability and to work towards the benchmarking of usability. However, no data was collected regarding the benchmarking of usability, as such benchmarking formed part of the literature that was surveyed as an initial step to associate CIUEs to benchmarking.

The main research question of this study was:

How can inter-organisational comparative usability evaluations be compiled into a useful methodology?

This was elaborated with the following sub-questions:

1. How can the information systems at organisations be compared in a manner that may convince organisations to invest into usability optimisation?

2. Which indicators of usability are objectively comparable?

1.4 RESEARCH METHODS

This research was based on the Design Science Research (DSR) paradigm. DSR typically adopts a pragmatic research approach to develop artefacts that are innovative and solve real-world problems (Hovorka, 2009). It was used in this study to develop a methodology to conduct CIUEs. This was done by following the guidelines for DSR presented by Kuechler and Vaishnavi (2008). As per these guidelines, the phases of DSR are 1) awareness of the problem, 2) suggestion of an artefact, 3) development of the artefact, 4) evaluation of the artefact and 5) presentation of the final artefact.

To implement the suggestion, in the development and evaluation phases of this study, a multiple-case study design was used (Yin, 2013). Eight South African organisations were included in the development and the evaluation of the methodology to conduct CIUEs. Three CIUEs were conducted on competing products of firstly three telecommunications organisations, then two insurance companies and finally three airlines. An eye tracker was used in each case to collect data relating to specific comparable usability indicators. The outcome of each CIUE was a report that illustrated the results of the comparison.

When the data collection and the data analysis associated with the first case were completed, a methodology to conduct CIUEs was presented. The two cases and the interviews that followed were used to evaluate and refine the methodology initially presented by the first case. With the completion of this process, a final version of the methodology to conduct CIUEs was produced.

1.5 CONTRIBUTION

The primary contribution of this study is a methodology to conduct CIUEs. From literature, comparative studies have been shown to improve the possibility for the development of benchmarks (Cragg, 2002; Pearlson & Saunders, 2009; Petuskiene & Glinskiene, 2015). Competitive studies of the usability of software products such as CIUEs may also encourage the adoption of UCD within organisations (Venturi et al., 2006). A further discussion as to why a CIUE may be useful or beneficial to an organisation continues in Section 2.4.

Checkland (1981) describes a methodology as a set of principles of method, which in any situation should be reduced to a process uniquely suited to that situation. A methodology is formulated in a generic manner to allow for the application of the methodology in diverse scenarios (Checkland, 1981). The distinction between a methodology and a method thus lies in that a methodology is universal and can be applied in different scenarios to produce appropriate methods.

In the course of selecting a methodology, it is useful to investigate possible philosophies, approaches, strategies, time horizons and methods (Saunders & Lewis, 2009). The investigation of these elements has been shown to be useful in defining research methodologies because it may help to identify the developed methodology as useful for application (Easterby-Smith, Thorpe, & Jackson, 2012). When creating a methodology (whether it is a research methodology or one developed for non-research purposes), it is therefore useful to define a philosophy, approach, strategy, time horizons and methods as this will guide the selection and application of the methodology.

The methodology that this dissertation presents, was thus generically defined to enable the selection of custom methods for unique application domains. An appropriate philosophy, approach, strategy, time horizon and methods were identified in the development of the methodology that will assist in the selection of the methodology.

A CIUE may encourage the investment into the usability of information systems (Venturi et al., 2006). The enhanced usability of an information system has been shown to be beneficial for organisations in various ways, including more effective and efficient completion of tasks by employees and more frequent use of information systems by customers (Shneiderman, 2010). Such a methodology may thus have a range of practical contributions regarding the enhancement of usability for such organisations.

The study will contribute a comparable set of usability indicators, which can be used in the future to conduct comparative studies of usability other than CIUEs. Further secondary contributions are the CIUEs that were done as part of the development and evaluation phases of this study. The first CIUE compared the usability of information systems at three mobile communications companies, the second CIUE compared the usability of information systems at two medical insurance companies and the third CIUE does so for three aviation companies. These three CIUEs are described in Table 1.1. The results will be given to the organisations that were involved in the research.

Table 1.1: The three CIUEs

CIUE case	Purpose	Associated industry
1 (pilot case)	Development of the methodology	Mobile communication
2	Evaluation and refinement of the methodology	Medical insurance
3	Evaluation and refinement of the methodology	Aviation

1.6 DEMARCATION

The first limitation in scope is that this study endeavoured to present a single methodology. Further research could be done to present competing methodologies. Furthermore, there are multiple ways that usability studies could be conducted. This study included one method of conducting usability studies that uses quantifiable data and thus contributes to the objectivity of a CIUE.

The methodology presented by this study could be used to conduct a comparison between elements of the usability of information systems at different organisations. The differentiator is thus "the organisation" and this study will mark a boundary here. Other differentiators were not investigated, although the application of CIUEs to select software products may be useful to research (Fenton & Bieman, 2014).

The methodology was only tested on the public facing information systems of eight South African organisations that constituted the multiple-case study. The methodology could be applied to other organisations and other forms of information systems to test it further.

1.7 AWARENESS OF THE PROBLEM AND A SUGGESTION FOR A SOLUTION

The methodology used to conduct the research presented in this dissertation is DSR. Gregor and Hevner (2013) suggest that, in DSR, the awareness of the problem and a suggestion for a solution should be highlighted. Kuechler and Vaishnavi (2008) suggest that this happens in the introductory chapter of a research paper or dissertation. Following this advice, the problems and suggestions that emerged thus far can be summarised as follows:

1. It has become easy for the customers of businesses to opt for competition in businesses that have an online or otherwise digital presence or store front (Srinivasan, Anderson, & Ponnavolu, 2002). However, no methodology exists to compare the usability of a computer interface to that of a competitor to evaluate whether it is better and whether a customer may rather opt for the competing business for reasons relating to usability.
2. The possibility exists for organisations to create benchmarks when comparative studies are conducted (Petuskiene & Glinskiene, 2015). Benchmarking could be used to identify opportunities for improvement, define best practices and create a competitive environment (Hasan, Morris, & Probets, 2013). No formal methodology was found in literature to compare the usability of information systems and therefore organisations will find it difficult to create benchmarks for the usability of information systems in a standardised manner.
3. The competitive analysis of the usability of systems has been shown to increase the likelihood that UCD will be adopted within organisations (Venturi et al., 2006), however no methodology was found that could be used to do such competitive analyses.

The suggestion made to address the problems above is the development of a methodology to conduct CIUEs.

1.8 DISSERTATION OUTLINE

This dissertation contains seven chapters. Chapter 1 is the introduction chapter, which gives a background to the dissertation, presents the research questions, states what the expected contribution of the study, outlines the boundaries of the study and gives the awareness of the problem and suggestion as described in the approach to DSR by Kuechler and Vaishnavi (2008).

Chapter 2 is the literature review chapter. Here an investigation was done into literature relevant to this study. Various topics were investigated, including usability, methodologies, organisations, comparative studies, methods to test usability and current comparative studies that organisations conduct. In chapter 2, the suggestion for a methodology to conduct CIUEs is further elaborated and the relevance of the methodology is investigated.

Chapter 3 is the research methodology chapter, which describes the research approach used and aligned it with a research philosophy, strategy, time horizons and data collection methods. This chapter also introduces the DSR approach that was followed and the reasons why it was selected. Chapter 4 discusses the pilot case study that was used to develop the methodology to conduct a CIUE. The CIUE involved the competitive analysis of the usability of information systems of three telecommunications companies in South Africa. Here the suggestion for the artefact was further developed.

Chapter 5 elucidates the evaluation and further development of the methodology to conduct CIUEs, which was produced by this study. Here, CIUEs were done for two other cases, and were then evaluated through interviews. Chapter 6 presents the final methodology to conduct CIUEs based on recommendations made in Chapter 5. Chapter 7 is the concluding chapter where the research findings are summarized, the knowledge contribution is evaluated, future research is suggested and the applied research methodology is reflected upon. The layout of the chapters is depicted in Figure 1-8.1.

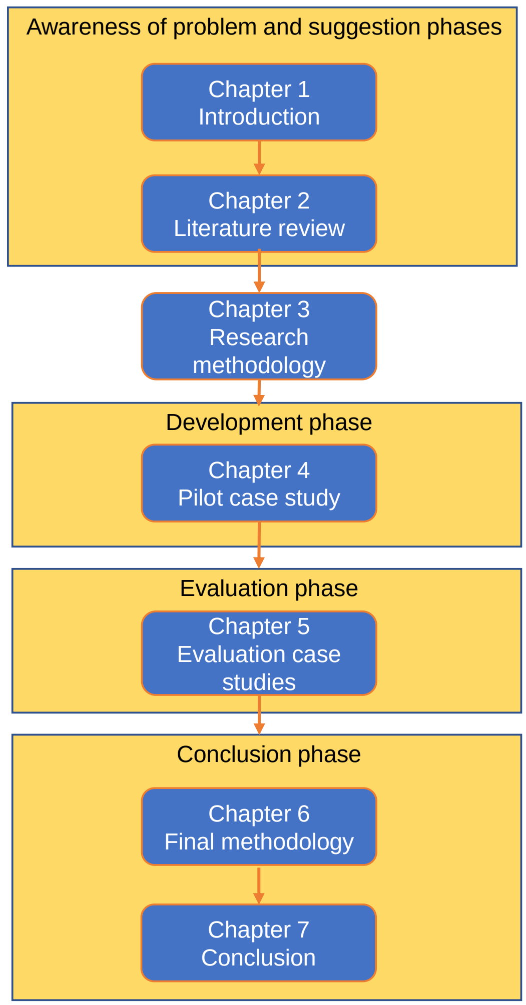


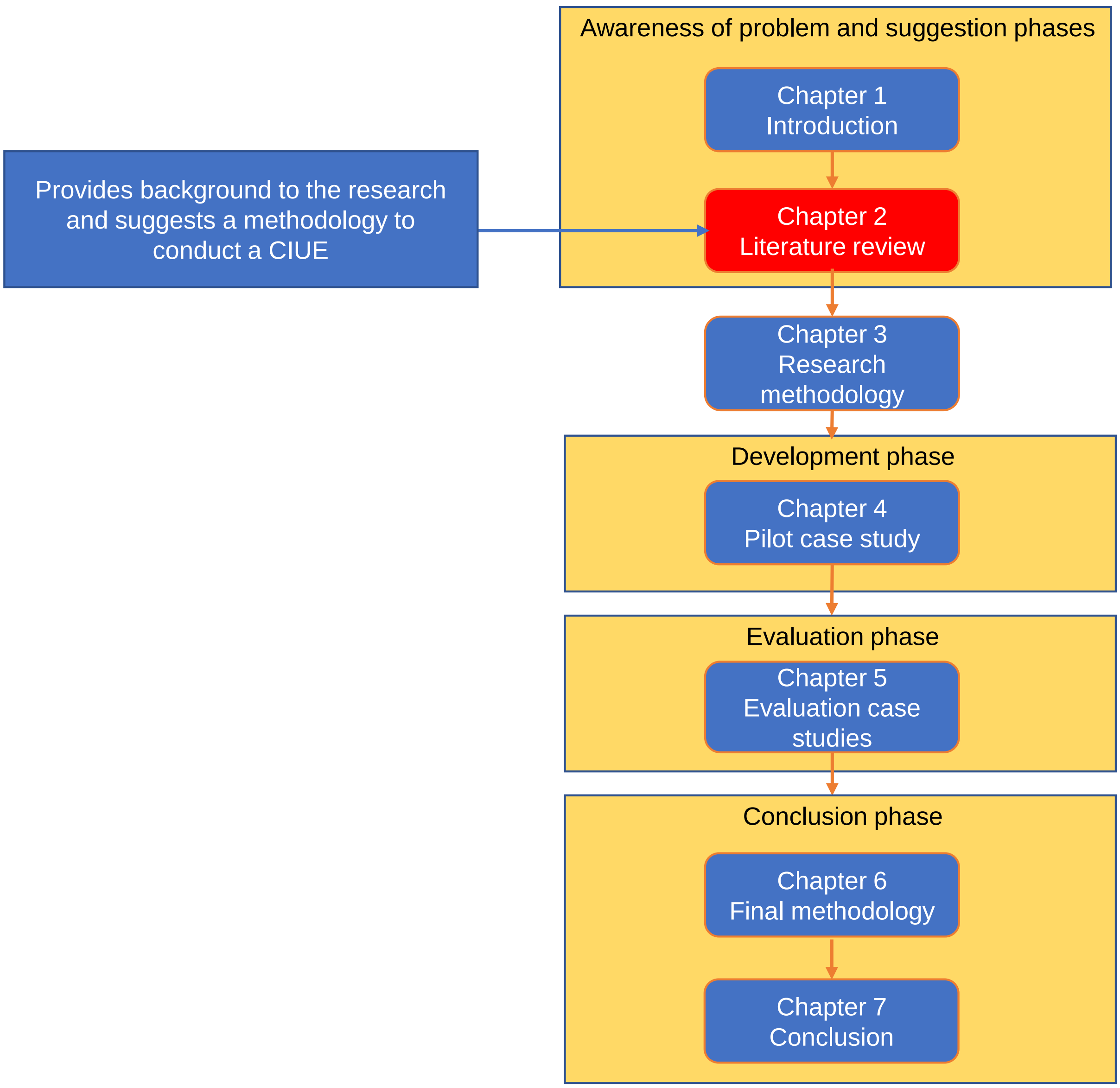
Figure 1.8.1: Layout of chapters.

1.9 CONCLUSION

This chapter served as an introduction to this study. The research questions were stipulated in this chapter. The main research question of this study is: “how can inter-organisational comparative usability evaluations be compiled into a useful methodology?” Design science research was introduced as the approach that was followed in designing a methodology to conduct CIUEs. The boundaries of the study were also discussed and as per Kuechler and Vaishnavi (2008) the awareness of the research problem was outlined and a suggestion for the research artefact was made, namely a methodology to conduct CIUEs.

CHAPTER 2: LITERATURE REVIEW AND BACKGROUND TO THE PROBLEM

The position of chapter 2 in the dissertation



2.1 INTRODUCTION

This chapter investigates literature relevant to the development of a methodology that could be used to conduct a Comparative Inter-organisational Usability Evaluation (CIUE). In Section 2.2 the Technology Acceptance Model (TAM) by Davis Jr (1986) is discussed as it is the theoretical backing of this study. Further discussions in the chapter relate to giving context to the outcomes of this dissertation. For example, an investigation is done into the South African information system and usability environment. A discussion around usability evaluation methods that will be used to conduct CIUEs is also provided in this chapter.

The contextualisation of this dissertation, done in this chapter, highlights the gap in knowledge that is addressed. This is emphasized using an explanation of what a methodology to conduct CIUEs is and why it may be useful for organisations. Additionally a study of similar research artefacts is done to address Gregor and Hevner (2013) requirements for Design Science Research (DSR).

2.2 THE TECHNOLOGY ACCEPTANCE MODEL

The theoretical backing of this study is the technology acceptance model (TAM) developed by Davis Jr (1986). TAM is depicted in Figure 2.2.1.

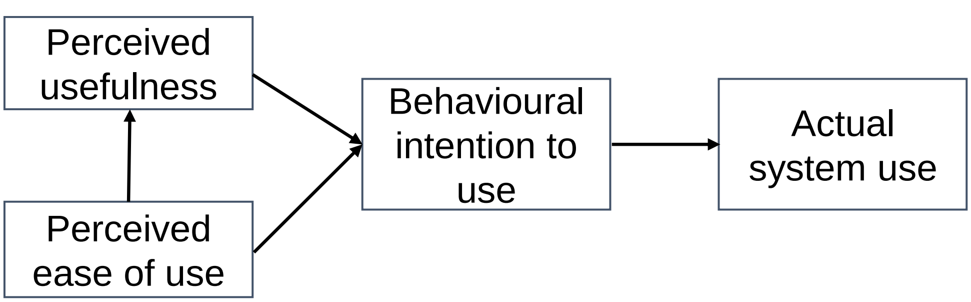


Figure 2.2.1: Technology acceptance model by Davis Jr (1986)

The TAM states that the perceived ease of use and the perceived usefulness of an information system will determine an individual's behavioural intention to use that system and ultimately that user's actual system use (Davis Jr, 1986). TAM also states that the perceived ease of use of a system will have a direct impact on the perceived usefulness of the system (Davis Jr, 1986).

Perceived usefulness of a system is described as the extent to which users believe that it will help them perform their job better (Davis, 1989). Perceived ease of use is related to perceived usefulness – a user may perceive the system as useful, but if they find it too hard to use the performance benefits of usage can be out-weighed by the effort of using the system and then the user will not use the information system (Davis, 1989). If a user does not perceive that the system is useful or does not perceive that the system is easy enough to use, this will reduce their behavioural intention to use the system and will limit their actual use of the information system.

This dissertation focuses on the approach used to conduct CIUEs. A methodology whereby CIUEs could be conducted was suggested and tested through a range of cases of application and interviews. The reasoning behind this research is based on the TAM, specifically because a goal of usability enhancement is to improve the actual use of an information system. The research question posed in Section 1.3 is: "How can inter-organisational comparative usability evaluations be compiled into a useful methodology?". It relates to TAM in that a CIUE may lead to the objective assessment of users' perceptions of an information system's ease of use, thereby providing organisations with insight into how to strengthen users' intention to use the system and finally improve its actual use.

The adoption of User Centred Design (UCD) and HCI practices improves the user experience and usability of information systems (Pretorius et al., 2015; Wilkinson & De Angeli, 2014). Although this is the case, the adoption of UCD and HCI practices in organisations has been limited (Wilkinson & De Angeli, 2014). The comparison of information system usability may encourage the adoption of these practices (Venturi et al., 2006). A methodology that could be followed to conduct such comparative studies may thus contribute to the perceived usefulness and ease of use of information systems as the adoption of UCD practices may be encouraged from the comparison. Following TAM, this will ultimately contribute to the actual use of an information system.

2.3 USER CENTRED DESIGN

User Centred Design (UCD) as mentioned in Chapter 1, is a broad term to describe the design process where the end user has an influence on the design (Abras et al., 2004). In UCD, the efficient and effective possibility of use, user environment, user opinions, user characteristics, tasks and workflow of products are given extensive consideration in the design process. The adoption of UCD practices in organisations has been limited although its use has been shown to be beneficial in the improvement of the user experience of information systems (Pretorius et al., 2015; Wilkinson & De Angeli, 2014).

The involvement of users to influence the design of information systems has also shown to increase the number of breakthrough ideas in research and development projects (Pallot, Trousse, Senach, & Scapin, 2010). Furthermore, the production of concepts and scenarios regarding the adoption of information systems, that may otherwise not have been thought of, was typically also improved when UCD was adopted in a number of projects (Pallot et al., 2010). The limited adoption of UCD has been attributed to the business potential in the use of an information system and technological limitations being considered before users are involved (Pretorius et al., 2015; Wilkinson & De Angeli, 2014).

It was also shown that if competitive analysis is done on the usability of software products it may increase the adoption of UCD (Venturi et al., 2006). There is a strong emphasis on the usability of information systems when UCD is adopted. Some of the basic principles that relate to UCD in terms of usability are as follows (Norman, 1988):

1. It should be made easy to determine which actions are possible for a user of an information system at any time. This means that there should be a manner to escape the current workflow, a manner to go back and a manner to go to the next step. The architecture of information is also important here as the structure of links and the language used will influence the perception of the possible steps that could be taken.

2. Things should be made visible. This includes actions that could be taken and the actions that have already been taken. There are a variety of considerations to be made here. The current trends in design will have an influence on the way an information system is viewed and which elements seem to be visible and which not. The information architecture and the language used on the interface have a similar effect. The architecture of information refers to the structural design of environments that contain shared information (Shneiderman, 2010).

3. It should be built into the design that the user can easily examine what the current state of task completion is. In a wizard for example, there should be an indication of the number of steps completed and how far the user is from the total completion of the task. It should further be clear what the effect of actions was.

4. It should be intuitive to follow natural mappings between intentions and the required actions, between actions and the resulting effect, and between the information that is visible and the interpretation of the system state.

When these principles are followed, the user is put at the centre of the design process (Norman, 1988). The adoption of UCD will improve the ease of use of systems and may also have an influence on the perceived usefulness of the system. From a TAM perspective, the adoption thereof may thus improve the actual use of systems at organisations.

With all this in mind, UCD is often regarded as a philosophy of information system design. When UCD is applied, it influences the entire design process.

2.4 HUMAN COMPUTER INTERACTION AND USABILITY

The domain of this study is Human Computer Interaction (HCI), which is a field of research concerned with the design, implementation and evaluation of interactive systems in the context of the user's task (Dix, 2004). "Human" in HCI does not refer to an individual human being or a group of people, it refers to whoever is trying to get a task done using a technological interface. The "computer" refers to any technology that the human has an interaction with when performing a task and the interaction refers to the communication between the human and the computer.

HCI research should be focused on the following generic areas to be fruitful (Shneiderman, 2010):

1. Reducing the anxiety and fear of computer usage.
2. Working towards making it easier for computer novice users to become experts.

3. The designing of tools to assist with the development and specification of user interfaces.
4. Areas that relate to the direct manipulation of users.
5. The vast amount of different input devices and their interfaces.
6. The architecture of information, especially to improve navigation.

The study of usability in HCI relates to a number of the research directions above as it is concerned with the effectiveness, efficiency and satisfaction with which users achieve goals in computer systems (Bevan, 2009a). The improved usability of an information system may reduce the anxiety and fear of computer usage, improve learnability, could direct users more concisely and may improve the navigation of systems (Shneiderman, 2010).

Usability relates to the quality of computer interfaces, particularly regarding to how easy they are to use (Nielsen, 2003). A user's interaction with an information system is not only bound to that user's interaction with information systems. The user's interaction with information systems may include the interaction with human system actors (Stair & Reynolds, 2013) such as secretaries or clerks. The term user interface, however, commonly refers to the technological elements through which users interact with computer systems (Laudon & Laudon, 2013). The quality measure of interactions that users have with the broader information system, other than computer systems, is not covered under the term usability.

To improve the usability of information systems, designers should consider the end user and therefore employ UCD (Norman, 1988). It should be remembered that what may be an effective design for one group of users may be an ineffective design for the next group (Majchrzak, Markus, & Wareham, 2016). For example, computer science practitioners may prefer the use of a terminal interface above a graphical user interface as the shortcut keys are more intuitive to them. Regarding the practical evaluation of a design, it is recommended that the time to learn how to use an information system, the speed of user performance, the rate of system errors, the rate of user errors, user retention over time and the subjective satisfaction of the user is considered (Shneiderman, 2010).

One way to design for the usability of an information system is to consider theories and guiding documents (Zapata, Niñirola, Idri, Fernández-Alemán, & Toval, 2014). Various organisations that develop technologies have compiled sets of generic usability considerations. Typically, these documents address graphics, screen layout, input and output devices, sequencing of actions, and training (Shneiderman, 2010). Complying with these recommendations may improve the time a user takes to learn how to use an information system, the speed of user performance, the rate of user errors, user retention over time and the subjective satisfaction of the user. However the dangers of only following principles of design is that the design is not tested for a variety of user groups and that the design may not be very unique (de Queiroz Pierre, 2015).

It is also important to test information system designs. Because it is an expensive process to change an information system when it is complete, prototypes should be created and tested before a design is accepted (Helander, 2014). Various software tools are available to create prototypes of information systems. They vary in the amount of detail they produce. It may be useful to design prototypes at various levels of detail to inform the design process at different stages.

Expert evaluation and usability testing with actual users are also recommended during the design process and once the process is completed. As stated earlier, an effective design for one group of users may be an ineffective design for another group of people. This is a problem that can typically not be addressed using generic usability principles nor by using untested prototypes (Shneiderman, 2010). To test the usability of a prototype or an information system different methods of usability evaluations may be used (refer to 2.9 for a discussion on various methods and how they may be appropriate for use in a CIUE).

Another field that falls under HCI that relates to usability is User Experience (UX). UX encompasses all the interactions that a user has with a company, its services and its products (Nielsen & Norman, 2014). The usability of the user interfaces present at companies has an influence on the overall user experience that customers will have when interacting with the companies as it is a measure of the quality. This discussion regarding UX is only for contextual purposes and does not form a further part of this study.

2.5 THE ROLE OF USABILITY EVALUATION

The goal of usability evaluation is to test and improve the usability of interfaces on computer systems (Nielsen, 2003). There are several ways to conduct usability evaluations, including focus group sessions with users and individual user evaluations. A first step to improving the overall usability of a computer interface is typically to find usability errors. The benefits of enhanced usability can be summarised as follows (Nielsen & Gilutz, 2013):

1. Reduced training cost with relation to the information system.
2. Limited user investment risk.
3. The enhanced performance of users that are completing a task.
4. Enhanced user efficiency and satisfaction.
5. Lower personnel cost for organisations.

6. Reduced software development cost and length of the development lifecycle.
7. Easier to use online interfaces for enhanced customer interactions.
8. Improved competitive advantage.

Further motivations for the adoption of usability practices come to the fore when considering the environment of the technology. Retention of how to use a life critical system such as an air traffic control system is important because it will need to be applied effectively and efficiently (Shneiderman, 2010). In commercial use, a reduction in system transaction time could have direct cost benefits. For example, suppose an administrator registers customers at one minute per customer, if the usability of that system is improved, that time may change to 30 seconds per customer and the administrator can register double the number of customers. The usability of entertainment technology may influence the perceived value of that product and thus also a customer's willingness to buy the product (Shneiderman, 2010).

If usability design and evaluation practices are not executed correctly, it may have negative consequences for stakeholders involved. This includes inflated costs of usability evaluation and wasting time on studies that do not bring benefit to the organisation involved (Dumas & Redish, 1999; Shneiderman, 2010). It is therefore desirable to do proper investigations before investing into the usability of computer systems.

2.6 INFORMATION SYSTEMS AND USABILITY IN SOUTH AFRICA

An extensive evaluation of the South African information system industry falls outside the scope of this study, but an overview of this environment is useful to place the research into context. Some areas in South Africa have first world-like exposure to technology while others do not. Typically, information systems are widely adopted by large South African organisations (Van Grembergen & De Haes, 2017). Information systems can be valuable to organisations, but the extent and the dimensions of this value may differ depending on internal and external organisational factors (Melville, Kraemer, & Gurbaxani, 2004). The limited exposure to technology in some South African areas serves as a limiting factor in extracting value from information systems. This limited exposure has been attributed to a lack of availability of low level technical requirements, such as electricity or an internet connection.

Mobile phone adoption has been phenomenal in Africa, and initial growth forecasts for South Africa have been greatly exceeded (Ochara & Mawela, 2015). South African organisations also use various forms of information systems including ERPs, mobile apps, banking applications, decision support systems, data warehouses and social media (Van Grembergen & De Haes, 2017). Despite this, there has been a limitation in citizen participation in e-government and other social initiatives (Ochara & Mawela, 2015).

A range of initiatives that attempt to enable small to medium sized enterprises in South Africa to participate in the global economy using information systems have also been conducted (Francis & Willard, 2016; Mann, 2017). Some research, however, reports that the adoption of modern Information Communication Technology (ICT) in South Africa and other developing countries has been limited due to the cost of internet and the limited bandwidth of mobile connections (Brown & Molla, 2015).

The factors that influence the benefit of information systems in organisations are affected by various environmental factors such as economics, politics, society and supplier availability. Different organisations also have different strategies regarding the leveraging of benefits that information systems may bring (Melville et al., 2004). Some organisations use big data analytics to determine future environmental influences and some change their digital marketing strategy to change their business environment (Armstrong, Kotler, Harker, & Brennan, 2015; Kwon, Lee, & Shin, 2014).

Furthermore, information systems have been shown to have an influence on business level strategy because they effect mechanisms of value generation (Drnevich & Croson, 2013). It is, however, regarded of utmost importance to have a strategy to fit information systems into an organisation (Saunders & Lewis, 2009).

The leveraging of the potential benefit that information systems may bring for an organisation is a complex task, especially in a developing market such as some of the markets present in South Africa (Francis & Willard, 2016; Mann, 2017). Except for a privileged minority in South Africa, most learners under-perform severely in literacy and mathematics at their grade levels (Spaull, 2013). Furthermore, the quality of education in South Africa has been ranked worse than many third world countries in Africa and only twelve percent of learners qualify to study at university after grade 12 (Spaull, 2013).

A manner to enhance the adoption of technology in an environment where education is lacking may be to adopt the UCD philosophy in information system design (Davis, 1989; Sayed, Badroodien, Hanaya, & Rodríguez, 2017). Furthermore, the literacy levels of system users should be a consideration in usability studies (Venkatesh, Hoehle, & Aljafari, 2017) especially in an environment where literacy levels are generally low. This will require competencies in related fields such as HCI, business analysis, systems analysis and systems development.

With regard to HCI competencies in South Africa, Beukes, Gelderblom, and van der Merwe (2016) found that designers of websites at large organisations in South Africa disregard basic web design principles. Beukes et al. (2016) conducted interviews with designers and confirmed that they were not well informed about usability guidelines and that they are not concerned with their lack of knowledge. This may indicate a lack of concern by the large organisations where the interviewed designers were employed to adopt usability practices as part of their application design and development process.

Pretorius et al. (2015) evaluated the UX landscape in South Africa. They found that the lack of UX buy in, time constraints, the lack of skilled UX staff, process challenges and budget were regarded as the top challenges by the UX practitioners surveyed. The lack of concern regarding UX practices was further highlighted by the lack of UX activities conducted at organisations (Beukes et al., 2016). It was shown that only half of the organisations surveyed by Pretorius et al. (2015) had any form of formal reporting regarding their UX capability and furthermore there was a lack of training in two thirds of the organisations surveyed. It was concluded that this also showed that there was a lack of UX education in South Africa.

Gelderblom, Adebesin, Brosens, and Kruger (2017) suggested that eye tracking can be used as an effective tool to teach HCI in tertiary institutions. They also found that there was a lack of adoption of UX practices in South Africa and one of the key reasons for this was a lack of education and therefore a shortage of skills.

From the above, it is shown that the development in the information system sector of South Africa still requires a lot of work, although technology adoption has been good in certain areas. The limitation of adoption in other areas has been attributed to various factors including the cost of internet and the limited bandwidth of mobile connections (Brown & Molla, 2015).

It was also found that there is a lack of adoption of usability design principles and other HCI practices by South African organisations (Beukes et al., 2016; Pretorius et al., 2015). If information system professionals in South Africa become more concerned with the usability of computer interfaces, it may advance the quality of information system products and the adoption of technology in the area.

2.7 INFORMATION SYSTEM RELATED METHODOLOGIES

The research artefact produced by this research was a methodology to conduct a CIUE. The concept of a system development methodology was investigated by Avison and Fitzgerald (2003). Avison and Fitzgerald (1999, p. 254) gave a general definition of a methodology as "a recommended series of steps and procedures to be followed during developing an information system". Various Information Systems Development Methodologies (ISDMs) were developed that work differently and suit different scenarios.

The methodology produced in this study is not an ISDM but rather a methodology to evaluate the usability of information systems and compare the evaluations between organisations with similar systems. An investigation into ISDMs may nonetheless be useful as a lot of research has been done into ISDMs, making it a mature field. ISDMs are related to CIUEs in that both use methodologies to achieve a goal.

ISDMs are categorised as process orientated, blended, object orientated, rapid development, people orientated or organisational orientated (Avison & Fitzgerald, 2003). An example of a process orientated ISDM is the Structured Analysis, Design, and

Implementation of Information Systems (STRADIS). This methodology was first introduced by Gane and Sarson (1979). It is concerned with the selection and organisation of program modules and interfaces to solve problems.

The blended category of methodologies adopts a pragmatic approach and does not focus on one element of systems development (Avison & Fitzgerald, 2003). An example of a blended methodology is the Structured Systems Analysis and Design Method (SSADM). Weaver, Lambrou, and Walkley (1998) provide a description of SSARM in which application development is divided into various stages, steps and tasks, and guidelines are given as to how projects should be described to suit the management of projects.

The object orientated category of methodologies adopts a philosophy where real world artefacts and the relationships between them are defined using software classes, objects and processes (Avison & Fitzgerald, 2003). An example of an object orientated systems methodology is Object Orientated Analysis (OOA) (Coad & Yourdon, 1991).

The rapid development category of systems development methodology emphasizes the speed from conception to development (Avison & Fitzgerald, 2003). An example of this is Extreme Programming (XP) where no specific structure is prescribed. This is determined by the development team. It is typically used for smaller applications and is known for producing good software quickly. The drawbacks are that there is no documentation and the success of the project is very dependent on the quality of the development team.

The people orientated category of ISDMs encompasses the social technical view that, in order to be effective, the social and organisational factors of the application should determine its technological requirements (Avison & Fitzgerald, 2003). An example of this is is the Knowledge Acquisition and Documentation Structuring (KADS), which was an outcome of a European research project (Schreiber, Wielinga, & Breuker, 1993).

The final category of ISDMs is the organisational category (Avison & Fitzgerald, 2003). In this category, organisational issues are prioritised above technical ones. An example of this is Information Systems work and Analysis of Change (ISAC) (Lundeberg, 1982). ISAC adopts a number of phases that put emphasis on organisational issues.

Various methodologies exist to assist in the development of information systems. It can also be seen that these methodologies are produced in a generic manner to allow for the application of the methodology in diverse scenarios, they are in other words not solely applicable to a specific type of information system. A methodology is a set of principles of method, and is thus universal and can be applied in different scenarios to produce appropriate methods as is the case with ISDMs (Checkland, 1981).

2.8 WHY A CIUE MAY BE USEFUL FOR AN ORGANISATION

This section discusses the reasons why a CIUE may be beneficial for an organisation. Comparative studies conducted by organisations are a point of discussion in this section to find out where a CIUE may fit in the context of an organisation and highlight how comparable evaluations are useful in organisations. Furthermore, competitive advantage for organisations is regarded as one of the main outcomes that could be drawn from CIUEs. A discussion into what competitive advantage is and what competitive advantage may be gained from usability enhancement is discussed in Sections 2.8.2 and 2.8.3.

Another application of a CIUE may be to aid in the selection of information systems. Jadhav and Sonar (2009) found that there is a need for software selection methodologies and evaluation techniques to assist decision makers in software selection. This is further discussed in Section 2.8.4.

Benchmarking has also been shown to aid organisations in various ways (Cragg, 2002; Pearlson & Saunders, 2009; Petuskiene & Glinskiene, 2015). A discussion around how a CIUE could be used in benchmarking and how organisations currently apply benchmarking is done in Section 2.8.5. Finally, the buy in for UX practices has been regarded as limited. The application of CIUEs to aid with this problem is discussed in Section 2.8.6.

2.8.1 HOW ORGANISATIONS USE COMPARATIVE EVALUATIONS

Organisations often conduct comparative evaluations or studies to benchmark (Glackin, 2013). Benchmarking studies are further discussed in Section 2.8.5. Other than benchmarking tests, organisations typically conduct business potential and feasibility studies before inception (Glackin, 2013). In such studies, organisations endeavour to find out what the current state of the micro and macro environments are to see if there are potential business opportunities. They can determine whether there are gaps in the market and, if so, exactly which gaps they could address (Glackin, 2013).

To do this, an organisation will compare products or services that are already available in the market to the products or services that they plan to deliver. This may then inform the organisation of the potential services or products that may be profitable for them to deliver (Glackin, 2013).

Furthermore, a typical organisation may compare the current manner of operations to alternative ways in which these operations could be conducted. If the alternative promises to be an improvement and it is feasible to implement the better method of operations, the organisation may do so. This may have a variety of influences on that organisation, including greater profits or business growth (Slack, 2015). It is challenging for organisations to determine the value they will derive from making a change to an information system or its environment (Pearlson, Saunders, & Galletta, 2016). This leads to challenges regarding the study of the effectiveness of operations in software development teams. A/B testing, which is also called split testing or bucket testing, is a method of comparing different versions of an information system to determine the effectiveness of a change or the effectiveness of the operations of a development team (Xu, Chen, Fernandez, Sinno, & Bhasin, 2015). This is a comparative study conducted by organisations that is similar to CIUEs in that information systems are compared, but the different systems are versions of the same one within the same organisation. An A/B testing platform has been used at various organisations including LinkedIn, Facebook, Twitter etc (Xu et al., 2015).

When an organisation compares its accomplishments to goals it had at conception or at a prior strategic stage, they may realise that those goals have not been achieved. If that is the case and it does not seem likely to change in the future in a way that will cause the organisation to achieve those goals, it may have to close (Li & Sun, 2013). This may happen when it has become impossible for a business to compete within its environment. This can, for example, be due to lower competitor prices or higher quality of competitor products.

Comparative studies may be conducted at the inception, during operations and at the closure of an organisation. There are many more examples of such comparative studies that organisations could conduct. In terms of usability, the question of whether the usability of information systems can be compared to aid an organisation in remaining competitive in a digital market or for encouraging the adoption of UCD practices may be useful.

2.8.2 COMPETITIVE ADVANTAGE

Competitive advantage is a reason why a CIUE may be performed at an organisation. For organisation A to be successful in gaining competitive advantage over organisation B, it should offer a product or service that is perceived to be better by its customers (Teece, 2010). If organisation A succeeds in drawing the customers of organisation B, it may have adverse effects on sustainability of organisation B. On the other hand, if organisation B is drawing the customers of organisation A, it may make organisation B more sustainable than organisation A, as more people are buying products or services from organisation B (Teece, 2010).

With the advances in social media, it has become easier for an organisation to conduct directed one on one marketing. This has led to more organisations adopting customer relationship management practices and larger databases containing customer data (Peppers & Rogers, 2016). Typically, this data is used to improve the experiences that customers have when they interact with the organisation. Most of the interactions with organisations that customers have today are facilitated through various technologies, typically through a computer interface (Peppers & Rogers, 2016).

Usability relates to the ease of use of computer systems (Nielsen, 2003). As per TAM, this ease of use of system influences the behavioural intention to use a system and ultimately the actual use of that system. This means that if the usability of a system is improved it will improve the ease of use of that system, which may lead to increased actual use of that system (Davis, 1989). The more a customer uses an organisation’s customer interaction points, the more the possible data that can be collected about that customer, which causes the potential to build greater customer relations to grow (Peppers & Rogers, 2016); this may lead to competitive advantage.

"Anti-competitive practices refer to a wide range of business practices in which a firm or group of firms may engage in order to restrict inter-firm competition to maintain or increase their relative market position and profits without necessarily providing goods and services at a lower cost or of higher quality" (Co-operation & Dev., 1993, p. 147). When seeking competitive advantage, it is advised to avoid anti-competitive behaviour.

When anti-competition occurs, business is conducted in a manner that may eliminate competition to gain competitive advantage. To do this an organisation typically takes active steps against competing organisations, such as agreeing to exclusive dealing with suppliers (Whish & Bailey, 2015). Other examples of anti-competitive behaviour includes the fixing of prices between organisations, dividing territories where businesses agree not to compete and dumping. Dumping is where businesses sell products at a loss to gain market share only to sell products at a profit at a later stage.

Anti-competition is generally regarded as unethical because of the negative effects it may have on competing businesses, such as a limitation in the suppliers a business can use. These effects may also have adverse consequences on the socio-economic environment of organisations effected by the anti-competitive practices. As such, conduct by organisations that is classified as anti-competition is deemed illegal in many countries (Whish & Bailey, 2015).

After considering the above, one can see that, seeking competitive advantage may be beneficial to businesses. However, businesses that seek competitive advantage should ensure that they do not diverge to anti-competitive conduct (Porter, 2008). It would be better for an organisation to engage in competitive behaviour by attempting to improve itself.

It should therefore be classified a goal of CIUEs to attempt to improve the organisation conducting the study rather than to discredit another organisation or behave in an anti-competitive manner. If a study such as this is used in an anti-competitive manner, the behaviour may be classified as unethical and most often illegal. However, when a more competitive business environment is created from studies such as CIUEs, it may lead to improved products or services that those organisations deliver.

2.8.3 COMPETITIVE ADVANTAGE GAINED FROM ENHANCED USABILITY

A study done to evaluate the benefits drawn from enhancing usability of e-commerce or operational information systems showed that there was an improvement on sales, website traffic, user productivity and the use of specific target features as well as reduced training cost and limited user risk (Nielsen & Gilutz, 2013). For e-commerce websites, the return on investment of a usability improvement is easy to determine as an evaluation can be done on sales improvements. It is, however, challenging to determine the return on investment of a usability improvement of information systems used in operations, as the productivity of employees is not directly linked to monetary value for organisations (Nielsen & Gilutz, 2013).

Regarding sales, it may happen that a customer using an e-commerce site can become frustrated quickly if the site is not usable, which may cause them to leave. Furthermore, usability enhances the experiences that users have on a site, which may make them stay on the site for longer to browse, recommend the site or return to the site. This extra time that they spend on the site and the site recommendations (to friends or family) they make increases the possibility that more will be bought, improves site traffic, and allows for extra marketing time (Schneidermeier, 2015).

Furthermore, development teams that have used usability engineering techniques early in a development cycle have indicated that this reduces the cost of development. These techniques have shown to be effective in minimising usability errors of information systems, which leads to the reduction of costs associated with more development iterations, user training, system documentation, support and maintenance (Bevan, 2009b). Additionally, more than 80 percent of mobile applications that are developed for both Android and Apple smartphones are downloaded less than a million times and once they are downloaded, one in four of these applications are only used once. Market research suggests that usability is the biggest reason why mobile applications are rejected by customers (Hoehle & Venkatesh, 2015).

Information systems are a critical resource for businesses today (Arnott, Lizama, & Song, 2017). Enterprise Resource Planning (ERP) and modern custom software have brought about a time where the efficiency and effectiveness of most tasks that are conducted in business are enhanced by software (Mamoghli, Goepp, & Botta-Genoulaz, 2017). Business intelligence and decision support systems have become the technological priority for executive management worldwide today (Arnott et al., 2017). In marketing, social media is used to enhance the customer experience and perception of an organisation through personalised touch points (Peppers & Rogers, 2016).

When systems are easier to use, the efficiency and effectiveness of task completion on those systems are less dependent on the computer skills or literacy of system users (Venkatesh et al., 2017). In usability testing and development of information systems, the literacy levels of current and potential users should therefore be a consideration. A further benefit of an easier to use system is that the actual use of that system is more likely (Davis Jr, 1986). The ease of use of systems does not only apply to customer touch points as it may also improve the use of systems by internal staff or other stakeholders (Nielsen, 2003).

If all these business stakeholders can complete tasks more efficiently on information systems, it will save the time that they spend completing those tasks on the information system. Typically, when the usability of an information system is improved, it will also improve the collective efficiency of the people using that information system. This influences their effectiveness, in that they will be more likely to do tasks and are happier completing them, and efficiency, in that they will be able to complete more tasks quicker (Davis, 1989; Nielsen, 2003). A mathematical representation to calculate the amount of time stakeholders take to complete all their tasks is as follows:

$$T=\sum_{i=1}^{i=S}(\sum_{j=1}^{j=N} Z)$$

where T is the total time taken to complete all tasks by all stakeholders at a specific point in time, S is the number of stakeholders, N is the number of tasks and Z is the time taken to complete that task. There are a variety of variables that may have an influence on the above-mentioned variables, but the equation above demonstrates the exponential effect of a change on Z. An improvement on usability may cause an improvement on Z. The exponential effect on the improvement on Z may have a significant effect on the amount of time stakeholders take to complete tasks and thus the efficiency of those stakeholders.

For an organisation that presents products or services for potential customers on an information system, the usability will have an influence on the perceived ease of use of that system and thus also the actual use of that system (Davis Jr, 1986; Hasan et al., 2013).

This will also have a direct influence on that organisation's customers' willingness to purchase products or services on that system, because that would require actual system use (Davis Jr, 1986; Hasan et al., 2013).

A methodology for CIUE should address the possible advantage above, namely competitive advantage gained from the better usability of information systems.

2.8.4 COMPARATIVE STUDIES AND THE SELECTION OF SOFTWARE

Ayala, Hauge, Conradi, Franch, and Li (2011) suggest that the success of software products using third party components is dependent on the ability to select these third-party components correctly. Third party components are software products that are included in information systems as libraries that serve certain functions (Zaimi et al., 2015). In that way, development teams do not need to build those components. Ayala et al. (2011) also showed that there is a lack of knowledge regarding this selection process and that it would be useful for the successful production of software products if there were more formalised methods to do this selection.

Jadhav and Sonar (2009, p. 555) state that there is a need for a "framework comprising of software selection methodology, evaluation technique, evaluation criteria, and a system to assist decision makers in software selection". Jadhav and Sonar (2009) also suggested that analytical hierarchical processes are used to select software packages and that there may be a lack of evaluation criteria that could be used in this selection process. It is acknowledged that such a framework should address the selection of software components, packages and the selection of fully fletched off the shelf information systems.

The typical process would be to use evaluation criteria as follows: evaluate each software package using the criteria, rank each package and select the package with the highest rank

(Jadhav & Sonar, 2009). The criteria is usually set up on the basis of what is important for decision makers, which is typically the cost of the package or its reputation (Carvallo & Franch, 2009; Low & Chen, 2012).

Low and Chen (2012) suggested criteria for the selection of cloud based hospital information systems. This was done by consulting 42 experts in the field and using fuzzy methods (including fuzzy set theory, the fuzzy Delphi method and a fuzzy analytical hierarchy process) to rank their selection criteria. Ease of use was ranked twelfth and the experience they had from a professional perspective as second. The experts had a choice of 23 criteria to select from. It was a recommendation that these criteria be evaluated by decision makers before the selection of a product. Although the priority of criteria may vary for different systems, an evaluation method that could be used in the selection of software is vital to avoid incurring unexpected delays, costs and unsatisfactory scope in the future.

Carvallo and Franch (2009) studied the selection of third-party components for an ERP system at a telecommunications company. This was done through a call-for-tender process for the selection of these third-party components. They found that a set of requirements was compiled for the selection of these components, but no hands-on experimentation was done to validate whether these components satisfy the initial requirements.

They later also found that there was a mismatch between component capabilities and requirements. In a case like this, an in-depth study of the products based on a variety of criteria (including the evaluation of the ease of use of these components and how they would satisfy a need as set in the requirements document) would have gone a long way to avoid the eventual cost of wrong component selection (Carvallo & Franch, 2009; Low & Chen, 2012).

It is essential for cases like the above that a criteria based solution be used in the selection of third party information systems or software components. In such an approach, a selector would come up with a range of scores for the selected criteria and then rank the information systems or components based on the score they received for each criterion (different criteria may be weighed differently).

Currently, no formal methodology is available that allows a selector to go through a systematic process for the selection of an information system or component based on the usability of that software. Although CIUEs compare the usability of interfaces between different organisations, a methodology for CIUEs, and the application thereof, could contribute to more rigorous selection of software or software components based on usability as adaptions of the methodology could be made to compare the usability of products that are not present at different organisations.

2.8.5 DEVELOPMENT OF BENCHMARKS

There are various questions that stakeholders of an organisation may have that could be answered using benchmarking. These questions include aspects like why some competing organisations are doing better than others, what the strengths and weaknesses of an organisation are, and why certain organisations may easily mitigate threats and take advantage of opportunities (Cragg, 2002; Pearlson & Saunders, 2009; Petuskiene & Glinskiene, 2015). Similar questions could be asked about the organisation's competition and this may enable the organisation posing the questions to gain competitive advantage (Cragg, 2002; Pearlson & Saunders, 2009; Petuskiene & Glinskiene, 2015).

A typical method to answer these questions is for an organisation to compare the way it is conducting business operations and delivering products or services to industry competition or internal standards. This is called benchmarking, a process that may assist organisations to improve by granting insight into areas that require attention (Petuskiene & Glinskiene, 2015). For example, if an organisation is making a large profit on a specific product, this organisation may perceive that it is doing well. However, when they compare the profits to a standard set by other organisations, it may seem that they should be doing better.

In the information systems industry, there are various benchmarking tests available. An organisation could, for example, conduct a Measurement for Information Technology Effectiveness (MITE) assessment, which benchmarks that organisation against the best practices of many other organisations within that industry (Chin, Keese, Apple, & Folkert Herlyn, 1998). Benchmarking tests are also done on records and information management activities, staff and systems using surveys (O'Donnell, 2017) .

Several attempts to come up with a model to benchmark user interfaces have been made (Presley & Fellows, 2013; Rohrer, Wendt, Sauro, Boyle, & Cole, 2016), however, no method to benchmark user interfaces in a quantitative manner was found in the literature. These methods often depend on expert recommendation and the qualitative judgement of a range of usability indicators. However, organisations may doubt expert recommendation due to the possibly perceived subjective nature thereof (Law & van Schaik, 2012).

A manner to objectively evaluate usability, like the benchmarking of usability, may be more suited than expert recommendation. A goal of this dissertation is to present a methodology that will enable the users to compare the usability of computer interfaces objectively. These comparisons may be used to develop benchmarks for usability.

2.8.6 CIUES AND USER EXPERIENCE (UX) PRACTICE BUY IN

In a study done by Pretorius et al. (2015) in South Africa, it was found that the top challenge faced by the UX practitioners surveyed was that there was a lack of buy in and the deficiency of sufficient promotion of UX practices and UCD within their organisations. This is despite the fact that the usability of an information system and the buy into the UCD philosophy has been shown to contribute significantly to potential positive experiences a user of that information system may have (Venturi et al., 2006).

The comparison of information system usability to that of other organisations may improve the perceived value of the UCD philosophy within an organisation because of the potential competitive advantage that may be gained there from (Venturi et al., 2006). The study of potential competitive advantage gained from enhanced usability may also lead to the possibility of scaling the value of an investment made into usability enhancements.

2.9 METHODS OF EVALUATING USABILITY AND THEIR USE IN CIUES

Various methods of usability evaluation are discussed in this section, namely focus groups, heuristic evaluation, remote usability evaluation and controlled environment usability evaluation. This subset of the numerous available methods has been chosen for its applicability to the research reported in this dissertation. This section also discusses whether the respective methods of usability testing are appropriate for use in CIUEs.

2.9.1 FOCUS GROUPS

The focus group, or a special interest group, is a particularly selected group of people in terms of purpose, size, composition and procedures that may be used for a usability evaluation (Krueger & Casey, 2014). Focus groups are used to gather people's opinions. The purpose of conducting such a focus group is to better understand how people feel or think about the subject involved. This subject may be a computer interface (Krueger & Casey, 2014).

In terms of usability evaluation, focus groups, typically a group of six to nine users, are gathered informally and they discuss their feelings, opinions and ideas about a computer interface (Jurca, Hellmann, & Maurer, 2014). There is always a moderator present to keep the focus of the discussion on the computer interface in question.

The disadvantages of focus groups include various participatory challenges. This includes conflict, which may influence the quality of data collected. It may also cause lack of interest, which limits the amount of data collected. The information gathered from users is subjective in nature. The social nature of focus groups may limit participants from raising their view and semantics can cause ambiguity and lack of understanding (Acocella, 2012).

There are also several advantages of using focus groups for usability evaluations, including the low cost associated with such a study because there is no specialised equipment necessary. They may also include some analysis of data, as a discussion may form part of a focus group. The participants of a focus group may highlight issues that would otherwise not have been found by a usability expert (Acocella, 2012).

2.9.2 HEURISTIC EVALUATION

In heuristic evaluation a group of experts evaluate the interface based on a set of recognised usability principles, which are often referred to as heuristics (Jurca et al., 2014). There are also various advantages and disadvantages associated with the use of heuristic evaluations. It is a quick method to find the major problems with a user interface and there are limited costs associated with this type of usability evaluation (Otaiza, Rusu, & Roncagliolo, 2010).

Furthermore, when heuristics are used, there is most often a set of resolutions that already exists to resolve identified problems. This allows for the quicker resolution of problems (Nielsen, 1995). When usability experts frequently apply heuristics, they become good at identifying usability problems at first glance, which assists in the decision making regarding investment into usability optimisations (Preece, Rogers, & Sharp, 2015).

A major disadvantage is the subjective nature of the opinion of evaluators (Preece et al., 2015). For example, an expert may have a problem with a certain colour and as such may rate one interface better than another because of that certain colour being present on one of the interfaces. However, this influence may be minimised by getting different experts to evaluate the same interface and using results that are common between experts (Lilholt, Jensen, & Hejlesen, 2015).

However, it has been shown that experts typically only identify 20 to 51 percent of usability problems when applying heuristic evaluation (Nielsen & Molich, 1990). It has also been found that different usability evaluators find different problems. What is easy for one evaluator to identify may be more challenging for another. The use of a number of experts to evaluate the usability of an information system also assists in minimising this problem and identifying a broader spectrum of usability problems.

2.9.3 REMOTE USABILITY EVALUATION

Remote usability testing is regarded as the evaluation of the usability of systems without the examiner being in the same geographical location as the participant being evaluated (Alghamdi, Al-Badi, Alroobaea, & Mayhew, 2013). Various methods to conduct remote usability evaluation currently exist (Hertzum, 2016). These are classified into two categories, synchronous and asynchronous. The difference is that there is real time involvement with a facilitator in synchronous remote usability evaluation and not with asynchronous remote usability evaluation. The main advantage of this type of evaluation is that many users can be involved in the evaluation without any additional cost, especially in the asynchronous case (Bastien, 2010).

There is an extensive range of data that could be collected from remote usability evaluation including click patterns, number of system errors, number of user errors, time spent and interface navigation paths. This data may then be used to improve the user interface in areas where problems are discovered (Bastien, 2010).

Automation of usability evaluations may be applied using computer software analyses to evaluate what the user did and how the user interacted with the computer system. This allows for a lot of data to be collected with minimum effort and cost (Hertzum, 2016).

The main disadvantage of remote usability evaluation for quantitative analysis is the variable environment of the user. The user is usually at home or at work and various distractions may occur that will influence the data that is retrieved about the interface use. Furthermore, different users may use different software tools, such as browsers, to complete tasks. This variability in software will have an influence on the outcomes of usability evaluations (Hertzum, 2016).

There are other environmental factors that will influence the nature of the data such as the speed of processors, the size of hard drives, the speed of internet connections, memory capacity and the quality of the graphics card, which could all influence how the user interacts with the interface.

2.9.4 CONTROLLED ENVIRONMENT USABILITY EVALUATION

Another class of usability evaluations are called controlled environment usability tests or evaluations. Here users are typically invited to a laboratory where they are requested to use an information system to complete a set of predefined tasks (Lazar, Feng, & Hochheiser, 2017). These users may be recorded, the interface may be recorded, the user may be requested to do a think aloud, the user may fill in a questionnaire or an interview may be conducted after the evaluation (Bastien, 2010).

The drawbacks of this approach are that it may be costly because of the resources that are required, such as a laboratory, and the number of users that can be tested may be limited by the resources available. It may also be difficult to test a diverse group of users because usability laboratories are typically in one geographical location. The restrictions or limitations that are present in the laboratory will be present for all of users involved (Barnum, 2010).

Controlled environments also have several advantages. For example, the usability laboratory can be set up in a manner that is optimal for a specific test. The tools that the user uses to complete the task may be limited to avoid diversity between users. The browsers, internet speed, computer hardware and software can be kept constant to limit the capriciousness of the data. Different technologies such as eye trackers or audio visual equipment can be used to enhance the amount of data that could be retrieved from the user’s interaction with the computer system (Barnum, 2010).

2.9.5 APPROPRIATE METHODS FOR USABILITY EVALUATION IN CIUES

The main advantages and disadvantages of the various methods to evaluate usability are summarised in Table 2.9.1. The methodology that will be presented as the resulting artefact from the research reported in this dissertation will be used to conduct CIUEs. A goal of the methodology is to use objective data, therefore focus groups and heuristic evaluations are not suitable for use in CIUEs. Furthermore, the variable nature of the environment of remote usability evaluation renders remote user evaluation inappropriate for use during CIUEs because it is impossible to account for all the variables in such possibly unknown and vastly different environments in a quantitative study.

Table 2.9.1: Advantages and disadvantages of methods to evaluate usability

Method of evaluation	Advantages	Disadvantages
Focus group	Low cost. Possible quick analysis.	Subjective opinion of focus group. Participatory challenges.
Heuristic evaluation	Low cost. Quick.	Subjective opinion of usability experts.
Remote usability evaluation	Low cost. Large amount of data. Possible automation.	Variable environment
Controlled environment evaluation	Objective quantitative data. Optimal configuration. Large amount of observation tools.	High cost. Limited diversity of users.

Controlled environment evaluations are more suited for CIUEs because the variability in data and environment can be controlled. As such, the focus can be placed on the difference between the usability of the different competing computer systems. This can also be optimised by using specialised equipment, such as an eye tracker or a video camera, to increase the amount of data and the number of indicators that could be used for the CIUE. This, in turn, increases the value and quality of the comparison. This extensive range of indicators minimises the need for possibly subjective data such as expert recommendations or user opinions.

2.10 EXISTING METHODOLOGIES FOR INTER-ORGANISATIONAL COMPARATIVE USABILITY STUDIES

Gregor and Hevner (2013) suggest that when DSR is done, an assessment of what is already known and what could be drawn from existing knowledge is useful for the grounding of the research. This section therefore evaluates artefacts similar to a methodology to conduct CIUEs that were found in the literature.

A comparison of methods to study usability was done by Molich, Ede, Kaasgaard, and Karyukin (2004). These researchers studied the methods that nine organisations used to evaluate the usability of Hotmail, an emailing tool developed by Microsoft. Molich et al. (2004) revealed that most of these organisations used different methods and a lot of usability problems were missed by the studies that the individual organisations conducted. This is, however, not the same as a CIUE because it compares the methods organisations used and not the usability of competing products across the organisations.

Another comparative study to evaluate methods used to assess the usability of mobile applications (Beck, Christiansen, Kjeldskov, Kolbe, & Stage, 2003) found that the techniques were not comparable, but the effectiveness of the techniques depended on a range of statistical criterion (Beck et al., 2003). Again, this is not the same as a CIUE because it compares usability evaluation techniques.

AlRoobaea and Mayhew (2014) conducted a comparative study of e-government websites in the United States of America (USA). The authors claim that usability studies of e-government websites in the USA are usually conducted using content based evaluations. They also claim that these e-government websites either use dichotomous measures or generic scales to construct indexes.

These indexes are used to construct benchmarks and comparative reviews are then conducted based on these benchmarks. This study goes on to suggest a content-based analysis methodology based on a Guttman-type scale that has several benefits over the current methods used by the e-government websites. A Guttman scale presents many items and respondents have to indicate which items they agree with or not. This is often accomplished through a series of yes/no questions (AlRoobaea & Mayhew, 2014).

A critique on the above is that usability does not necessarily relate to the content of a website (according to Nielsen's (2003) definition of usability). Furthermore, a CIUE as suggested in this dissertation is a method to compare the usability of computer interfaces of information systems at competing organisations, not one that assesses the usability based on content analysis of websites (AlRoobaea & Mayhew, 2014).

From the above it follows that a range of comparative evaluations of aspects relating to usability have been conducted, but they do not correspond to the CIUE as discussed in this dissertation. From the research done thus far, examples of such studies have not been found. A comprehensive search to find methodologies to conduct CIUEs was unsuccessful.

2.11 TENTATIVE DESIGN OF METHODOLOGY TO CONDUCT CIUES

In line with DSR, a suggestion for the methodology to conduct CIUEs is required as a starting point that will lead to the final design thereof (Kuechler & Vaishnavi, 2008). Avison and Fitzgerald (2003) describe a methodology as a series of steps and procedures followed during developing an information system. When this definition is applied to CIUEs it becomes a series of steps and procedures followed during conducting a CIUE.

When defining a methodology to conduct research, it is useful to investigate the research philosophy, the research approach, the research strategy, the time horizons and the data collection methods (Saunders & Lewis, 2009). The investigation of these elements has been shown to be useful in research because it may help the researcher to identify which research methodologies could work and which will not (Easterby-Smith et al., 2012). Since usability testing can be regarded as a form of research these concepts can also be applied to a CIUE methodology.

2.11.1 THE PHILOSOPHY OF CONDUCTING A CIUE

The overall goal of a CIUE is to improve the usability of an information system by evaluating and comparing its usability to that of information systems of competing organisations. This evaluation is done through usability studies involving actual users. The philosophy of UCD is about promoting a design process in which the end users influence how an information system is designed (Abras et al., 2004) .

To follow UCD, the user of a design should be placed at the centre of the design process (Norman, 1988). This is the philosophy followed in the conducting of a CIUE because users – their behaviour and opinions – are at the centre of the usability evaluation of information systems when a CIUE is done.

2.11.2 THE APPROACH FOLLOWED WHEN CONDUCTING A CIUE

A goal of the methodology to conduct CIUEs should be that it is done in an objective manner as it may be less believable if it were done through potentially subjective methods. Expert recommendations are for example doubted due to the perceived subjective nature thereof (Law & van Schaik, 2012). Therefore, it is recommended that a quantitative approach is followed in the collection and analysis of the usability data collected in a CIUE. This quantitative approach will have the epistemological assumption that an objective reality exists in which the usability of information systems can be measured and analysed deductively (Bryman & Bell, 2015). This objectivity may also contribute to the development of methods to benchmark usability.

2.11.3 THE STRATEGY AND TIME HORIZONS FOLLOWED WHEN CONDUCTING A CIUE

The strategy followed by a CIUE to compare the usability of information systems is the execution of a multiple-case study where various organisations are selected and the usability of their information systems are compared (Yin, 2013). Furthermore, a CIUE is done once at a particular point in time. A CIUE represents a single instance of the comparison of information system usability between organisations. It is therefore regarded as a cross-sectional evaluation. If a longitudinal CIUE is to be done, the CIUE could be performed multiple times at different time intervals, which will make it multiple CIUEs. Longitudinal CIUEs fall outside the scope of this study.

2.11.4 STEPS AND PROCEDURES WHEN CONDUCTING A CIUE

It was explained in Section 2.8.5 that a controlled environment for usability evaluations is best suited for conducting CIUEs because the environmental influence on the usability evaluations can be regulated. This places the focus on the variability of information systems involved in the CIUE. The steps and procedures associated with controlled environment usability studies are therefore applicable to CIUEs.

Furthermore, it was found that the data analysis and collection associated with a CIUE should be done in a quantitative manner. The steps and procedures to compare the usability between the information systems selected should thus be done deductively. The detailed steps and procedures will be investigated in Chapter 4. Figure 2.11.1 outlines the tentative design of the methodology to conduct a CIUE.

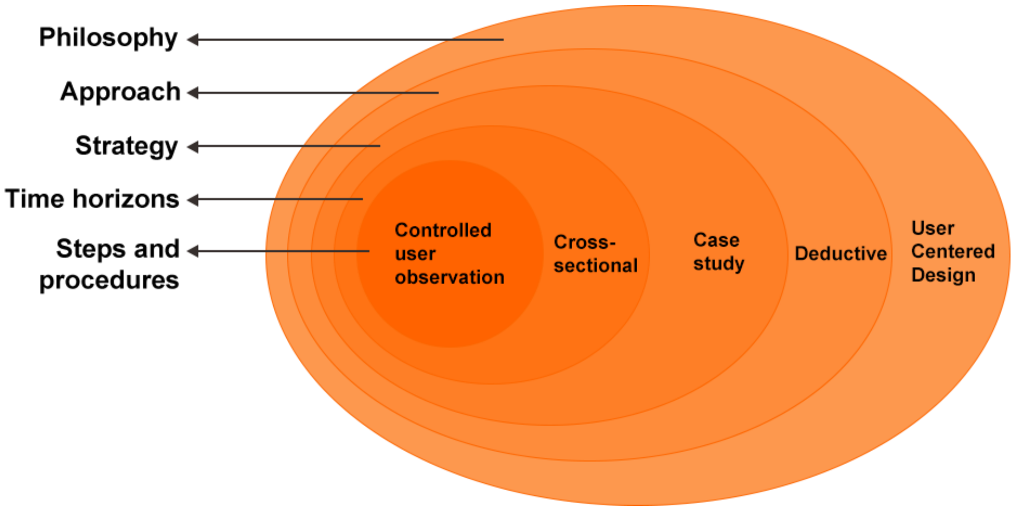


Figure 2.11.1: Tentative design of methodology to conduct a CIUE.

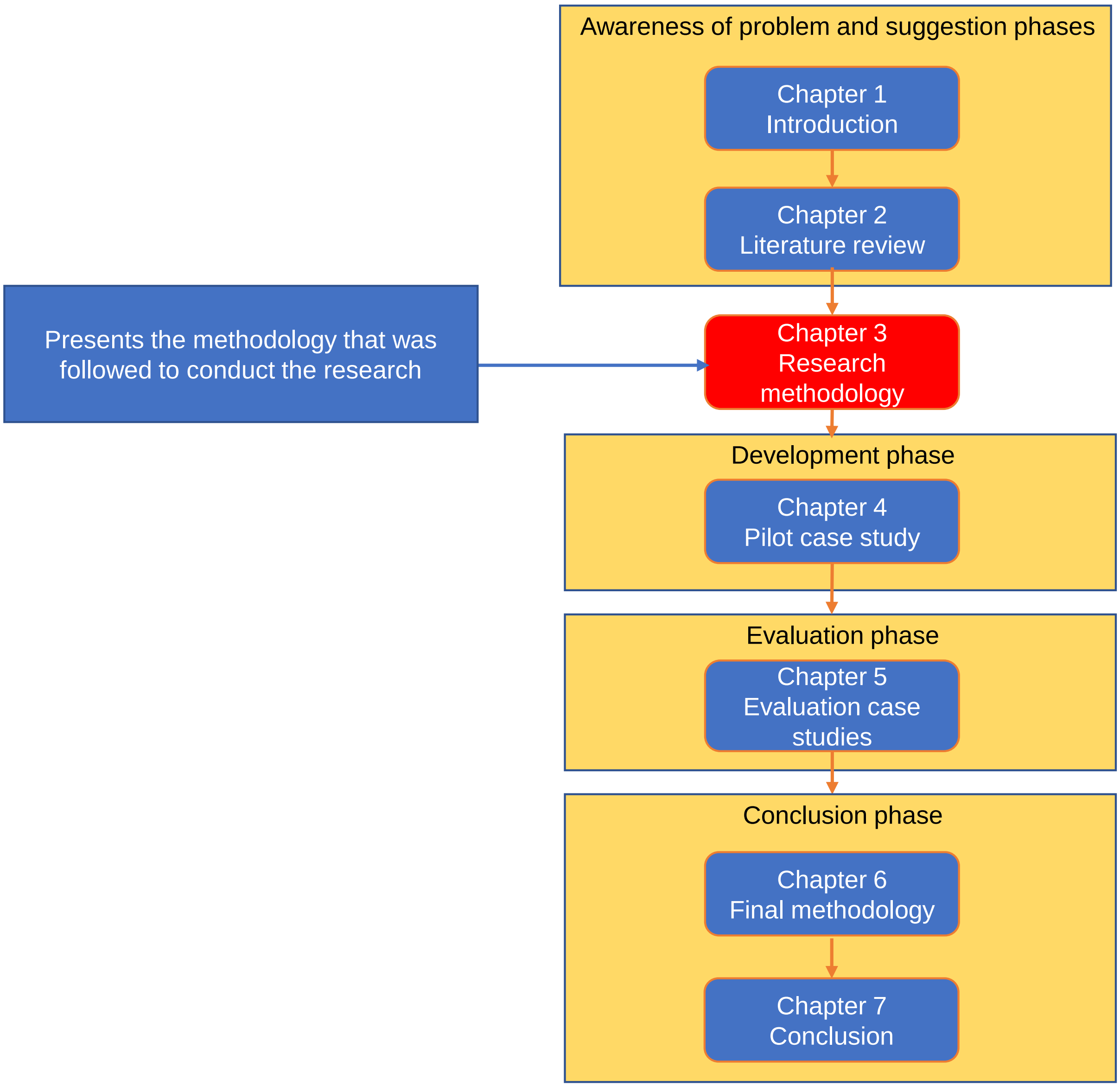
2.12 CONCLUSION

This chapter investigated the literature related to the research reported in this dissertation. Initially, TAM was brought forward as the theoretical backing used for this study (Davis, 1989) and UCD as the philosophy behind the methodology to conduct CIUEs. The reasoning behind what a methodology to conduct CIUEs is and why it may be useful for organisations was also described in this chapter.

Furthermore, this research was contextualised by discussing the South African IT and usability environment. A comparison was also made between different methods of usability evaluation as they will form part of a CIUE. Gregor and Hevner (2013) suggest that the evaluation of similar artefacts may be useful in DSR; this was done in Section 2.10. Finally, a tentative design of the methodology to conduct CIUEs was outlined in Section 2.11.

CHAPTER 3: RESEARCH METHODOLOGY AND APPROACH

The position of chapter 3 in the dissertation



3.1 INTRODUCTION

This chapter describes the methodology and approach that was followed in this research study. The purpose of this section is to indicate the validity, replicability and reliability of this study. This chapter thus also contributes to the scientific method used in the study in that it satisfies the replicability characteristic of scientific method (Bhattacherjee, 2012).

The focus of this chapter is on the methodological approach, the research design and the research philosophy employed in this research study. There is also a discussion regarding the data collection and analysis strategies and tools that were used as part of this research study.

3.2 RESEARCH DESIGN AND PHILOSOPHY

Bryman (2015) states that a research philosophy refers to the set of beliefs concerning the nature of the reality being investigated. Saunders and Lewis (2009) developed a “research onion” to describe the stages that researchers should go through to formulate an effective methodology. The most outer layer of the onion is about the research philosophy that a research study follows. This research onion is depicted in Figure 3.2.1.

According to the Saunders and Lewis (2009) research onion, the philosophy adopted by a research study could be positivism, pragmatism or interpretivism. Bryman and Bell (2015, p. 729) define positivism as “an epistemological position that advocates the application of methods of the natural sciences to the study of social reality”. The epistemological assumption made by positivism is that there is an objective reality that can be studied. Typically, a study that adopts a positivist research philosophy will follow a deductive approach using experiments to test hypotheses.

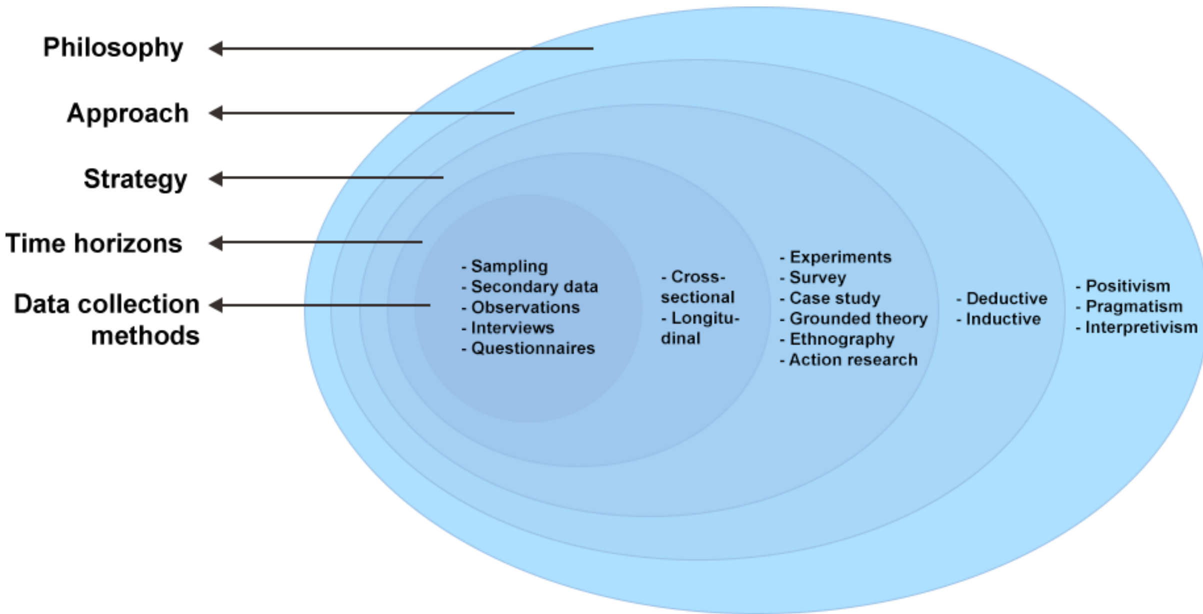


Figure 3.2.1: Saunders and Lewis (2009) research onion.

In contrast, the interpretivist epistemological position finds that the meaning of social phenomena is created by each observer (Bryman & Bell, 2015). The difference between positivism and interpretivism originates not only from the stance taken on reality but also from the goals of research done from each perspective. The goal of interpretivism is to study the meaning of social action whereas positivism attempts to study social reality. Typically, interpretivist research follows an inductive approach, often using qualitative methods (Bryman & Bell, 2015).

Realism is another research philosophy which is the belief that objects exist ontologically independent of human thoughts or beliefs (Mingers, 2004). Like positivism, realism also takes the ontological stance that there is an objective reality. Quantitative methods can also be used to measure this objective reality using a deductive approach (Mingers, 2004).

Another paradigm that is appropriate for various forms of research including social research is pragmatism (Morgan, 2014). The primary concern of pragmatism is to find a method that works to achieve research goals. The pragmatist may thus use inductive or deductive approaches to research in an attempt to achieve the research goals. Pragmatism acknowledges the use of quantitative, qualitative and mixed research methods in the effort to find a way that works to achieve the research goal.

Various strategies could be used to conduct research (Saunders & Lewis, 2009). These are often classified as either quantitative or qualitative methods. Saunders and Lewis (2009) define a research strategy as “how the researcher intends to carry out the work”. Some examples are case studies, experiments, surveys, grounded theory, ethnography or action research.

Each of these research strategies is applicable in several scenarios. The selection of an appropriate research strategy depends on the research methodology and philosophy adopted by the researcher. These strategies may accept several types of data collection methods like interviews, questionnaires, observation, secondary data or sampling.

The Saunders and Lewis (2009) research onion further specifies that time horizons should be a consideration in the formulation of an effective research methodology. The time horizons of a research study may either be longitudinal or cross-sectional. In cross-sectional research studies, data is collected at a certain point in time whereas in longitudinal studies the influence of time on a study is measured and so data is collected at various points in time.

The research discussed in this dissertation adopted a pragmatic research philosophy and used the Design Science Research (DSR) methodology. DSR is described in Section 3.3 and how it was applied in this research study is described in Section 3.4. Regarding time horizons, this research study was a cross-sectional study, data was only collected once, not at several points in time. A multiple-case study and experiments were used as research strategies in this study. User observation and interviews were used as data collection methods. The positioning of this research is summarised in Figure 3.2.2.

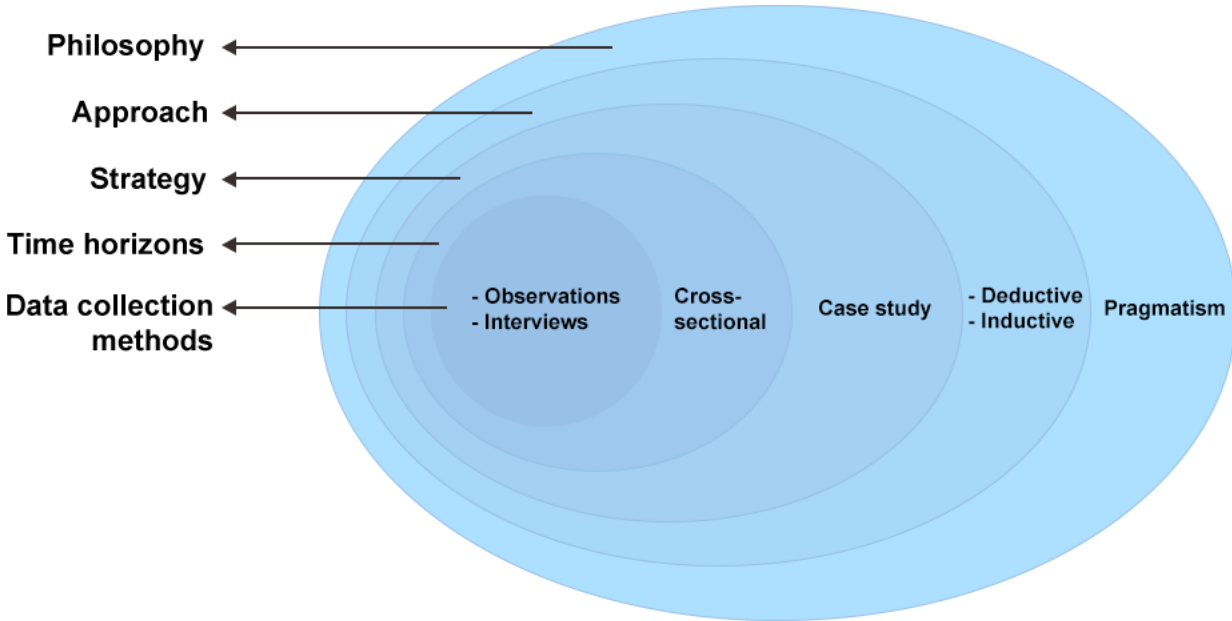


Figure 3.2.2: Positioning of this study

3.3 DESIGN SCIENCE RESEARCH

The research methodology that was adopted in this dissertation was DSR. According to Hovorka (2009), DSR is a pragmatic research methodology that is used to develop artefacts that are innovative and solve real world problems. The artefact that was developed in this study was a methodology to conduct Comparative Inter-organisational Usability Evaluations (CIUEs). The methodology that was produced as an artefact should not be confused with the methodology that was used to conduct the research.

Hevner and Chatterjee (2010) find that DSR is highly relevant to information systems research because it relates directly to two related issues, namely the role of an information system artefact in research and the perceived lack of professional relevance of information system research.

Various other sources conclude that artefacts relating to methods, such as a methodology, are viable artefacts and could be produced through DSR. For example, Offermann, Blom, Schönherr, and Bub (2010) conducted a literature review that included a large range of artefacts that may be produced by DSR, which included methods and methodologies. Winter (2008) and Peffers, Tuunanen, Rothenberger, and Chatterjee (2007) also recognised that methods may be a DSR output. Peffers et al. (2007) define an artefact associated with DSR as "any designed object with an embedded solution to an understood research problem". This definition was also adopted by this research study.

The approach that was followed in the conducting of the DSR reported in this dissertation was that of Kuechler and Vaishnavi (2008) as depicted in Figure 3.3.1.

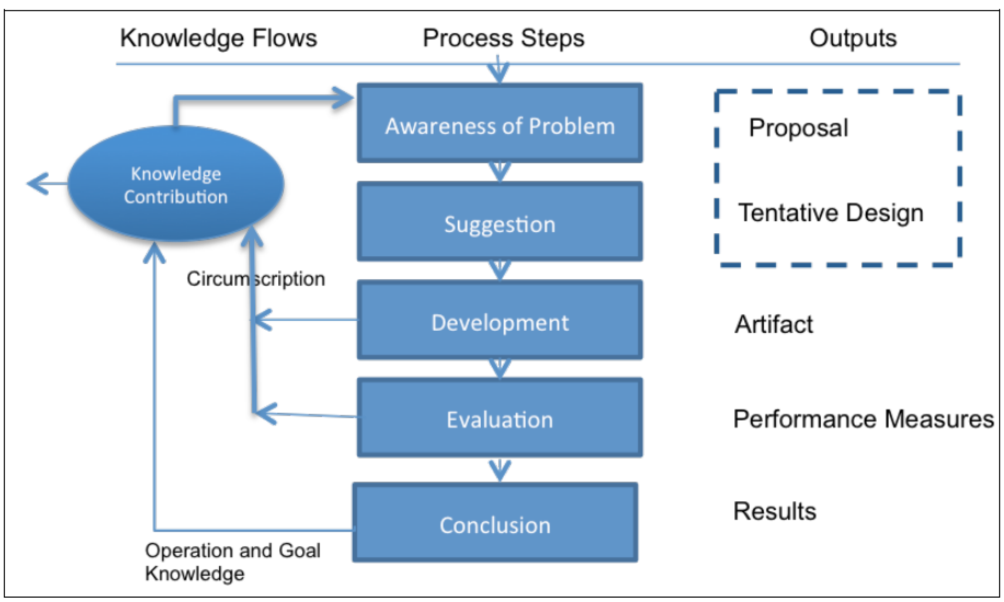


Figure 3.3.1: Phases of DSR

The phases of DSR as defined by Kuechler and Vaishnavi (2008) are as follows:

1. Awareness of the problem – The output of this phase is a proposal for research that is to be done. The awareness may originate from various sources including industry and academia.

2. Suggest a design – The researcher undergoes a creative process where a suggestion for the research artefact is made. The output from this phase is a tentative design.

3. Develop an artefact – The artefact is created by the researcher from the tentative design. Various research methods may be applied to develop the artefact. The output from this phase is a version of the artefact.

4. Evaluate the artefact – Further research methods are applied to evaluate the performance of the artefact. The outcomes of this evaluation feed into a next iteration of DSR, which is used to refine the artefact, and this continues until the artefact's performance is deemed good enough.

5. Conclude and present results – The final phase of DSR is the conclusion that presents the artefact after it was deemed to be good enough. The presented artefact might not be perfect, therefore future research considerations are suggested and the artefact is added to the pool of knowledge.

DSR can be applied in different ways and may be iterative to improve the quality and applicability of the artefact produced (Kuechler & Vaishnavi, 2008). The application of DSR for this research study is discussed in Section 3.4.

3.4 APPLICATION OF DESIGN SCIENCE RESEARCH IN THIS RESEARCH

This section relates to the how DSR was applied in this study to develop a methodology that could be used to conduct CIUEs. The research design is depicted in Figure 3.4.1.

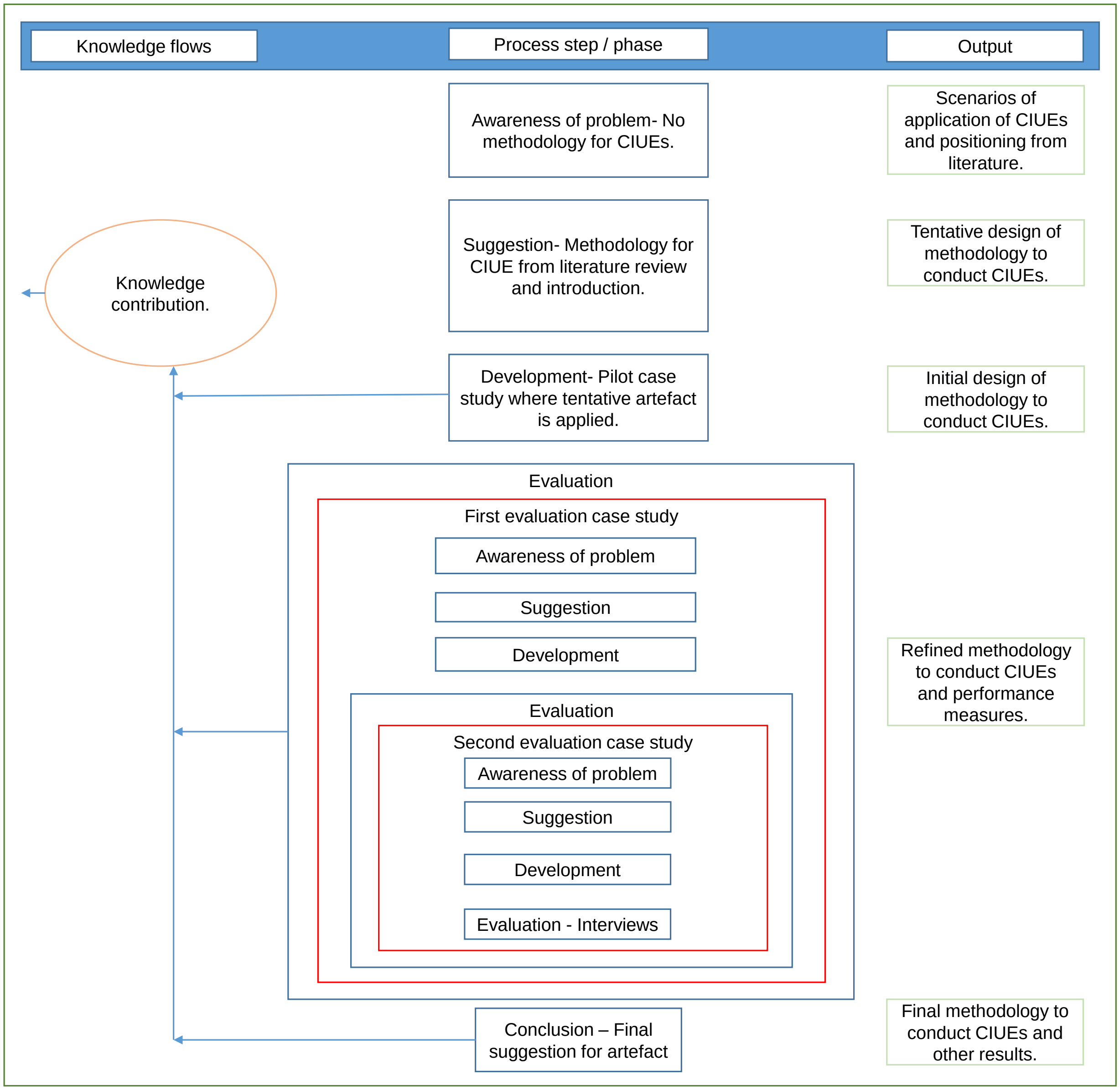


Figure 3.4.1: Application of DSR in this dissertation.

The first phase of this dissertation was the initial awareness of the problem. The main problem identified was that no formalised methodology to conduct CIUEs existed. There were also several sub-problems identified as part of this initial awareness of the problem phase. The problem and its sub-problems, as described in Chapter 1, were positioned by identifying several uses for a CIUE, namely:

1. To compare the usability of user interfaces in information systems when selecting information system products.
2. To work towards the benchmarking of the usability of user interfaces.
3. Creating a method to convince stakeholders of information systems to invest in the enhancement of the usability of an information system.
4. To scale the value of an investment made in usability enhancements.
5. To encourage the use of usability enhancement mechanisms to contribute to the attainment of competitive advantage from user interfaces of information systems.

The suggestion derived from the above discussion was to develop a formalised methodology to conduct CIUEs. The identified uses of the suggested methodology guided the development of its requirements. This suggestion was elaborated through a literature study and a tentative design of a methodology to conduct CIUEs was produced. As part of the tentative design, a description of the possible outcomes of such a CIUE was also formed.

The methodology to conduct CIUEs was further developed using a pilot case study. This produced the initial design of the methodology to conduct CIUEs. This initial design was regarded as the first knowledge contribution made from this research study and was presented at the CONF-IRM conference in Cape Town in 2016 (Kruger, Gelderblom, & Beukes, 2016), the research paper that was presented is available in appendix G. The methodology to conduct CIUEs was developed with the use of the following DSR steps:

1. Evaluating the practical feasibility of the tentative design as produced in the suggestion.

2. Suggesting changes to the methodology to conduct CIUEs to make it more useful and more practically feasible.

The methodology to conduct CIUEs was then evaluated by applying it in further cases and interviews. These cases were used to:

1. Test the methodology to conduct CIUEs in various scenarios.
2. Make suggestions as to how the methodology should be changed to accommodate for more scenarios and so making it more applicable.
3. Study performance measures of the methodology to conduct CIUEs though applying it in diverse scenarios and evaluating the usefulness of each CIUE produced and how this may be improved.

The results of the CIUEs were then presented to selected relevant parties at the organisations where the CIUEs were conducted on to study whether they found the CIUE useful and to gather any suggestions about how the methodology to conduct CIUEs may be further improved. These interviewees were selected on the basis that they play a role in system design processes at their organisations. The way data was collected and analysed are discussed in Section 3.4.2 and Section 3.4.3 respectively. Figure 3.4.2 provides an overview of the research process.

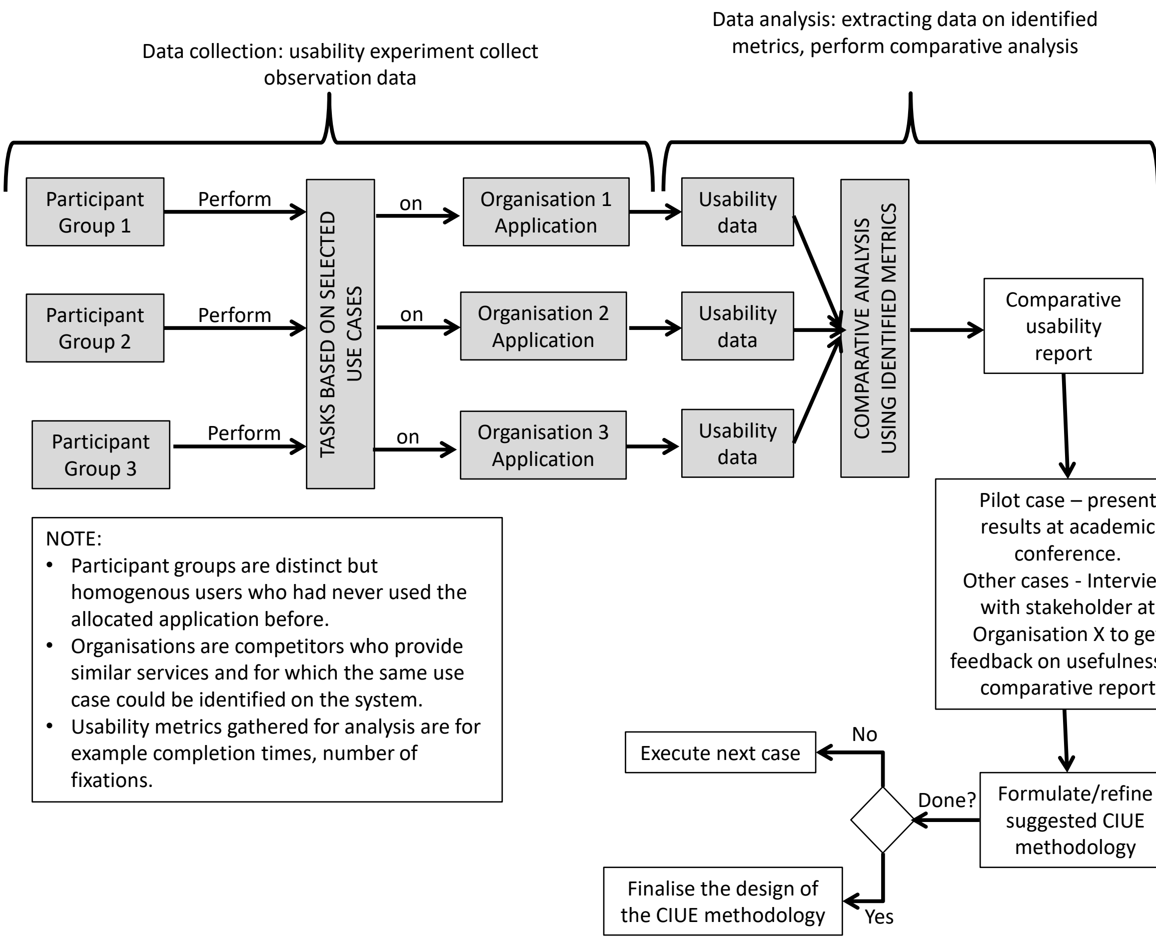


Figure 3.4.2: Overview of the research process

The steps in the research process are divided between data collection and data analysis processes. For data collection, participants were divided into groups. A group was created for each organisation that was evaluated. The form of this division of participants differs for each case that was done as a result of lessons learnt during each case. This is discussed in Chapter 4 and 5. The participants were requested to complete a task on the information system selected for the organisation they were grouped into. They were then observed using a camera, audio recorder and an eye tracker.

When the data collection was complete, usability data was extracted from the collection tools and analysed. This was done by grouping the data according to identified usability metrics and comparing the values for each of the information systems used in the study.

The results from the comparison were compiled into a report. In the pilot case, this report formed part of the research paper that was presented at an academic conference and for the other cases, interviews were conducted at organisations that participated in the study. From data collected at the interviews, the suggested methodology was refined. The feedback gathered at the conference was used to validate that research into CIUEs was a useful endeavour. The refinement of the methodology to conduct CIUEs was done in a DSR cycle until the methodology to conduct CIUEs was finalised.

3.4.1 RELATIONSHIP BETWEEN THE RESEARCH METHODOLOGY AND THE METHODOLOGY TO CONDUCT CIUES

It should be noted that elements of the data collection and analysis for this research study also serve as part of data collection and analysis phases in the methodology to conduct CIUEs.

For each case conducted by this study a CIUE was done. Regarding the data collection, participants were used in the conducting of the CIUE. These participants also served as participants in the research methodology adopted by this study, as they form part of the data collection that was done. The participants were also observed in each case to investigate their role in the methodology to conduct CIUEs.

Furthermore, a CIUE entails data analysis in the form of comparing the results of the different organisations. This CIUE data analysis in the three cases was done as part of this study and thus also forms part of the data analysis component of this study. A meta-analysis of the collective results of the three case CIUEs led to the refinement of the proposed CIUE methodology.

To summarise: The research methodology was used to suggest, produce and test a methodology to conduct CIUEs. The data collection and analysis methods for each case conducted in this study, may also form part of the methodology artefact produced in this study.

3.4.2 DATA COLLECTION STRATEGY

A multiple-case study consisting of three cases was used in the development and the evaluation of the methodology to conduct CIUEs produced in this research study. Case studies have been shown to be useful and applicable to quantitative and qualitative research (Yin, 2013). This study uses the DSR methodology, data collection methods that are applicable in quantitative and qualitative research are suitable for use in studies that apply the DSR methodology. Multiple CIUEs were conducted, one for each case. One was used in the development of the methodology to conduct CIUEs and two were used in the evaluation and refinement thereof. User observation was used to collect data. This was done in a quantitative manner by selecting indicators that could be measured.

For each CIUE case, two or three competing organisations were selected. For each set of competing organisations, a task that a user would typically do on a system at the organisation was selected for comparison. A variety of metrics could be used to compare use cases or software functions (Fenton & Bieman, 2014).

The way competitors were selected for the CIUE was that intended outcomes for the user of the systems were similar. Participants were requested to complete the selected tasks while being observed. These participants were selected through the convenience and snowball sampling methods. The number and the similarity (homogenous vs heterogeneous) of the participants selected varied per case and this variation formed part of the refinement of the methodology to conduct CIUEs. A total of 95 participants were observed as part of the cases in this research study. Details regarding the participants in each case will be provided in Chapter 4, 5 and 6.

Multiple metrics were selected that were measured, summed and compared between the competing organisations' user interfaces. The exact number of metrics used differed between cases. The metrics included time to complete the allocated task, the time taken until first click, the number of fixations until first click and the number of fixations until the task was completed. A fixation is defined as a point that a user looks on the screen (Shneiderman, 2010). The selected indicators are listed in Appendix A. These indicators were selected because they can give an indication of the usability of the user interfaces and are comparable through the measurement of values (Shneiderman, 2010).

The number of fixations gives an indication of the unavailability of data regarding the elements that users are looking for (Djamasbi, Siegel, & Tullis, 2010). Fixations have been linked to intense cognitive processes and as such give an indication that a user is putting in a lot of effort to complete a task if the number of fixations are high (Pan et al., 2004). The reasons for a high number of fixations vary, however, it is ideal for a user to complete the task using a low number of fixations as this shows that it was done using the minimum amount of effort.

The data collected though the observation of the participants was analysed by comparing them using descriptive statistical methods. Participant data was collected through post evaluation interviews (Patton, 2005). In these interviews, demographic questions and questions regarding the participant's technological ability were asked. This data was used to indicate the diversity of participants.

When the data analysis of each case was completed, further data was collected through semi-structured individual interviews (Patton, 2005). These interviews were aimed at assessing the perceived usefulness of CIUEs and refining the methodology to conduct CIUEs. The interviews were conducted with the head of business analysis of one of the organisations studied in the first evaluation case, and the head of IT and innovation at one of the organisations studied for the second evaluation case. The CIUEs were presented to the selected parties and the following questions were asked in the interviews to guide the discussion:

1. What is your opinion of the presented CIUE?
2. What did you learn about your organisation from the CIUE?
3. What changes would you recommend to the CIUE that was done?
4. Depending on the results of the CIUE, comment on your willingness to invest in a usability evaluation.

The following tools were used for data collection:

1. A Tobii Pro X3-120 eye tracker and the Tobii Studio Professional software were used to collect data associated with the selected usability indicators (Tobii, 2016b). Eye trackers have been shown to be useful in the collection of usability indicators (Djamasbi, 2014). The sampling rate of the eye tracker is specified to be 120 Hz and has a 97% track ability (Tobii, 2016b) and has been used successfully to conduct various types of studies. An eye tracker may be used to measure the user eye fixations, the length of the fixations and the saccades between them. These fixations are located on x and y coordinates of screen recordings and lines are drawn between these coordinates that represent saccades. The Tobii Studio Professional software also records the coordinates

of clicks, the time a user spent between clicks and the total time that a user spends on an interface. The recorded fixations can be observed using gaze plots and heat maps.

2. The same Lenovo desktop computer with a 4th generation i7 processor, 16GB of RAM and a touch screen was used for all the usability evaluations. This consistency is an attempt to maximise the commonality of the observation environment and present data that is adequately comparable (Yin, 2013).

3. Two mobile phones were used to do voice recording during the interviews.

3.4.3 DATA ANALYSIS APPROACH

Data analysis was conducted separately for each case conducted as part of this study. This included data analysis relating to the evaluation of the usability of various user interfaces as well as interview data. The first manner in which data was analysed was to directly compare the usability indicators collected (Fenton & Bieman, 2014).

An indicator is defined as "(a) a specification of how we will recognise or measure a concept and (b) a variable that can be used to study another variable because it affects or is correlated with it" (Vogt & Johnson, 2011, p. 179). The comparison of usability indicators was conducted using the following steps:

1. Identifying the indicator to compare.
2. Collecting data relating to this indicator from the observation of the use of the different computer interfaces by participants.
3. Comparing that data and determining the achievement of the computer interface based on the comparison.

For example, the average time taken to complete the tasks on competitor one's application was a compared to the average time taken to complete the tasks on competitor two's application. Where efficiency is desirable in that task, the computer interface with the faster completion time would have performed better. After the comparison process was completed for each case study (meaning each organisation and each indicator), reports were generated that presented the results.

Referring to the process outlined in Figure 3.4.1, when the data collection and the data analysis associated with the pilot case was completed a methodology to conduct CIUEs was formulated. The subsequent two cases and the interviews were used to evaluate the CIUE methodology. The overall results of the two evaluation cases were used to refine the CIUE methodology. The interviews were semi-structured and were analysed by summarising and interpreting answers that could be used to evaluate the usefulness of the methodology or that could contribute to the refinement thereof. When this process was completed, a final version of the methodology to conduct CIUEs was produced and presented in CHAPTER 6: of this dissertation.

3.4.3.1 Tools used for data analysis

The Tobii Pro Studio software has been shown to be a useful tool for data analysis of metrics collected from the Tobii Pro X3-120 eye tracker (Tobii, 2016a), and was used in this dissertation for the initial data analysis. Microsoft Excel 2013 was used to record and compare all the data extracted from the Tobii Pro Studio software. This comparison was presented in a bar chart for each indicator where it was important to highlight differences. The comparison included statistical functions such as summing, calculation of averages and median determination.

Microsoft Excel 2013 has been shown to be sufficient for data collection and analysis tasks such as the descriptive statistics necessary for this study (Liengme, 2015). The simplicity of using Microsoft Excel 2013 may also contribute to the perceived usefulness of the methodology presented by this study as the use thereof does not require the attainment of overly specialised statistical software.

3.5 ETHICS

Ethics is defined by Shamoo and Resnik (2009) as “standards of conduct that distinguish between right and wrong, good or bad”. The University of Pretoria’s committee for ethical research and integrity commit to attempting to ensure that researchers at the University of Pretoria conduct research that is true to the ethical principles of justice and credibility. The committee guides research to be done in an ethical manner by assessing research proposals and granting ethical clearance for research they deem to be ethical. Ethical clearance was granted for this study; the ethical clearance letter is attached in Appendix B. The research complied with all the conditions set out in the letter.

3.6 CONCLUSION

DSR is described as an approach that adopts a pragmatic research paradigm to develop artefacts that are innovative and solve real-world problems (Hovorka, 2009). It was used in this study to develop a methodology to conduct CIUEs.

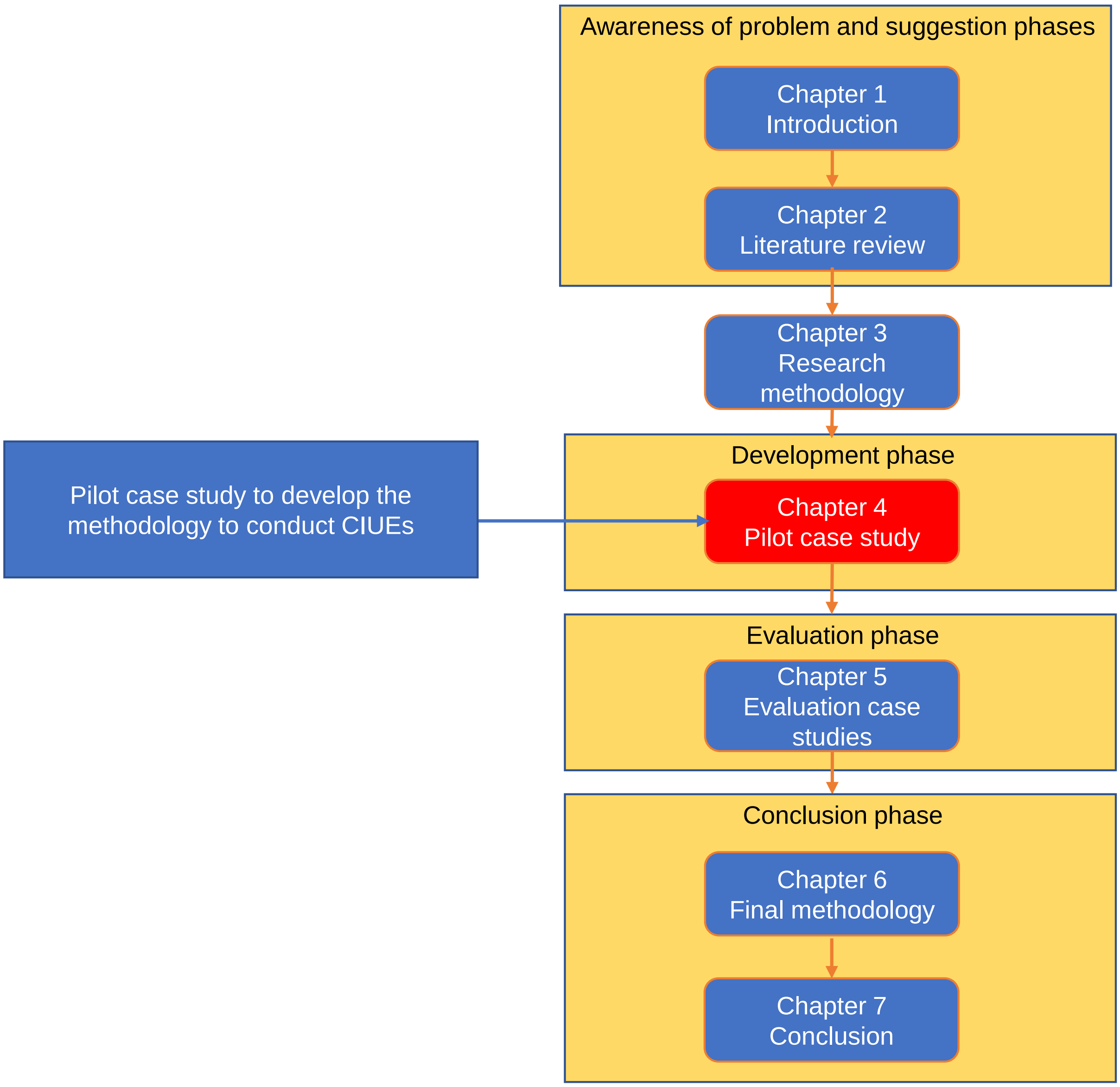
This was done by following the guidelines to DSR presented by Kuechler and Vaishnavi (2008). As per these guidelines, the phases of DSR are 1) awareness of the problem, 2) suggestion of an artefact, 3) development of the artefact, 4) evaluation of the artefact and 5) presentation of the final artefact.

Multiple-case study research was applied in the development and the evaluation of the artefact produced by the DSR methodology. One case was used in the development of the artefact and two cases in the evaluation of the artefact. The artefact produced was a methodology to conduct CIUEs. An eye tracker was used in each case to collect data relating to specific usability indicators that could be compared. The outcome of each case was a CIUE. These CIUEs were used to refine the methodology and to determine performance metrics.

When the data collection and the data analysis associated with the first case were completed, a pilot methodology to conduct CIUEs was presented. The two cases and the interviews that followed were used to refine and evaluate the pilot methodology initially presented by the first case. When this process was completed, a final version of the methodology to conduct CIUEs was compiled. In the next three chapters, the development of the pilot methodology to conduct CIUEs, the evaluation thereof and the final version of the methodology to conduct CIUEs will be discussed.

CHAPTER 4: THE PILOT CIUE CASE

The position of chapter 4 in the dissertation



4.1 INTRODUCTION

This chapter discusses the pilot Comparative Inter-Organisational Usability Evaluation (CIUE) case. The data collected and analysed from this CIUE was used to develop a pilot methodology to conduct CIUEs. Section 4.2 explains the procedures followed in this case study. Sections 4.3 and 4.4 respectively describe the data collected and data analysed for this CIUE case. Finally, changes to the methodology to conduct CIUEs, which was suggested in CHAPTER 2:, are proposed in Section 4.5.

The first case involved a CIUE of the websites of three mobile communications companies in South Africa. The task that was chosen for users to complete was to top up their mobile data balance. A group of participants was selected to complete the task for each of the companies' websites. Data was collected from 45 participants and analysed by comparing usability evaluation indicators collected from the three mobile communications companies' websites.

4.2 PILOT CIUE CASE PROCEDURES

These procedures were based on those in the findings of Holmqvist et al. (2011). Adaptions, such as the involvement of competing organisations, were made where necessary. The procedures for collecting data for this CIUE case were as follows:

1. The competing organisations that were involved in the CIUE were selected.
2. A data collection plan was set up. The plan included the participants that were selected, the technologies that were used, which tasks to test on each information system and how the participants contributed to the evaluation of the usability of tasks on the information systems.

3. Participants were selected based on the data collection plan, typically using the snowball or convenience participant selection method.

4. The usability laboratory was prepared before the participants arrived. This included setting the lighting, testing the eye tracking computer, setting up the eye tracker, testing the eye tracker, arranging participant consent forms and checking the participation rewards (treats) for quality.

5. The Tobii studio project was set up for the pilot CIUE case and a separate test was created for each company. During the set-up process, a test participant was invited to the usability laboratory. The facilitator ran a test evaluation and calibrated the eye tracker with the test participant to make sure that everything is in order.

6. The participants engaged in the study individually. On arrival, the participant was invited into the laboratory, greeted and asked to sign a consent form. The consent form is attached in Appendix C.

7. The participant was informed of the task that he or she had to complete. They were then requested to sit in front of the computer where the eye tracker was calibrated according to the Tobii guidelines (Tobii, 2016b).

8. The user then performed the task and the facilitator pressed escape when the predefined endpoint was reached.

9. If interested, participants were given the opportunity to watch the recording made of their use of the information system.

10. The facilitator then interviewed the participant to gather participant data. The interview questions in the interview are available in Appendix D.

11. Finally, the participant received a treat as a token of appreciation.

When all the participants for a test had gone through the process described above, the data analysis phase commenced. The data analysis procedures were as follows:

1. For each participant, the data elements identified as applicable to the evaluation were gathered from Tobii Pro Studio and inserted into the appropriate section of a Microsoft Excel workbook. This included data such as time taken to complete the task and number of fixations on a specific page. All the metrics collected in this way are listed in Appendix A.
2. The researcher analysed the post-test interview recordings for additional data that pertained to the usability of the application. This data was then inserted into appropriate sections of the Microsoft Excel workbook. The workbooks for the respective cases are attached in Appendix E. Each workbook contains sheets for each respective company's data and a sheet that contains combined comparative data.
3. The comparative data was then analysed and visualised on an indicator by indicator basis. For example, the average time to complete the task for Company 1 vs Company 2 vs Company 3 was represented on Microsoft Excel charts.

4.3 DATA COLLECTION: CASE 1

In the first case, data was collected from 45 participants. Users were requested to top up their data balance using a mobile communications companies' website. Data relating to the indicators described in Appendix A was collected from the participants.

4.3.1 PARTICIPANT DISTRIBUTION DATA

Tables 4.3.1, 4.3.2 and 4.3.3 contain participant data for each company evaluated in the first case respectively. These tables include columns for the participant identification number, age, gender and the participants' qualification(s).

The Company 1 participants' ages ranged from 14 to 60. Of the 15 participants eight were female and seven were male. The participants' qualifications ranged from high school students without qualifications to participants with master's degrees (see Table 4.3.1). The participant identification numbers used in this chapter come directly from Tobii studio and may therefore not be sequential.

Table 4.3.1: Company 1 participants

Identification	Age	Gender	Qualification
3	23	Male	Informatics Honours Student
4	22	Female	Admin Assistant
5	21	Female	Informatics Student
6	22	Male	BIT Honours Student
7	60	Female	Secretarial Diploma
9	24	Male	Degree
11	60	Male	Engineering Degree
12	27	Male	Engineering Degree
13	38	Male	Qualified Project Manager
14	33	Female	Master's Degree
15	14	Female	School
16	17	Female	School
18	37	Male	MBA
19	37	Female	Degree
20	58	Female	Business Owner

For Company 2 16 participants participated of which seven were male and nine were female. Their ages ranged from 20 to 47 and the qualifications that participants had ranged university students with a high school qualification to professors. See Table 4.3.2.

Table 4.3.2: Company 2 participants

Identification	Age	Gender	Qualification
5	20	Male	Student
6	36	Female	Lecturer
7	43	Male	Lecturer
8	21	Male	Student
9	20	Female	MBCHB Student
10	21	Female	MBCHB Student
11	47	Female	Project Manager
12	43	Female	Business Analyst
13	32	Female	Business Analyst
14	36	Male	Professor
15	25	Female	Finance Student
16	22	Female	Medical Science Student
17	24	Male	Lawyer
18	21	Male	Engineering student
20	32	Female	Administrator
21	20	Male	Student

The 14 participants used for Company 3, included five males and nine females. Their ages ranged from 14 to 42 and their qualifications ranged from high school students without qualifications to participants with master's degrees. Details appear in Table 4.3.3.

Table 4.3.3: Company 3 participants

Identification	Age	Gender	Qualification
1	21	Male	Student
2	41	Female	Medical Professional
3	14	Female	High School
5	42	Female	Business Analyst
6	22	Male	Honours BIT Student
7	22	Male	Student
9	22	Female	Admin Staff
10	27	Female	Medical Student
12	19	Female	Finance Student
13	21	Female	Finance Student
14	22	Male	Student
15	22	Female	Drama Teacher
16	31	Female	Engineering Degree
17	42	Male	Archaeologist

4.3.2 USABILITY DATA COLLECTED FROM PARTICIPANTS

This section discusses the data collected from evaluating the use of the systems for the three different companies in the first CIUE case. To ensure that the data is comparable, the same indicators of usability were collected from all three companies. For each company, there are two tables containing summaries of the data collected (Tables 4.3.4 through 4.3.8 and appendix F).

The first table contains columns for user identification, the sample percentage and the total time spent on the information system. The sample percentage column relates to the percentage of time that the eye tracker was able to detect the user's fixations or saccades (Tobii, 2016a). The third column refers to the total amount of time that the user spent to complete the task.

Furthermore, the table contains data relating to the home screen and the top up screen. The top up screen is the final screen where the user completes the task. The sections relating to the home screen contains columns that are titled "Time until first click" and "Number of fixations until first click". These columns contain the amount of time each user took to click the first time on the home screen as well as the number of fixations that were tracked before the user clicked the first time.

The top up section of the table contains the amount of time that the top up page took to load before the user could start working on completing the task (topping up) and a column relating to the number of fixations that were tracked before the user successfully completed the assigned task.

The second table for each company contains raw data flow analysis. This is the path that each user took to complete the task to top up his or her airtime balance. Again, there is a column to identify the user and there are two columns for every page that the user visited – one containing the number by which the page can be identified and one for time the user spent on each page. The URLs of the pages that were visited are not mentioned to secure the anonymity of the companies involved. The raw data flow analysis tables for Company 1 and 3 are large and are therefore appended in Appendix F.

The sample percentage that is mentioned frequently in the upcoming sections refers to the percentage of the total time that the eye tracker successfully recorded the participant's eye fixation or saccade (Tobii, 2016a). The sample percentage for the Company 1 usability data collected ranged from 78% to 97%, all the data collected could be used as the minimum threshold was set at 60%. The total time spent on the site to complete the task for Company 1 ranged from 50 seconds to 351 seconds. The average for the time to complete task was 164 seconds.

Furthermore, for Company 1 the quickest participant to decide where to click the first time was participant 11 who clicked the first time after 5.64 seconds. The number of fixations recorded from this participant was also the least and was recorded at 12 fixations. The participant that took the longest to decide where to click was participant seven who took 25.40 seconds. The participant with the most number of fixations was participant 4 with 117. On average users took 14 seconds and 48 fixations before deciding to click the first time.

On the final screen, the participants took an average of 161.67 seconds to finish the task, which was to top up their data balance. The participant with the most fixations had 517. It must however be said that the page took the longest for that participant to load at three minutes and six seconds. The participants that fixated the least on the final screen were participants 5 and 12 for who 69 fixations recorded. The loading times of the page were also the least for those two participants.

Table 4.3.4: Company 1 general data collected

User identification	Sample percentage	Time to complete task	Time until first click	Number of fixations until first click	Top up screen: Loading time	Top up screen: Number of fixations until top up
3	87%	00:05:51	13.96	48	00:01:39	280
4	92%	00:05:06	32.98	117	00:00:40	85
5	88%	00:01:35	8.11	34	00:00:26	69
6	92%	00:03:38	9.21	36	00:00:54	112
7	90%	00:03:39	25.40	90	00:00:41	83
9	97%	00:01:10	10.04	38	00:00:38	105
11	94%	00:05:18	5.64	12	00:03:06	517
12	93%	00:01:11	7.96	32	00:00:31	69
13	97%	00:03:28	21.31	70	00:00:53	160
14	96%	00:03:59	16.41	58	00:01:03	148
15	78%	00:02:42	7.06	28	00:00:31	162
16	96%	00:00:50	5.93	23	00:00:28	77
18	96%	00:01:18	22.10	75	00:00:34	88
19	86%	00:03:17	7.71	28	00:01:15	200
20	94%	00:05:32	12.04	34	00:01:47	270

For Company 2, the lowest sample rate recorded was 40% and the highest was 98%. There was only one participant's sample percentage under the minimum of 60%, so the data collected from that participant was not used. Participant 14 took the longest to complete the task at five minutes and 30 seconds. Participant 15 completed the task the quickest at 21 seconds. The average time spent on the Company 2 site to complete the assigned task was 95 seconds.

Regarding the time until first click, participant 20 took the longest to decide where to click the first time and did it in 82.50 seconds. The minimum time taken to click the first time was that of participant 14 who was, incidentally, also the participant that took the longest to complete the task. The average time taken to decide where to click the first was 19.47 for Company 2. Participant 20 was also the participant for which the sample rate was only 40%, however, this does not have an influence on the usefulness of indicators regarding time, as the sample rate does not have an influence on the data related to the time spent by the users to complete each goal. However, it may be that the participant was distracted and did not focus on the screen. As such, the participant may have taken longer to decide where to click.

Regarding fixations, participant 20 must be disregarded as the sample percentage has an influence on the number of fixations that were recorded. The maximum value was that of participant 10 which was 158 fixations, the minimum value was that of participant 14 who only registered 7 fixations before clicking the first time. The average user fixated 57.54 times before clicking the first time. For all the participants, the page loaded in 26 seconds or less. Participant 16 had the least number of fixations on the final screen to complete the task and participant 6 had the most number of fixations with 73 fixations. The average user completed the task using 35.61 fixations on the final screen.

For the Company 3 site, all the sample percentages were more than 60%, The entire task was completed in 14 seconds by participant 6, who was the quickest participant to complete the task. The participant who took the longest was participant 7, who completed the task in four minutes and 20 seconds. The average participant took 44 seconds to complete the task on the Company 3 website.

Furthermore, regarding the time that participants took to click the first time on the home screen, the quickest participant was participant 7 who did it in 2.12 seconds. Participant 5 took the most amount of time and spent 19.88 seconds before clicking the first time. The average participant took 11.03 seconds to click the first time. The minimum and maximum values for the time spent before clicking the first-time correlates to the number of fixations recorded until clicking the first time. The highest number of fixations was recorded for participant 5 before clicking the first time which was 67 fixations. The least number of fixations was recorded for participant 7 which was five fixations.

Table 4.3.5: Company 2 general data collected.

User Identification	Sample percentage	Time to complete task	Time until first click	Number of fixations until first click	T Top up screen: Loading time	Top up screen: Number of fixations until top up
5	88%	00:01:56	25.07	83	00:00:18	46
6	85%	00:00:48	17.46	60	00:00:26	73
7	98%	00:02:44	21.98	57	00:00:13	48
8	72%	00:00:45	16.73	56	00:00:26	67
9	76%	00:01:01	16.38	55	00:00:07	17
10	78%	00:01:34	45.26	158	00:00:08	17
12	92%	00:01:49	20.50	68	00:00:10	34
14	93%	00:05:30	2.48	7	00:00:09	37
15	76%	00:00:21	10.48	14	00:00:07	33
16	80%	00:01:24	16.39	40	00:00:04	13
17	87%	00:00:39	14.31	38	00:00:06	16
18	98%	00:01:31	25.88	64	00:00:14	45
20	40%	00:01:37	82.50	244	00:00:04	4
21	95%	00:00:33	20.27	48	00:00:05	17

On the final screen, it took 13 fixations for participant 15 to finalise the task, which was the minimum, and it took 44 fixations for participant 17 to complete the task, which was the maximum. The average number of fixations to complete the task on the final screen was 26.92 and the sites final screens all loaded in less than 20 seconds.

Table 4.3.7: Company 3 general data collected.

User Identification	Sample percentage	Time to complete task	Time until first click	Number of fixations until first click	Top up screen: Loading time	Top up screen: Number of fixations until top up
1	82%	00:00:37	15.74	48	00:00:10	38
2	76%	00:00:21	11.65	39	00:00:06	30
3	94%	00:00:32	12.36	40	00:00:07	24
5	93%	00:00:34	19.88	67	00:00:09	40
6	94%	00:00:14	6.19	17	00:00:05	19
7	84%	00:04:20	2.12	5	00:00:04	15
9	87%	00:00:23	16.81	53	00:00:05	20
10	90%	00:00:38	9.67	32	00:00:06	17
12	86%	00:00:22	14.56	39	00:00:06	21
13	86%	00:00:24	13.86	43	00:00:08	31
14	94%	00:00:18	9.23	20	00:00:07	31
15	76%	00:00:44	2.79	11	00:00:04	13
16	93%	00:00:20	9.79	28	00:00:09	34
17	83%	00:00:25	9.85	30	00:00:12	44

The table relating to raw data flow analysis for Company 1 and 3 is attached in Appendix F. Participant 20 visited a total of 23 pages to finally complete the task and participant 3, 5, 9, 12, 13, 14, 16, 18 and 19 visited only nine pages, this was the minimum. The average number of pages visited on the Company 1 website was also nine.

The Company 2 raw data flow analysis can be found in Table 4.3.8. For Company 2 the average number of pages that participants visited was four. The participant that visited the most number of pages was participant 7 who visited 10 pages. All the other participants only visited two pages to complete the task.

Finally, regarding the raw data flow analysis for Company 3, the average number of pages that participants visited was also four. The participant that visited the most number of pages was participant 14 who visited 10 pages. Participants 6, 8, 10, 15 and 20 visited two pages to complete the task, which was the minimum.

Table 4.3.8: Company 2 raw data flow analysis.

Raw data Flow Analysis																				
Steps	1		2		3		4		5		6		7		8		9		10	
	Page	Time Elapsed on page	Pag e	Time Elapsed on page	Pa ge	Time Elapsed on page	Pa ge	Time Elapsed on page	Pag e	Time Elapsed on page	Pa ge	Time Elapsed on page	Pa ge	Time Elapsed on page	Pa ge	Time Elapsed on page	P a g	Time Elapsed on page	Pa ge	Time Elapsed on page
User1	1	00:00:26	2	00:00:11																
User2	1	00:00:13	2	00:00:08																
User3	1	00:00:20	3	00:00:11																
User5	1	00:00:24	2	00:00:10																
User6	1	00:00:09	2	00:00:05																
User7	1	00:00:38	4	00:00:18	5	00:00:20	6	00:00:28	1	00:00:28	7	00:01:25	5	00:00:07	6	00:00:13	1	00:00:18	2	00:00:05
User9	1	00:00:18	2	00:00:05																
User10	1	00:00:31	2	00:00:07																
User12	1	00:00:16	2	00:00:06																
User13	1	00:00:16	2	00:00:08																
User14	1	00:00:11	2	00:00:07																
User15	1	00:00:39	2	00:00:05																
User16	1	00:00:11	2	00:00:09																
User17	1	00:00:12	2	00:00:13																

4.4 DATA ANALYSIS

This section contains the analysis of the data collected for the first case study that was used to develop the methodology to conduct CIUEs. Each of the usability indicators selected were compared across the three companies that were evaluated. The comparison is demonstrated with the use of a bar chart where appropriate. The indicators that were analysed included the time to complete the top-up task, the amount of time until first click, the number of fixations until first click, and the number of pages visited.

4.4.1 TIME TO COMPLETE TASK

The averages, medians, maximums and minimums regarding the time to complete the top up task appear in Table 4.4.1. The average participant took 194 seconds to complete the task for Company 1, 95 seconds for Company 2 and 44 seconds for Company 3. The average participant took twice as long to complete the task for Company 1 as compared to Company 2. The average participant took twice as long to complete the task for Company 2 when compared to Company 3.

This indicates that the average participant could complete the task most efficiently on the Company 3 site. If Company 1 and Company 2 have the desire to optimise the efficiency of their system use, they will realise from the results of this comparison that it is possible.

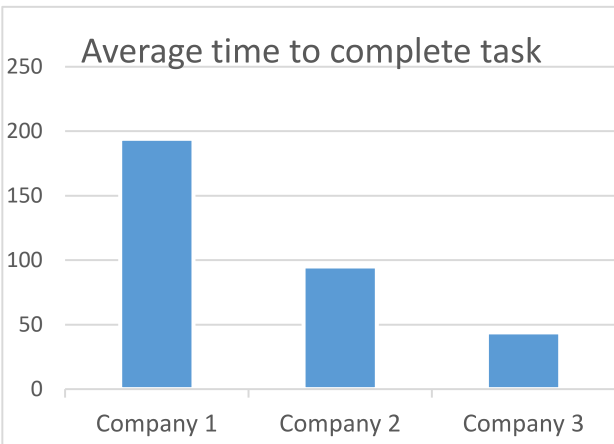


Figure 4.4.1: Average time to complete task.

For Company 1 and 2 the median values correlate to the averages. However, in Company 3's case the median is notably lower than its average. This can be explained by participant 7's data. This participant took 260 seconds to complete the task, which is a lot longer than all the other users and thus skews the average. The median values are demonstrated in Figure 4.4.2. When the median is used instead of the average to eliminate the effect of this outlier, Company 3 performed even better than was previously reported.

Table 4.4.1: Analysis of time to complete task

Company	Average	Median	Minimum	Maximum
1	194	208	50	351
2	95	88	21	330
3	44	25	14	260

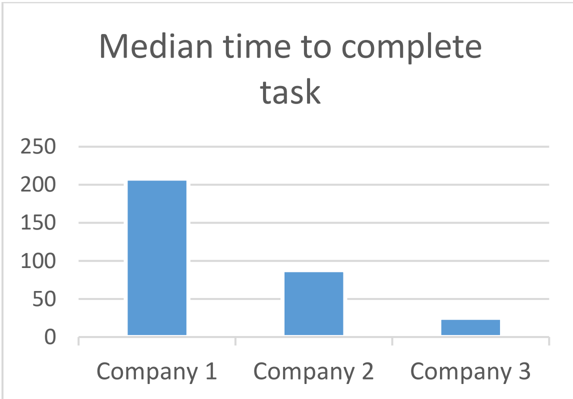


Figure 4.4.2: Median time to complete task.

4.4.2 TIME UNTIL FIRST CLICK

The averages, medians, maximums and minimums of the time until first click appear in Table 4.4.2. The average and the median user for Company 2 spent the most time before clicking the first time when compared to Companies 1 and 3. These values are almost double for Company 2 in comparison to Companies 1 and 3. The average user for Company 3 spent 11 seconds before clicking the first time and the average user for Company 1 spent 14 seconds before clicking the first time, which is only a difference of 3 seconds. This insignificant difference is also shown by the median values that tell a similar story.

Table 4.4.2: Analysis of time until first click.

Company	Average	Median	Minimum	Maximum
1	14	10	6	33
2	24	19	2	20
3	11	11	2	82

The comparison of the averages and median values is demonstrated in Figure 4.4.3 and 4.4.4. The long time until first click for Company 2 may have been due to the Company 2 home screen being cluttered. There were many screen elements that the users needed to analyse before clicking the first time. This is clear when compared to the values of its competitors. This may be a place to start if Company 2 wants to optimise the usability of their website.

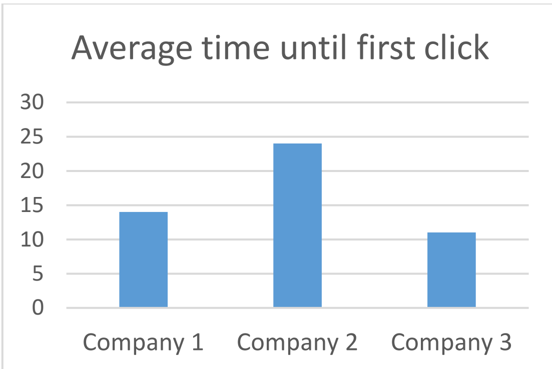


Figure 4.4.3: Average time until first click in seconds

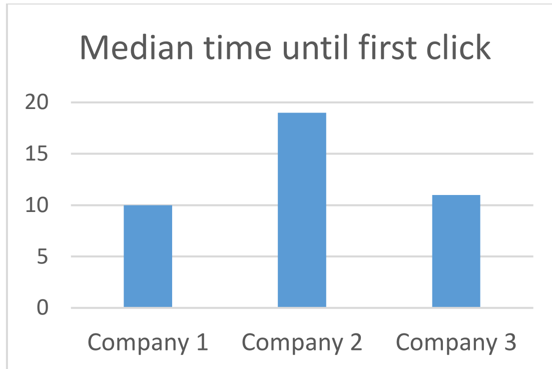


Figure 4.4.4: Median time until first click in seconds

4.4.3 FIXATIONS UNTIL FIRST CLICK

The number of fixations before a participant clicks the first time confirms the results based on the time to first click. The average number of fixations before clicking the first time for Company 2 was 58 which is close to double the 34 fixations of Company 3, which in turn had 14 fixations less on average than Company 1. The median number of fixations of Company 1 and 3 is the same, so the slight difference in the average may be regarded as insignificant. However, the difference between Company 2 and the rest is significant.

Table 4.4.3: Analysis of fixations until first click

Company	Average	Median	Minimum	Maximum
1	48	36	12	117
2	58	56	7	158
3	34	36	5	67

All the values gathered regarding the number of fixations until first click are outlined in Table 4.4.3. The comparison of the averages and median values is also demonstrated in Figure 4.4.5 and 4.4.6. Similarly, to the time until first click, the values for Company 2 are high because the Company 2 home screen was cluttered.

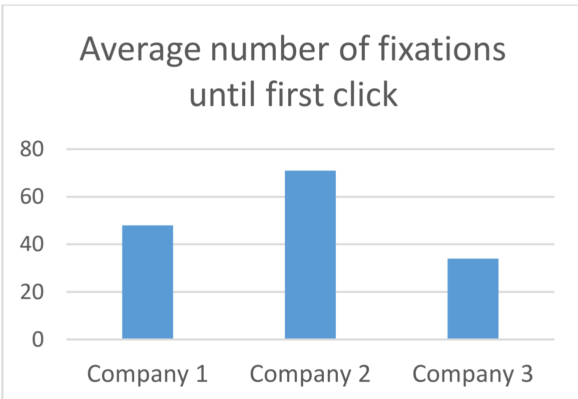


Figure 4.4.5: Average number of fixations until first click.

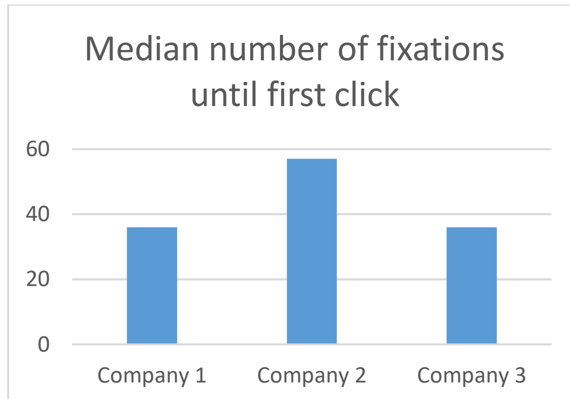


Figure 4.4.6: Median number of fixations until first click.

4.4.4 NUMBER OF PAGES VISITED

Table 4.4.4 outlines the average number of pages visited by participants for each company. This is also demonstrated by Figure 4.4.7. It was found that in the average case for the Company 1 participants, the number of pages visited was three times the number of pages visited to complete the same task for companies 2 and 3 respectively. This result correlates with the average number of fixations and the amount of time spent to complete the task.

This shows that either there is a problem with the information architecture or the clarity of paths on the Company 1 website or it is not possible to complete the task without visiting that number of pages. However, participant 12 completed the task by visiting four pages, so it is possible to complete the task visiting far fewer pages than the average and median values suggest. This indicates that there is a clear problem in navigating through the optimal path to complete the task on Company 1's website.

Table 4.4.4: Analysis of the number of pages visited.

Company	Average
1	9
2	3
3	3

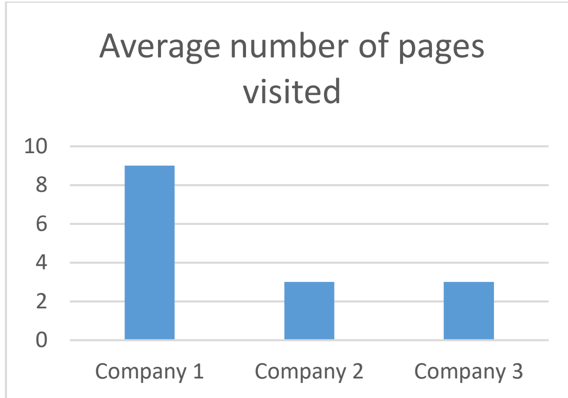


Figure 4.4.7: Average number of pages visited.

Minimum and maximum values relating to all the metrics, the results are not viable as metrics to analyse the performance of an information system as they are participant dependent. The participants vary in age, technological exposure and expertise, and this could have a significant influence on their performance. This simply means that it would be difficult to compare minimums and maximums between different company interfaces because of the significant role one participant could play on the results.

What maximum and minimum values do show, however, are the possibilities in terms of the participant group. For example, regarding the time to complete the task, for Company 1 it was possible for a participant in that participant group to finish the task in 50 seconds, for Company 2 it was possible for a participant to finish the task in 21 seconds and for Company 3 it was possible for a participant to finish the task in 14 seconds.

This correlates to the average amount of time it took a participant to complete the task, the number of fixations and the number of pages visited. The maximum amount of time it took a participant to complete the task was also the highest for Company 1. This correlates with the findings from the other indicators.

With regard to the amount of time until first click, for Company 1, 2 and 3 it was possible for a participant in that participant group to decide to click the first time in six, two and three seconds respectively. The maximum here diverted from the pattern seen thus far – the maximum amount of time was the highest for Company 3 and the lowest for Company 2. The maximum values refer to individual participants' usage of the site and it is therefore not advised to read too much into it.

The time to load the page was taken into consideration on the top up page. The values as reported above, in sections 4.4.1 to 4.4.4, were recorded from the moment the page had loaded completely. With regard to the number of fixations, the analysed values are as in Table 4.4.5.

Table 4.4.5: Top-up page total number of fixations.

Indicator	Company 1	Company 2	Company 3
Minimum	69	13	4 (invalid value)
Maximum	517	44	73
Average	109	27	34
Median	162	27	33

The average and the median number of fixations for Company 1 are far higher than that of companies 2 and 3. This also correlates with the minimum and maximum values. The minimum number of fixations for Company 3 is invalid because the fixation sample percentage for that candidate was below the minimum accuracy level of 60%.

The raw data flow analysis is an indication of how far a participant dwelled from the optimal path to complete the task. Again, the results are consistent with the results relating to the metrics discussed above. None of the participants for the Company 1 site completed the task in the optimal number of pages. For Company 2, only one user did not complete the task in the optimal number of pages. For Company 3, all but three of the participants completed the task in the optimal amount of time. This may indicate an information architectural problem with the Company 1 site regarding the user task to top up their data balance.

The above is confirmed by the average number of pages visited for Company 1 that are three times that of Company 2 and 3.

4.5 A METHODOLOGY FOR A CIUE

In this section, the development of the research artefact (a methodology to conduct a CIUE) is discussed. This development is based on the suggestion made in section 2.10 and the results of the case where a CIUE was applied as discussed in Sections 4.1 to 4.4 above. In Section 2.10 the philosophy, approach, strategy, time horizons and steps and procedures to conduct CIUEs were discussed. This section will add to or refine the steps and procedures defined in section 2.10. The whole methodology to conduct a CIUE, including the elements from section 2.10, will not be described here again to avoid repetition.

It was recommended in section 2.10 that a controlled user evaluation environment be used for CIUEs. This was suggested to minimise the influence of variables that are not related to the information systems being evaluated. For example, the lighting, smell and the sound in the room should be similar for all the experiments. The state of the computer being used should be the same for all the participants. This includes the operating system and other software that may influence the user's interaction.

Furthermore, it is recommended that as much data is collected as possible. One can always remove unnecessary data, but once the participant has completed the experiment, new data cannot be collected from that participant again (the data will be influenced by his or her previous interaction). Additionally, it is also acknowledged that it is important to gather data that could potentially confirm or contradict other data. This helps determine the accuracy of the experiments.

To comply with these guidelines, an eye tracker was used in all the case studies of this research study. Eye tracking is a not requirement to conduct a CIUE but is a useful tool to optimise the amount of data gathered. An eye tracker allows you to record time accurately, record the fixations of the participant, record and manage videos of each participant meticulously and inspect system navigation patterns accurately. There are other kinds of data that could be gathered from the eye tracking software, but when data is selected for a CIUE, quantitative data should be favoured to avoid the influence of a subjective opinion.

From above, the guidelines to conducting a CIUE are as follows:

1. The competing organisations that will be involved in the study should be selected.
2. Initially a plan should be set up for the CIUE. This involves determining the data that will be collected, the participants that will be used, where the experiments will take place and a timeline of events. The planning phase should also contain the selection of tasks that participants will be requested to complete on each of the information systems that are to be compared. The tasks should be similar in terms of user outcome.
3. The controlled environment should be set up in a manner that will ensure that it remains in a consistent state throughout all the experiments. You can never ensure that the environment is precisely the same for all the participants (e.g. time changes) but it should be kept as similar as possible. The recording tools should also be set up.
4. Participants should be grouped to ensure that there is diversity between participants for each information system being evaluated. The number of participants used for each

information system should be similar, an attempt should be made to keep it the same. It is advisable to recruit more than the required number of participants to cater for those who fail to arrive and those whose eye tracking recordings fall under the required recorded fixation percentage.

5. The experiments should be conducted in a manner that is replicable for each participant.

6. Once the experiments are completed, the data should be extracted from the recording tools that were used.

7. The data should be imported into a data analysis tool that will be used. Ensure that the tool can be used to do the statistical analysis that will be required.

8. Suitable statistics should be used to compare the data gathered.

9. Present the compared data in a manner that is easily perceivable by readers. The use of simple charts may be appropriate here, then compile the results in the form of a report or a presentation. The above process is outlined in Figure 4.5.1.

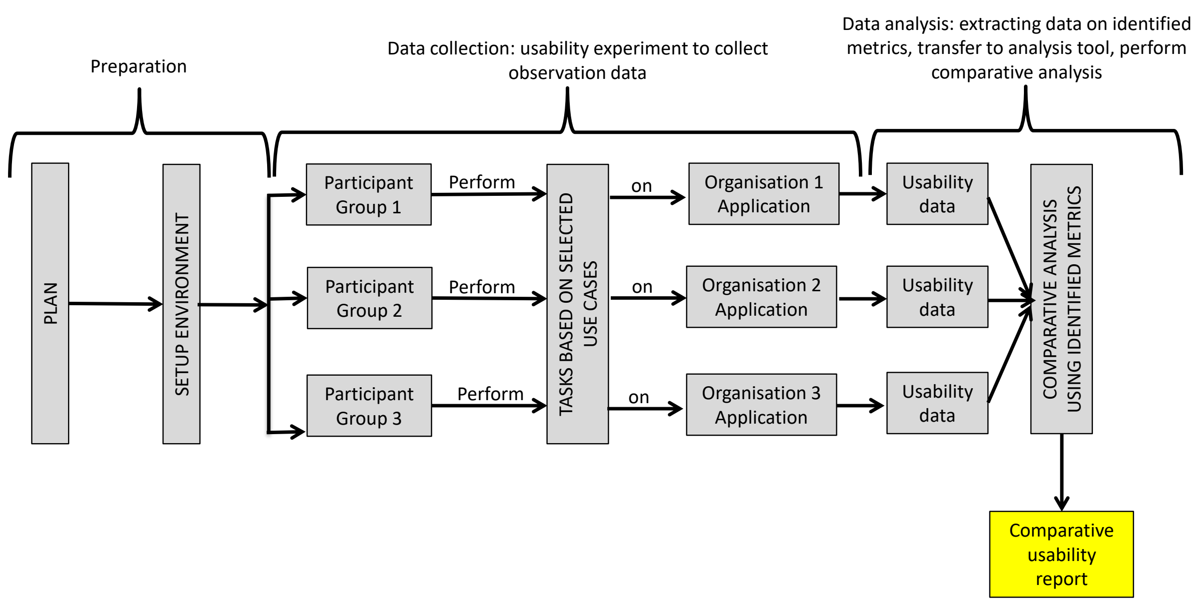


Figure 4.5.1: Representation of the CIUE process.

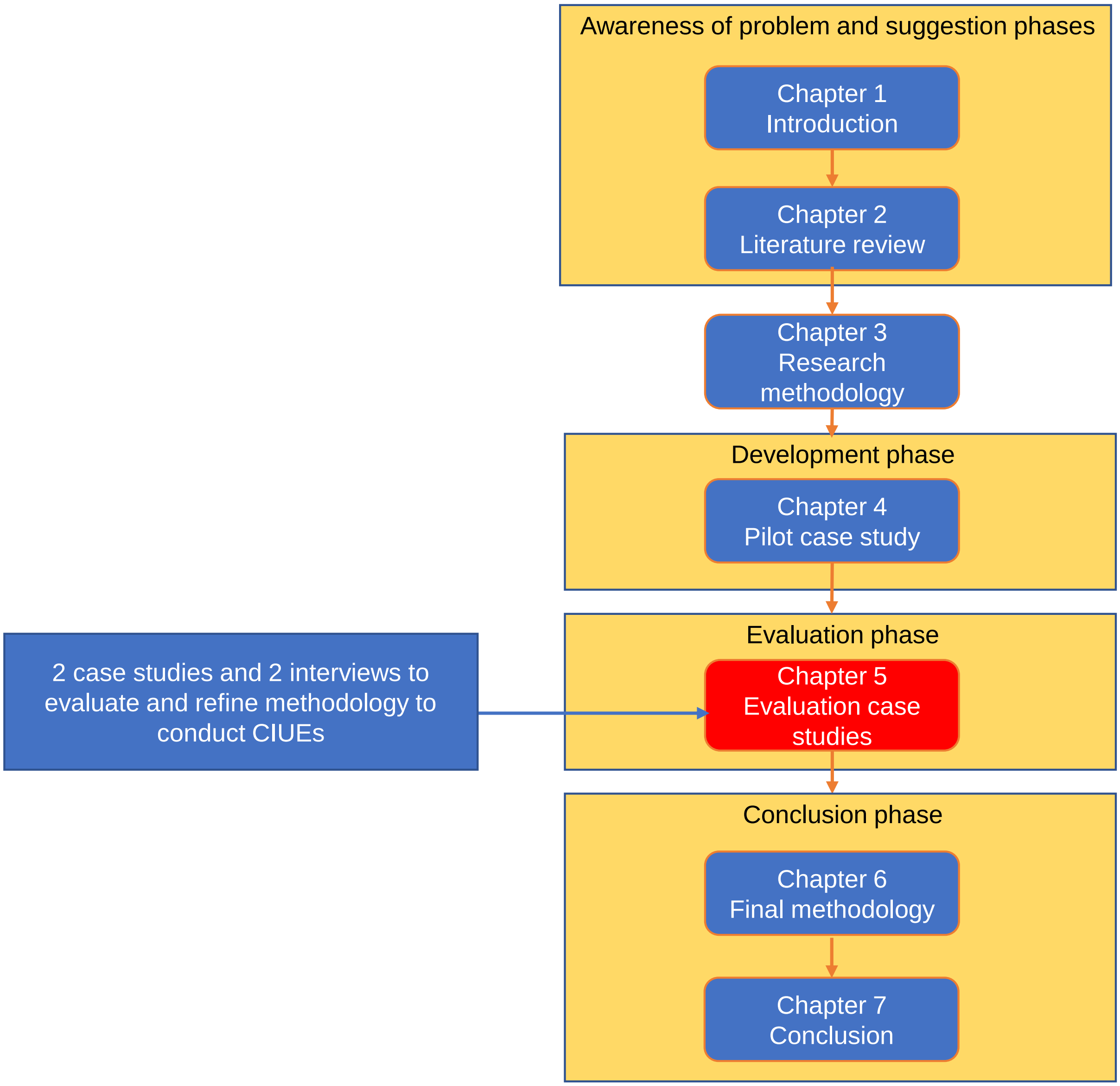
4.6 CONCLUSION

In this chapter, the methodology to conduct a CIUE was further developed. This development was based on the initial suggestion made in Chapter 2 and a case of application that was conducted prior to the development of the artefact in this chapter. This case involved a heterogenous group of 45 participants and three mobile communications companies.

The participants were observed using an eye tracker. The data that was analysed included the time to complete an assigned task, the number of fixations to complete the task, the time until first click, the number of fixations until first click and the number of pages visited. This data was compared between each organisation that was selected and various deductions were made that contributed to the process of creating the methodology.

CHAPTER 5: EVALUATION OF THE ARTEFACT

The position of chapter 5 in the dissertation



5.1 INTRODUCTION

The purpose of this chapter is to evaluate the methodology to conduct CIUEs that was presented in Chapter 4. This is done to determine whether the methodology to conduct CIUEs is useful and to refine the methodology (Gregor & Hevner, 2013). This process involved two evaluation cases and two evaluation interviews.

Sections 5.2 and 5.3 respectively discuss the two cases, each with data collected, data analysed and conclusion sections. These discussions are organised in a similar way as the discussion of the pilot case study in Chapter 4. After each case discussion, suggestions regarding the alterations to the artefact are explained. The interviews were aimed at assessing the adapted methodology to conduct CIUEs and are discussed in Section 5.4. Some changes were recommended after the interviews.

5.2 THE FIRST EVALUATION CASE STUDY

The first evaluation case involved a Comparative Inter-organisation Usability Evaluation (CIUE) of the information systems of two medical insurance companies in South Africa. Twenty participants were used for this study. These participants were selected using a snowball or convenience sampling method. It was discovered during the pilot case that when the participants differ in terms of, for example, exposure to technology, it can influence the reliability of the evaluation results. A more homogenous participant group was therefore selected for the first evaluation case. Only students with a high technological proficiency from the same university were selected. This way the focus could be on the comparison between the information systems being tested, rather than on the participants.

Tables 5.2.1 and 5.2.2 contain data relating to the participants that were used to conduct the usability evaluations of the respective companies' websites. These tables contain the users' age, gender, educational level, technological devices they frequently use and self-rated technological proficiency.

Participants were interviewed after the usability evaluation. The interview questions are in Appendix D. The answers to the questions from the interviews are summarised in Tables 5.2.1 and 5.2.2.

Table 5.2.1: Company 1 participants

Age	Gender	Educational Level	Technological Devices	Self-rated Technological Proficiency
23	Male	Undergraduate degree	Smart phone and laptop/PC	Very good
22	Female	Undergraduate degree	Smart phone, laptop/PC and tablet	Average
22	Female	Attended university	Smart phone, laptop/PC and tablet	Excellent
22	Male	Undergraduate degree	Smart phone, laptop/PC, video game console and tablet	Very good
20	Male	Undergraduate degree	Smart phone, laptop/PC and tablet	Excellent
21	Male	Undergraduate degree	Smart phone and laptop/PC	Excellent
21	Male	Attended university	Smart phone, laptop/PC, video game console and tablet	Excellent
21	Male	Undergraduate degree	Smart phone, laptop/PC and tablet	Excellent
23	Male	Attended university	Smart phone and laptop/PC	Very good
23	Male	Undergraduate degree	Smart phone and tablet	Very good

Of the 10 participants for Company 1, two were female and the rest were male. The ages of the participants ranged from 20 to 23 and they were all either attending a university or had an undergraduate degree. Furthermore, they all rated their technological proficiency as either very good or excellent except for one participant that self rated her proficiency as average. All the participants owned a smart phone and a laptop or a PC with the exception for one participant that did not own a laptop or a PC but owned a tablet in addition to the smart phone. Of the 10 participants seven owned a tablet and two owned a video game console.

The same number of participants was selected for Company 2. The ages of the participants also ranged from 20 to 23. Of the 10 participants, nine were male and one was female. They all either attended university or were busy with an undergraduate degree, except for one of the participants who was busy with a post-graduate degree. All the participants owned a smart phone and a laptop, four of the participants also owned a video game console and four also owned tablets. Their self-rated technological proficiency ranged from very good to excellent.

Table 5.2.2: Company 2 participants

Age	Gender	Educational Level	Technological Devices	Self-rated Technological Proficiency
20	Male	Undergraduate degree	Smart phone, laptop/PC, video game console and tablet	Very good
20	Male	Attended university	Smart phone, laptop/PC and video game console	Excellent
20	Male	Attended university	Smart phone and laptop/PC	Very good
21	Male	Post graduate degree	Smart phone and laptop/PC	Excellent
21	Male	Undergraduate degree	Smart phone and laptop/PC	Excellent
23	Male	Attended university	Smart phone, laptop/PC and tablet	Very good
20	Male	Undergraduate degree	Smart phone	Very good
20	Female	Attended university	Smart phone and laptop/PC, video game console and tablet	Excellent
21	Male	Undergraduate degree	Smart phone and laptop/PC and video game console	Excellent
22	Male	Undergraduate degree	Smart phone and laptop/PC	Excellent

5.2.1 DATA COLLECTED

This section discusses the data collected in the first evaluation case involving two medical insurance companies in South Africa. There were 20 participants that participated in the study. The users were requested to complete a task with the use of a scenario. The scenario was:

"Imagine you are a father or a mother and you are part of the rewards program at company X, find all the rewards that may motivate your children to live a better lifestyle".

The outcome of the task was to find data relating to rewards that could be gained from joining the rewards programs of the medical insurance companies evaluated. It should be noted that the sample percentage data was not collected in this case as in the pilot case described in chapter 4. Where in the pilot case all the participants were included, participants in this case were filtered out if they did not meet the 60% threshold (Tobii, 2016a). The page load times were measurement activity was also removed as it was decided to manage the internet speed in the controlled environment to be consistent, which renders that attribute irrelevant.

The data was collected in a variety of ways, the first being with the use of an eye tracker. Tables 5.2.3 and 5.2.4 contain data that was collected for the two companies respectively. The tables list the participant number, which is used for identification. The time to complete the provided task, the total number of fixations recorded for the whole task, the time until the first click, the number of fixations recorded before the first click, the number of pages visited, the number of breaks and an indicator of optimal completion (i.e. could they perform the whole task successfully or not) are also displayed.

The number of pages visited replaces the raw data flow analysis from the pilot study. This simplifies the process of conducting a CIUE. Flow analysis is also very particular to the interface being evaluated and may introduce a subjective bias in interpretation when a comparison is done. For example, the flow between the screens of Company 1 could depend heavily on the nouns used in links, whereas Company 2 used verbs in the links on their screens. This cannot be compared as the analysis would depend on the interpreter's opinion of the use of nouns or verbs in links.

The task that the participants had to complete was an information search. There are a variety of dynamics that need to be considered when evaluating an information search. The first dynamic that needs to be catered for is that of optimal completion. With an information search, a user may feel that they have completed the task when they have only retrieved partial information. The user may feel satisfied that the task is complete even though there is still a vast amount of information that could be collected.

On both the websites evaluated in the pilot case, there was a page that contained all the information relating to the search that the users were tasked with. The columns titled "Optimal Completion" relate to these pages. If the user arrived at the page containing all the information as outlined in the outcome of this task, then the optimal completion value for the row relating to that user is set to "Yes". If the user was satisfied with the information that he or she found before arriving at the page containing the full set of information, then this value is set to "No".

Another variable applicable to an information search is the number of times the user experienced a software malfunction or otherwise called a system break. This is recorded to evaluate the viability of the assessment. If a system malfunction occurs, the user is hindered in their execution of the task. As such, indicators relating to time and number of fixations can no longer be used for users that experienced such a software malfunction.

For Company 1, 10 participants were involved in the study. The time they used to complete the assigned task ranged from 120 seconds to 318 seconds. The average completion time was 218.1 seconds on the Company 1 site. The number of fixations recorded for each participant to complete the entire task ranged from 395 to 1172 fixations. The average number of fixations recorded for the participants was 770.5.

Regarding the time until first click for Company 1, this ranged from 3 to 26 seconds, the number of fixations until first click ranged from 17 to 69. The average number of seconds until first click was 11.5 and the average number of fixations was 42.9. Furthermore, the minimum number of pages that participants visited on the Company 1 site was five and the maximum was 14. The average number of pages visited by the participants was 9.2 and a system break did not occur for any of the participants. Finally, of the 10 participants, three participants completed the task optimally and seven did not.

For the Company 2 site, 10 participants were selected for the study. All the participants completed the task optimally and system breaks did not occur for any of the participants. The number of pages that was visited by participants ranged from 2 to 11. The participants completed the task in a number of seconds ranging from 89 to 240. The average number of seconds that was taken to complete the task on the Company 2 site was 169.3.

The number of fixations recorded to complete the task ranged from 309 to 1602 for all the participants. The 1602 value was an outlier and so the average was 609.8 and the median number was lower at 502. Regarding the time until first click, the minimum value recorded was 6 seconds and the maximum value was 26 seconds. The average number of seconds until first click recorded was 15. This correlated to the number of fixations recorded for all the participants. This ranged from 24 to 89 and the average value was 53.3.

Table 5.2.3: Company 1 data collected

Participant number	Time to complete (in seconds)	Number of fixations	Time until first click (in seconds)	Fixations until first click	No pages visited	Number of breaks	Optimal completion
1	141	503	19	65	5	0	No
2	229	728	5	17	11	0	Yes
3	314	1098	12	42	10	0	Yes
4	120	395	26	93	6	0	No
5	204	690	6	27	8	0	No
6	318	1172	5	23	14	0	Yes
7	191	677	21	69	9	0	No
8	186	695	4	17	9	0	No
9	246	1039	3	18	12	0	No
10	232	708	14	58	8	0	No

Table 5.2.4: Company 2 data collected.

Participant number	Time to complete (in seconds)	Number of fixations	Time until first click (in seconds)	Fixations until first click	No pages visited	Number of breaks	Optimal completion
1	101	383	24	89	4	0	Yes
2	406	1602	6	24	11	0	Yes
3	108	382	15	53	6	0	Yes
4	170	591	26	86	11	0	Yes
5	129	515	20	78	7	0	Yes
6	175	639	17	60	4	0	Yes
7	153	489	14	41	6	0	Yes
8	89	309	10	42	2	0	Yes
9	122	432	9	30	3	0	Yes
10	240	756	9	30	3	0	Yes

5.2.2 DATA ANALYSIS

This section discusses the data analysis for the first evaluation case. This will be done on an indicator by indicator basis starting with time to complete task in section 5.2.2.1.

5.2.2.1 Time to complete task

As can be seen from Table 5.2.5 and Figures 5.2.1 and 5.2.2 respectively, the median and average times to complete the task were lower for Company 2 than for Company 1. This simply means that the average and median user took more time to find the information required using Company 1's information system as opposed to that of Company 2. A similar story could be told about the minimum values. It was possible for a participant to complete the task in 41 seconds less than the closest competitor from Company 1.

However, in the worst case the Company 2 participant took 88 seconds more to complete the task than the closest competitor from Company 1. It should be noted that the maximum and minimum values are participant dependent and should thus only be used to indicate possibilities. From the minimum values, we know that it is possible to complete the task in as little as 120 seconds on the Company 1 site and 89 seconds on the Company 2 site. However, we also know that it is possible for a user to take up to 318 seconds on the Company 1 site and 406 on the Company 2 site.

Table 5.2.5: Time to complete task.

Company	Average	Median	Minimum	Maximum
1	218,1	216,5	120	318
2	169,3	141	89	406

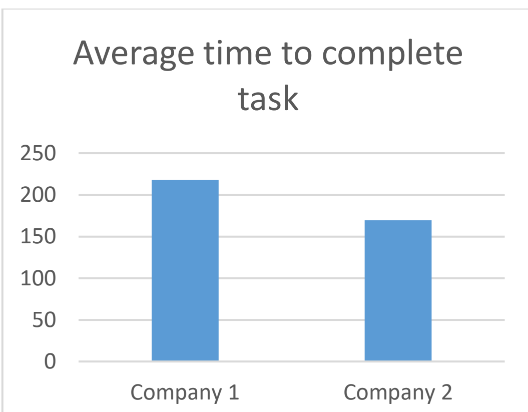


Figure 5.2.1: Average time to complete task.

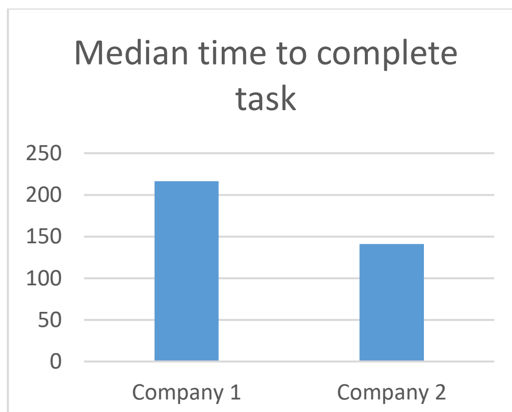


Figure 5.2.2: Median time to complete task.

5.2.2.2 Number of fixations to complete task

The data from the number of fixations to complete task correlates with the time taken to complete the task. The analysis of this data is summarised in Table 5.2.6. Graphical representations of the average and median values appear in Figures 5.2.3 and 5.2.4. A similar trend could be noticed in the average, median, minimum and maximum values when compared to the time to complete the task. The average number of fixations to complete the task was 770.5 for Company 1 which is 160.7 more than the average of 609.8 for Company 2. The difference in values was even greater for the median where it was 199.5 (701.5 vs 502). The maximum number of fixations for the Company 2 site was 1602, which was clearly an outlier.

Table 5.2.6: Fixations to complete task.

Company	Average	Median	Minimum	Maximum
1	770,5	701,5	395	1172
2	609,8	502	309	1602

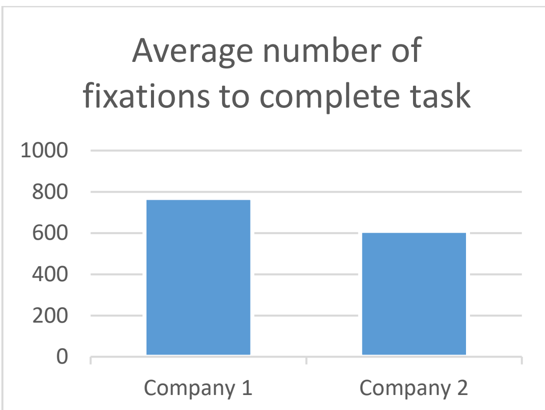


Figure 5.2.3: Average number of fixations to complete task.

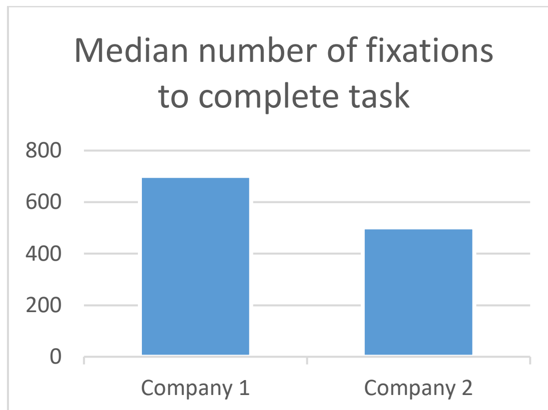


Figure 5.2.4: Median number of fixations to complete task.

5.2.2.3 Time until first click

Regarding the time until first click, the average and the median participant took more time to decide to click the first time on Company 2's website. This is demonstrated in Table 5.2.7 and Figure 5.2.5. There is a bigger difference between the median than the average for the time until first click. This is also shown in Table 5.2.7 and in Figure 5.2.6. The average and median values were 11.5 and 9 seconds for the Company 1 site vs. 15 and 14.5 seconds respectively for the Company 2 site.

It was possible for a participant to click the first time in 3 seconds on Company 1's website which is half the time it took the quickest user on Company 2's website. This may be due to the popularity of the Company 1 rewards program. On the home page, the sections that relate to the rewards programs on both sites are indicated by the name of the rewards program. This consistency in naming may be the reason for the quicker first clicks on the Company 1 site.

Table 5.2.7: Time until first click.

Company	Average	Median	Minimum	Maximum
1	11,5	9	3	26
2	15	14,5	6	26

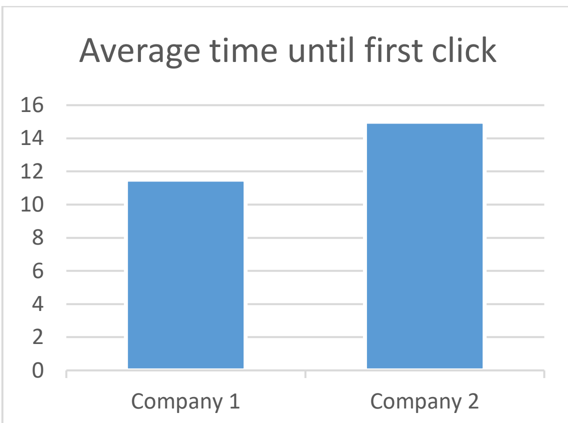


Figure 5.2.5: Average time until first click.

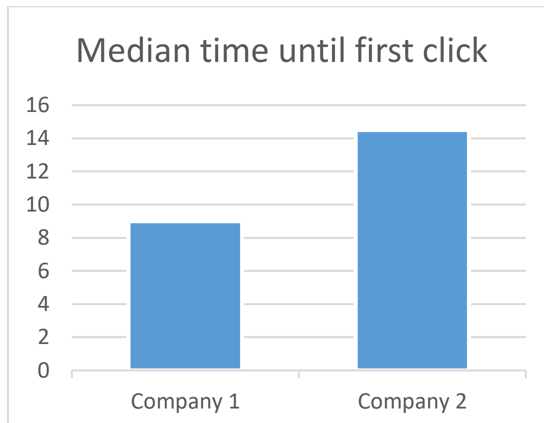


Figure 5.2.6: Median time until first click.

5.2.2.4 Fixations until first click

The data from the number of fixations until first click correlates to the time until first click. A similar trend could be noticed from the average, median and minimum values. Regarding the maximum value, there was a slight difference in values unlike the time until first click, but this is insignificant as it is participant dependent. The fixations until first click data is shown in table 5.2.8. The median and average values are depicted in Figure 5.2.7 and 5.2.8 respectively.

The average value for Company 1 was 42.9, which is 10.4 seconds less than Company 2's value of 53.3. The story is similar for the median values where the Company 1 value is 34.5, which is 13 less than the Company 2 value of 47.5.

Table 5.2.8: Fixations until first click.

Company	Average	Median	Minimum	Maximum
1	42,9	34,5	17	93
2	53,3	47,5	24	89

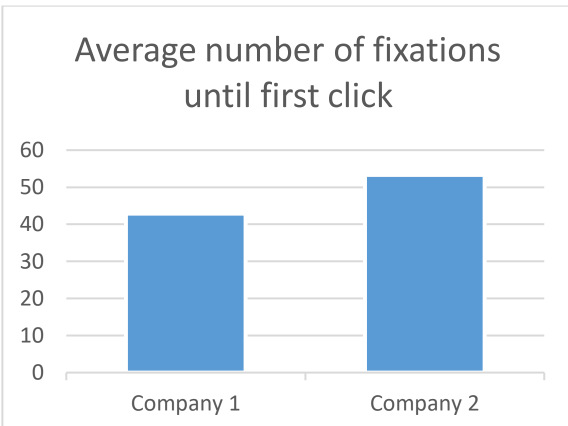


Figure 5.2.7: Average number of fixations until first click.

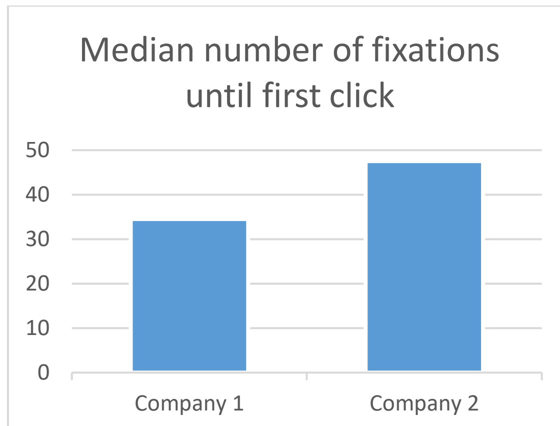


Figure 5.2.8: Median number of fixations until first click.

5.2.2.5 Number of pages visited

The average, median, minimum and maximum values for the number of pages visited are shown in Table 5.2.9. The average values are also demonstrated in Figure 5.2.9 and the median values in Figure 5.2.10. For both companies, it was possible for a participant to complete the task by visiting three pages. However, the number of pages visited by the average participant for Company 1 was 9.2 which is more than triple the optimal number. For Company 2 it was 5.7 which is slightly less than double the optimal number of pages.

The average and median values for Company 1 are larger than for Company 2 (9.2 vs. 5.7 and 9 vs. 5 respectively). This correlates with the results relating to time to complete the task. An investigation should also be done into what the influence of the quick time until first click on the home page is on the average number of pages visited. For example, could it be that the quick judgement by the participants sent them on a wrong path in the attempt to find the information? This may indicate a problem in the information architecture.

Table 5.2.9: Number of pages visited.

Company	Average	Median
1	9,2	9
2	5,7	5

Figure 5.2.9: Average number of pages visited.

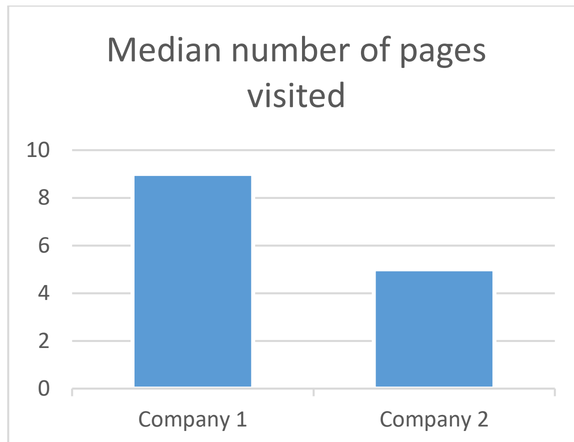


Figure 5.2.10: Median number of pages visited.

5.2.2.6 Optimal Completion

The optimal completion relates to the fact that different sources of information could be found regarding the task given to users. There are sections in both the sites that contain partial sets of information. To determine optimal completion, the full set of information in this instance was regarded as the page with the descriptions of all the rewards that could be gained if you joined the rewards program for a company. The evaluation task ended when the participant felt that they found what they were looking for. Optimal completion is a measure of how many participants found the complete set of information before deciding that they were satisfied.

The data relating to optimal completion is shown in Table 5.2.10 and the average values are depicted in Figure 5.2.11. Thirty percent (30%) of the participants accessed the full set of data before deciding that they were satisfied. On the contrary, all the participants from Company 2 decided they were satisfied when they accessed the full data set. This may have to do with the information architecture of the Company 1 site. There may be a lack of indications that more complete information is available elsewhere.

Table 5.2.10: Data relating to optimal completion

Company	Total	Percentage
1	3	30
2	10	100

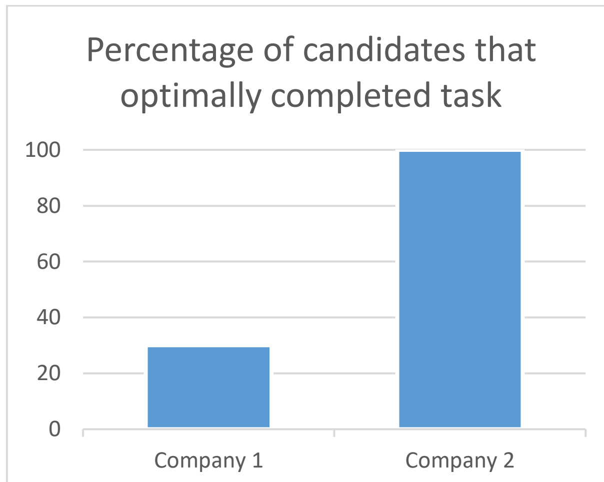


Figure 5.2.11: Percentage of candidates that optimally completed task.

5.2.3 SUGGESTED CHANGES TO THE METHODOLOGY

This section provides a refinement of the methodology to conduct CIUE based on the case described above. It was previously recommended that as much data as possible should be gathered because you can always remove data, but once the experiment is completed, data cannot be collected from the same participant again. However, to reduce unnecessary time and resources spent on the CIUE, another requirement to conduct a CIUE should be that in the planning phase, it should be stated what criteria should be met to eliminate data. For example, participants with a recorded fixation sample percentage of less than 60% as indicated by Tobii Pro Studio, were eliminated because data from those participants would not be an accurate representation of actual interaction.

Furthermore, data that is not quantitatively comparable should also be eliminated, as doubt may exist in subjective data as described in section 2.11.2 (Law & van Schaik, 2012). In the initial case, raw flow data analysis was done, but a study of flow data would be subjective to the interpreter's view of what a good path may be. To avoid spending unnecessary time and other resources, any form of subjective data should not be gathered. The subjective nature of data should also be considered in the planning phase.

The first evaluation case involved an information search. The website had portions of the information available at different areas of their website, and some participants felt they had completed the task when they had located the incomplete information. This thus brings about another requirement for the methodology. When there is more than one possible outcome of a task, the evaluator should assess which of these will qualify as successful outcomes and which not. This should then be accounted for in the evaluation plan. The indicator used in this case was 'optimal completion' which provided the percentage of participants that arrived at the full set of information before deciding that they had completed the task.

Another aspect that was not considered with the initial development of the methodology was the number of participants. The pilot case that was used to do the initial development of the methodology involved more participants (45) than the first evaluation case (20). The pilot case involved three companies and the first evaluation case involved two. Furthermore, in the pilot case, participants were not eliminated based on sample percentage and the population group was not controlled regarding previous exposure to technology. When you have one participant that is much older than the rest of the participants and has limited exposure to technology, their use of the information system may have a significant influence on the results which would make the results less useful as it is heavily based on that particular participant.

The principle to follow with regard to the number of participants should thus be that the greater the diversity of the represented user population, the more the participants that should be recruited. The danger of using a very specific population group is that the study may only be applicable to that population group. Seeing that the output of a CIUE is a comparison between different user interfaces, the variable indicators should relate to those interfaces and not to the participants that are used. The point should be made that participants will never be the same but the variability between them should be minimised if a large group of participants is not available.

Some organisations are bigger than others and some organisations do more marketing than others. As such some of their products may be more popular and more familiar to participants and the public. This was found in this first evaluation case to have an influence on the participants' use of the information systems. Exposure analysis may therefore also be beneficial to a comparative study. For example, participants that have used the computer interface before should not be used in an experiment that tests that same computer interface because they may then know exactly how to complete that task.

Sometimes because of the broad marketing of an organisation, certain terms may be more familiar to participants than terms on a competing organisation's user interface. This may make the interface easier to navigate for that participant. This does not make the data incomparable – it simply means that a marketing strategy contributed to the better usability of the information system. Participants' general exposure to the organisations should thus also be a consideration in the methodology to conduct a CIUE as it may assist in the interpretation of the results.

5.3 THE SECOND EVALUATION CASE STUDY

The second evaluation case involved three low cost aviation companies in South Africa. It included 10 participants per airline. This evaluation was conducted in a similar way as the first evaluation case. These participants were given the following scenario:

"Imagine you are a mother. Book a ticket for yourself and a baby from Durban to Cape Town leaving on the 29th of July 2017 and returning on the 3rd of August 2017. You have one bag that you would like to check in".

The task used for this case stopped the participant before payment for the ticket. Again, an attempt was made to select a homogeneous group of participants to minimise the effect of the participants' background on the results. An eye tracker was used to record the participants' use of the information systems. Participants that had a fixation sample percentage of less than 60% were eliminated.

An indicator that was included in the first evaluation case, namely, the optimal completion indicator, was removed from this case. This is because the task here required that the user book an airline ticket and there was only one possible outcome. One of the websites created a system break under certain conditions. This influenced the results and will be further examined in the analysis of the data.

Tables 5.3.1, 5.3.2 and 5.3.3 contain data relating to the participants involved in the study, one table for each airline. Only students with a high technological proficiency from a South African university were selected.

For the Airline 1 site, the ages of the participants ranged from 20 to 25 and of the 10 participants two were female and the rest were male. Furthermore, all the participants used for the Airline 1 site either attended university or had an undergraduate degree. All the participants owned a smart phone and a laptop or a PC. Two of the participants owned a tablet in addition to the above. One of the participants, who owned a tablet, also owned a video game console. Of the 10 participants, only one rated his technological proficiency as average, the rest claimed that they either had a very good or an excellent technological proficiency.

Table 5.3.1: Airline 1 participants.

Age	Gender	Educational Level	Technological Devices	Self-rated Technological Proficiency
20	Male	Attended university	Smart phone and laptop/PC	Very good
21	Male	Attended university	Smart phone, laptop/PC, video game console and tablet	Excellent
23	Male	Undergraduate degree	Smart phone and laptop/PC	Excellent
25	Male	Undergraduate degree	Smart phone and laptop/PC	Very good
22	Male	Undergraduate degree	Smart phone and laptop/PC	Very good
21	Female	Attended university	Smart phone and laptop/PC	Excellent
25	Male	Post graduate degree	Smart phone, laptop/PC and tablet	Excellent
22	Male	Attended university	Smart phone and laptop/PC	Average
21	Female	Undergraduate degree	Smart phone and laptop/PC	Very good
23	Male	Undergraduate degree	Smart phone and laptop/PC	Excellent

The ages for the Airline 2 site ranged from 19 to 26 and four of the 10 participants were female. Of the 10 participants, one was busy with a post-graduate degree, the rest were either busy with an undergraduate degree or had previously attended university. All the participants owned a smart phone and a laptop or a PC except for one who only owned a smart phone. Two of the participants also owned a tablet and one of the participants owned a video game console in addition to his smart phone, laptop or PC and tablet. Of the 10 participants, only one rated her technological proficiency as average, the rest claimed that they either had a very good or an excellent technological proficiency.

Table 5.3.2: Airline 2 participants.

Age	Gender	Educational Level	Technological Devices	Self-rated Technological Proficiency
26	Male	Undergraduate degree	Smart phone and laptop/PC	Very good
23	Male	Attended university	Smart phone, laptop/PC, video game console and tablet	Very good
19	Female	Undergraduate degree	Smart phone and laptop/PC	Very good
19	Female	Undergraduate degree	Smart phone and laptop/PC	Very good
20	Female	Attended university	Smart phone and laptop/PC	Average
21	Male	Post graduate degree	Smart phone and laptop/PC	Very good
20	Female	Attended university	Smart phone and laptop/PC	Excellent
20	Male	Undergraduate degree	Smart phone	Excellent
19	Male	Undergraduate degree	Smart phone and laptop/PC	Excellent
20	Male	Attended university	Smart phone, laptop/PC and tablet	Excellent

Five of the 10 participants selected for the evaluation of the Airline 3 site were female. The ages of the participants ranged from 19 to 23 and all of them were either busy with an undergraduate degree or had previously attended university. All the participants owned a smart phone and a laptop or a PC except for one who only owned a smart phone. Two of the participants also owned a tablet and one of the participants owned a video game console in addition to his smart phone and laptop or PC. Only two of the participants self-rated their technological proficiency as average and the rest claimed that they had either very good or excellent technological proficiency.

Table 5.3.3: Airline 3 participants.

Age	Gender	Educational Level	Technological Devices	Self-rated Technological Proficiency
23	Male	Attended university	Smart phone	Very good
20	Male	Attended university	Smart phone and laptop/PC	Very good
21	Female	Attended university	Smart phone, laptop/PC and tablet	Very good
21	Female	Attended university	Smart phone and laptop/PC	Excellent
20	Female	Attended university	Smart phone and laptop/PC	Very good
21	Male	Undergraduate degree	Smart phone, laptop/PC and video game console	Excellent
19	Female	Undergraduate degree	Smart phone and laptop/PC	Average
19	Male	Attended university	Smart phone and laptop/PC	Excellent
22	Male	Undergraduate degree	Smart phone, laptop/PC and tablet	Average
23	Female	Attended university	Smart phone and laptop/PC	Very good

5.3.1 DATA COLLECTED

For this case, the ‘optimal completion’ indicator used in the previous case was removed and the ‘number of breaks’ indicator was added. The reason for the removal of the optimal completion indicator is that there was only one possible outcome of the task, which was that the participant reached the payment page of the ticket purchasing process. The number of system breaks was included because system breaks were recorded on one of the information systems evaluated. Other differences from the previous case include that three companies were evaluated, not two and the line of business was low cost aviation, not medical insurance.

The data collected is contained in Tables 5.3.4, 5.3.5 and 5.3.6 – one for each airline included in the case. The metrics included in the tables are the participant identification number, the time to complete the given task, the number of fixations recorded while the participant completed the task, the time until the user clicked the first time, the number of fixations recorded until the user clicked the first time, the number of pages visited by the participant and the number of system breaks recorded.

Regarding the data collected from the Airline 1 site, the average time taken to complete the task was 301 seconds. The participant who finished it the quickest did it in 218 seconds and the slowest participant did it in 347 seconds. Furthermore, with regard to the number of fixations recorded, the minimum was 597 and the maximum was 1295. The average number of fixations that was recorded for the completion of the task on the Airline 1 site was 962.6.

For the time until first click, two of the participants on the Airline 1 site took one second and two of the participants took 12 seconds to click the first time. This was the minimum and maximum respectively. The average number of seconds that participants took until clicking the first time on the Airline 1 site was 4.7. The average number of fixations until clicking the first time was 19.6 for the Airline 1 site, the maximum was 50 and the minimum was five. Four of the participants visited seven pages to complete the task, three visited five pages, one visited six, one visited nine and one visited fifteen. The average number of pages visited was 7.3 for the Airline 1 site and none of the participants experienced a system break.

Concerning the time to complete the task for the Airline 2 site, the average was 240.2. Participant four completed the task in 120 seconds which was the minimum and participant five completed the task in 408 seconds which was the maximum. 495 fixations were recorded for participant one to complete the task which was the minimum and 1511 fixations were recorded for participant five which was the maximum. The average number of fixations recorded was 794.7 for the Airline 2 site.

The time until first click ranged from two to six seconds for the Airline 2 site and the average was 3.4. The number of fixations recorded until first click ranged from 6 to 20. The average number of fixations recorded until first click was 12.3. The total number of pages visited for the Airline 2 site ranged from two to four and the average was 2.5. Three of the 10 participants used for the evaluation of the Airline 2 site experienced system breaks.

None of the participants on the Airline 3 site experienced system breaks. The average time in seconds taken to complete the task on the Airline 3 site was 261.4, the minimum was 121 and the maximum was 325. The maximum number of fixations on the Airline 3 site was 1150 and the minimum was 461 and the average number of fixations recorded was 840.7.

With regard to the Airline 3 site, the minimum time it took participants to click the first time was two seconds, the maximum was eight seconds, and the average was 4.5 seconds. Six fixations were recorded until participant four clicked the first time and 34 before participant eight clicked the first time, which was the minimum and maximum respectively. The average number of fixations recorded for all the participants involved in the Airline 3 study was 17.7. All the participants on the Airline 3 site visited five pages to complete the task.

Table 5.3.4: Airline 1 data collected.

Participant No	Time to complete	Number of fixations	Time until first click	Fixations until First click	No pages visited	Number of breaks
1	218	597	3	12	5	0
2	282	966	2	7	5	0
3	291	936	11	41	5	0
4	303	1295	12	50	7	0
5	301	974	5	17	7	0
6	344	1223	4	18	15	0
7	256	800	1	5	7	0
8	322	648	4	19	6	0
9	346	1149	1	8	9	0
10	347	1038	4	19	7	0

Table 5.3.5: Airline 2 data collected.

Participant No	Time to complete	Number of fixations	Time until first click	Fixations until First click	No pages visited	Number of breaks
1	166	495	2	12	2	1
2	171	530	2	11	2	0
3	228	669	4	12	2	1
4	120	683	4	7	2	0
5	408	1511	3	12	4	0
6	287	887	5	14	2	0
7	238	723	6	20	2	0
8	309	983	3	14	4	0
9	238	752	2	6	2	0
10	237	714	3	15	3	1

Table 5.3.6: Airline 3 data collected.

Participant No	Time to complete	Number of fixations	Time until first click	Fixations until First click	No pages visited	Number of breaks
1	256	841	2	11	5	0
2	316	734	6	18	5	0
3	315	789	2	7	5	0
4	167	513	2	6	5	0
5	121	461	3	13	5	0
6	240	836	5	24	5	0
7	319	979	6	20	5	0
8	286	994	8	34	5	0
9	325	1110	4	14	5	0
10	269	1150	7	30	5	0

5.3.2 DATA ANALYSIS

This section contains a discussion around the analysis of the data collected for the second evaluation case. The data was analysed using descriptive statistics and is described in this section on an indicator by indicator basis.

Time to complete the task

The data relating to the time to complete the task is contained in Table 5.3.7. The averages are depicted in Figure 5.3.2 and the median values in Figure 5.3.3. The average participant for Airline 1 completed the task in 301 seconds, for Airline 2 in 240,4 and for Airline 3 in 261,4 seconds. The difference between the average and the median is negligible for all cases except for the median for Airline 3's site, which is 8.6 percent higher than the average. For Airline 3, it then means that there was a large difference in the values of each of the participants. There may be assorted reasons for this, one plausible explanation would be that if participants follow the optimal path on Airline 3's website, they finish quickly, but if they make a mistake or want to change it, it causes a significant difference in the time it takes to complete the task.

Participants for Airline 2's website took the shortest time to complete the task on average. It was also noticed that the Airline 2 website facilitates the entire booking on 3 pages. Since the task relating to this case was stopped before the payment process, the value for the number of pages visited for Airline 2 will always be two. Most of the data for Airline 2 was collected on the second page. This may be a reason for participants finishing quicker on that site but makes the page heavy in terms of the amount of data on that page. The Airline 2 information system also had system breaks on this page under certain conditions. The condition that was recorded was an unhandled exception where the age of the passenger was greater than the maximum age of an infant on that site but the participant selected that the passenger was an infant.

The minimum time to complete the task is almost the same for Airline 2 and Airline 3's websites but there is a 98 second difference between the minimum value on Airline 1's website and Airline 2's website. The opposite is true for the maximum values where Airline 2 scores the highest. But again, the minimum and maximum values demonstrate possibilities (it was for example possible for a participant to complete the task in 218 seconds on the Airline 1 site), and this should not be used for comparison because the influence from the individual participant is significant.

Table 5.3.7: Time to complete task.

Airline	Average	Median	Minimum	Maximum
1	301	302	218	347
2	240,2	237,5	120	408
3	261,4	286	121	325

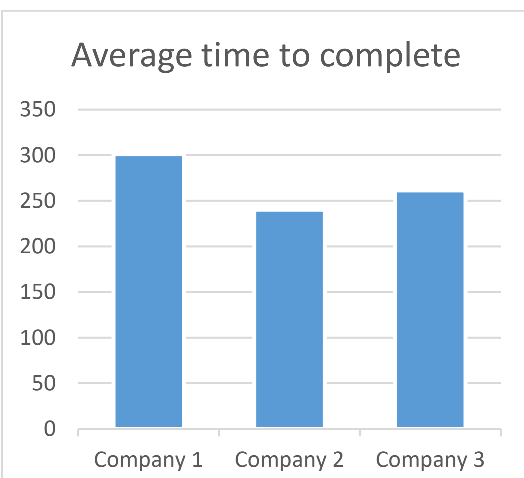


Figure 5.3.2: Average time to complete.

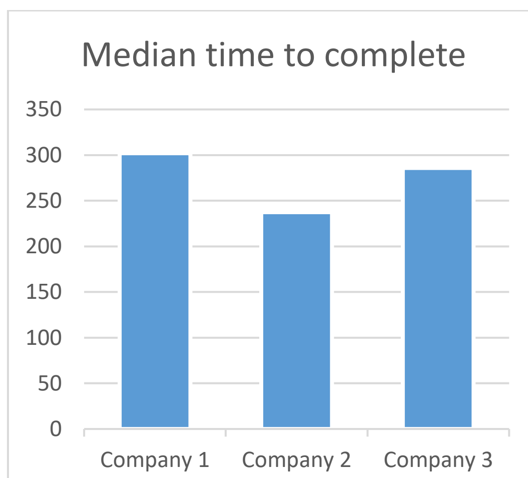


Figure 5.3.3: Median time to complete

Number of fixations to complete the task

The data relating to this section is contained in Table 5.3.8. The average number of fixations to complete the task is depicted in Figure 5.3.4 and the median values in Figure 5.3.5. The number of fixations to the completion of the task confirms the points highlighted under discussion of the time to complete the task. However, the difference between the median and average for Airline 2's site regarding fixations until first click is bigger. This may mean the following:

1. The standard deviation between participants is higher.
2. When participants dwell from the optimal path on the Airline 2 site, they need to fixate more to correct the mistake than on other sites.

This can again be explained from the second page on the Airline 2 site that collects all the data. If a participant needs to search for the place to correct the mistake, there are a lot of areas to fixate on before finding the area to fix.

The minimum value of the Airline 2 site also demonstrates that although the minimum time to complete task on the Airline 2 website is almost the same as the Airline 3 website (120 vs. 121), the difference between the minimum fixations is greater (495 vs. 461). This may be a symptom of the cluttered Airline 2 website.

Table 5.3.8: Number of fixations to complete task.

Airline	Average	Median	Minimum	Maximum
1	962,6	970	597	1295
2	794,7	718,5	495	1511
3	840,7	836	461	1150

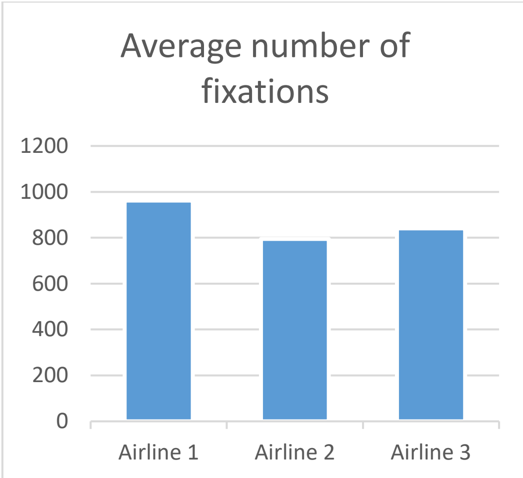


Figure 5.3.4: Average number of fixations.

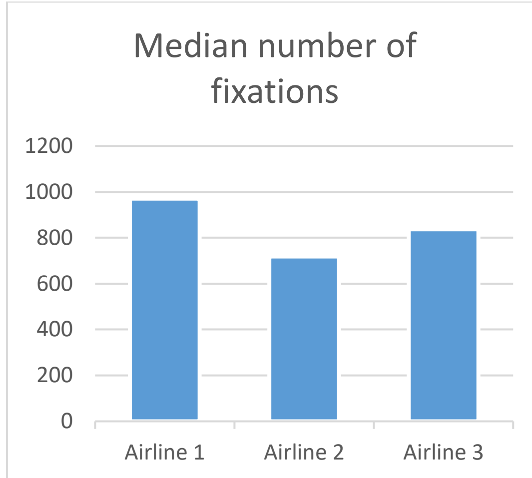


Figure 5.3.5: Median number of fixations.

Time until first click

The data relating to this section is contained in Table 5.3.9. The average number of fixations to complete the task is depicted in Figure 5.3.6 and the median values in Figure 5.3.7. Regarding the time until first click, the values are not very high for any of the websites evaluated. The average time in seconds for Airline 1 was 4.7, for Airline 2, 3.4 and for Airline 3 it was 4.5.

These values are all low, but Airline 1 and 3 can do an investigation as to why it was slightly lower for Airline 2. This may have to do with the wording or other elements of the interface on the home screen. The median values are slightly lower for each airline which may mean that more or more significant values are less than the average for each of the sites. An investigation into standard deviations may be valuable.

The only outlier regarding maximum and minimum values is the maximum value of the Airline 1 website. An investigation could be done into what caused this participant to take this long but the value may not be significant for the comparison as it relates to one specific participant.

Table 5.3.9: Time until first click.

Airline	Average	Median	Minimum	Maximum
1	4,7	4	1	12
2	3,4	3	2	6
3	4,5	4	2	8

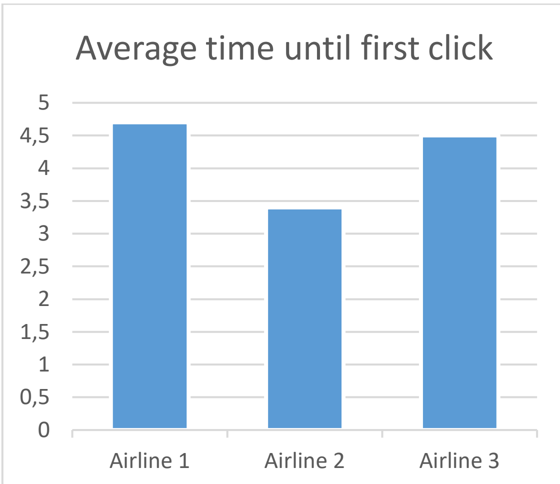


Figure 5.3.6: Average time until first click.

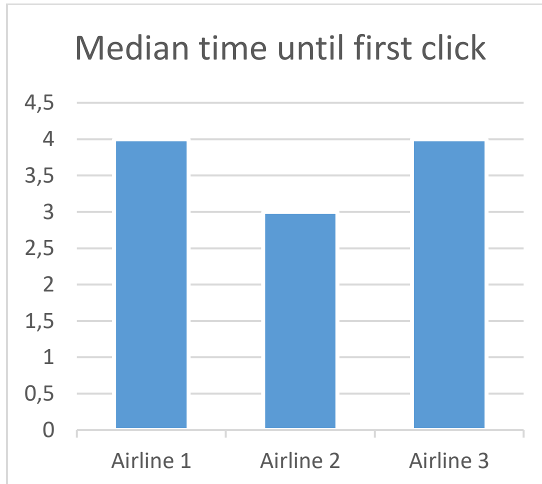


Figure 5.3.7: Median time until first click.

Number of fixations until first click

The data relating to this section is contained in Table 5.3.10. The average number of fixations to complete the task is depicted in Figure 5.3.8 and the median values in Figure 5.3.9. The number of fixations until first click confirms the discoveries from the time until first click.

Table 5.3.10: Fixations until first click.

Airline	Average	Median	Minimum	Maximum
1	19,6	17,5	5	50
2	12,3	12	6	20
3	17,7	14	6	34

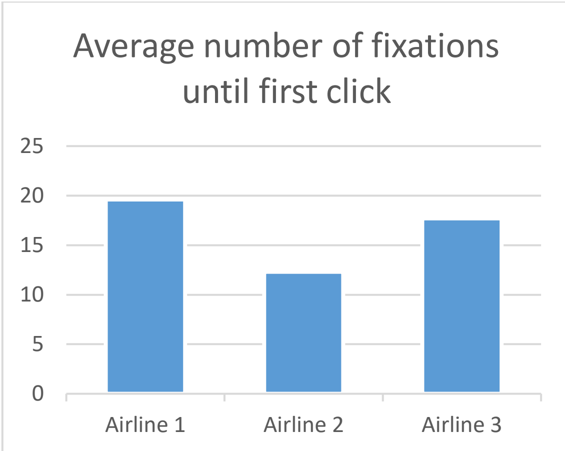


Figure 5.3.8: Average number of fixations until first click.

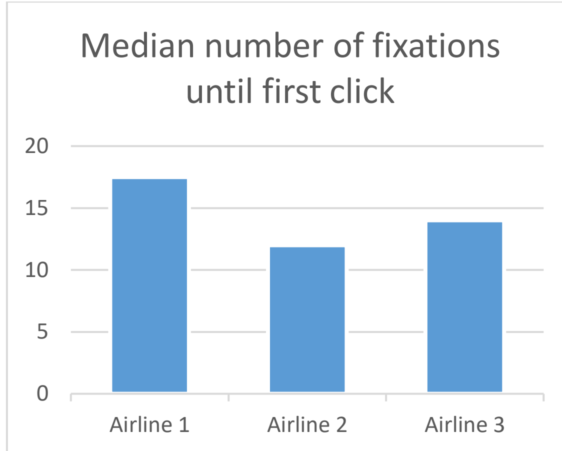


Figure 5.3.9: Median number of fixations until first click.

Number of pages visited

The data relating to this section is contained in Table 5.3.11. The average number of fixations to complete the task is depicted in Figure 5.3.10 and the median values in Figure 5.3.11. The number of pages visited is significantly lower for the Airline 2 website. This is explained by the second page of the Airline 2 website that collects all the data from the participant that is needed for the booking.

There was not a significant difference in the average and median number of pages visited for the Airline 1 site when compared to the Airline 3 sites. One of Airline 1 values was 15 which is a more than double the average for Airline 1. This shows the possibility that users may find it challenging to find their way back on the Airline 1 site once they have deviated from the optimal path.

The values for the number of pages visited on the Airline 3 website were all the same. The average and the median is thus also the same. The optimal route on the Airline 3 website contains 5 pages and the participants all followed the optimal route. This may indicate a good information architecture on the Airline 3 website.

Table 5.3.11: Number of pages visited.

Airline	Average	Median	Minimum	Maximum
1	7,3	7	5	15
2	2,5	2	2	4
3	5	5	5	5

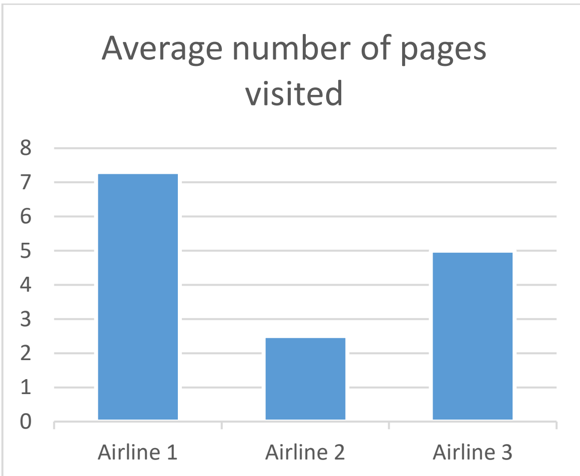


Figure 5.3.10: Average number of pages visited.

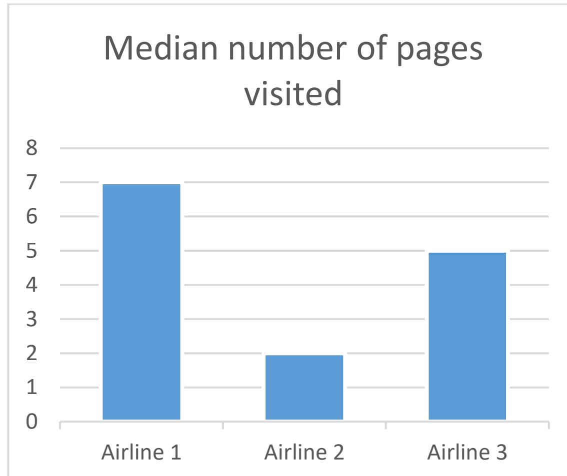


Figure 5.3.11: Median number of pages visited.

Number of system breaks

Another indicator that this case study evaluated was the number of system breaks. The Airline 1 and 3 websites did not crash or break but the Airline 2 website had a total of 3 system breaks. This has an influence on the time to complete the task, the number of fixations to complete task and the number of pages visited in that the experiment was terminated. The breaks never occurred on the first page and therefore do not have an influence on the time until first click and the number of fixations until first click.

The question that comes to mind is, what is to be done about the values for the time to complete tasks, the number of fixations to complete task and the number of pages visited when a system break occurs? There are various possibilities like the disqualification the associated company or elimination of those specific participants that should be investigated, but the methodology to conduct CIUEs should have rules that address this.

5.3.3 SUGGESTED CHANGES TO ARTEFACT

One of the noteworthy themes of this case was a system break count. One of the candidate systems that were tested had to be terminated for three participants before the participant completed the task due to system failure. The question is how do we deal with this when a CIUE is conducted? System failure has an influence on most indicators that were measured for usability. Due to this, the rule that is suggested is that if system failure occurs during an experiment, the candidate company that was being evaluated should be eliminated from the comparison and the data collected about that candidate should be regarded as unusable. There may be alternatives, but the alternatives all influence the value of the comparison. That candidate should then be advised that if they would like the CIUE to be done, they should first fix the system breaks.

Another consideration is possible resource constraints. In the ideal situation many participants can be used for the CIUE, many tasks can be evaluated, and a lot of data can be collected, analysed and compared. However, as the amount of data, the number of participants and the number of tasks increases, the amount of resources that are required increases. This should thus also be a consideration in the planning phase of a CIUE.

Regarding the resource constraint, it can be noticed that in the first evaluation CIUE, two candidates were selected for comparison and in the second evaluation CIUE there were three selected. The selection of candidates when CIUEs are done should be based on the closest competition of an organisation. The number of candidates selected for comparison has an influence on the amount of resources required for the CIUE and the results of the CIUE. As the amount of time, participants, and other resources increase so does the cost of conducting a CIUE. This cost significantly increases with every competitor that is added to the study.

The advantage of using more candidates is, however, that it becomes easier to determine how well or poorly a candidate performed, as the results of the evaluation can be compared to more candidates. As such, the CIUE becomes more useful. Furthermore, to develop industry benchmarks for indicators such as the time to complete a task, it would be ideal to compare as many of the players in that industry as possible (Zhu, 2014). As it is a goal of the methodology to conduct CIUEs to work towards benchmarking of the usability of computer interfaces, it is recommended that as many competitors as possible be included when the resources are available.

Regarding data, an assessment should be done of which indicators are valuable for comparison and which indicators are only valuable to show possibilities. For example, maximum and minimum values will only consider one participant and as such results are drastically influenced by those participants' use of the system. They are not very valuable for comparison, but they do indicate what values are possible for that indicator on that information system for that task.

Finally, it should not be forgotten that a substantial portion of the data collected can be repackaged and reused to conduct a study into how to improve the usability of the evaluated information systems. This should be considered in the planning of the CIUE. The question should be asked, how do we do the study in a manner that the data can be reused to conduct a study into improving the usability of the evaluated information systems? To contribute to this, the user interface that was evaluated can be examined to determine why a particular result occurred when a comparison was made.

5.4 EVALUATION INTERVIEWS

As discussed in Chapter 3, interviews were conducted for the evaluation cases. The goal of the interviews was to assess the perceived usefulness of the proposed CIUE and to gather recommendations as to how the methodology may be improved by presenting it to relevant parties at organisations whose applications were evaluated. The interview questions are available in appendix D.

5.4.1 INTERVIEW REGARDING THE FIRST EVALUATION CASE

The person interviewed for the first evaluation case was the head of business analysis at the medical insurance company referred to as Company 1 in Section 5.2. The researcher presented the first evaluation CIUE to the interviewee before asking the interview questions. This presentation included the data gathered, the analysis and the results of the comparative study. In other words, the interviewee could see how the usability of the information system of her company (for which she was responsible) compared to that of a competing company. The identity of the competing company was not disclosed to the interviewee for ethical reasons.

When the interviewee was questioned about the usefulness of the CIUE presented, she responded that she found it useful and interesting. She was particularly interested in the optimal completion of tasks and said that she would take the company’s bad performance regarding that up with the digital team.

When questioned about changes that she would make to the CIUE she responded that she would like to know what the participants thought about the look and feel of the site. She could also see from the results that the competitor had a marketing head start although she believed that her company had better products.

She also said that they were re-doing the entire website and was therefore not currently interested in investing in usability studies. She did not know that there was a division within her organisation that studies the user experience and usability of information systems. She also did not know what A/B testing was so that rendered those questions irrelevant.

5.4.2 POSSIBLE CHANGES TO METHODOLOGY FROM THE INTERVIEW

Even though it was suggested that the user opinions about look and feel would have been interesting to know, the user's opinion would be subjective in nature and can thus not be used for comparison. The user's opinion may however be recorded for usability recommendations that may come from the evaluations done in a CIUE. It may be a recommendation that when a CIUE is done, the data gathered should be used for a usability report.

Another recommendation would be that timing is also considered when a CIUE is conducted. If, for example, the CIUE is conducted on an information system that is soon to be replaced, it may not be very useful.

5.4.3 INTERVIEW REGARDING THE SECOND EVALUATION CASE

Interviews aimed at assessing the usefulness of CIUEs and gathering recommendations for the improvement of the methodology to conduct CIUEs were conducted for both the evaluation cases. The person that was interviewed for the second evaluation case was the head of information technology and innovation at Company 3.

When he was questioned about the usefulness of a CIUE he stated that it made a lot of sense, since it was measured in the right way and he admired that it was kept simple. He would have liked the entire ticket booking user task to be evaluated instead of stopping before the payment process. He did recognise the limitations that were associated with this. He also stated that clutter is a big thing for them and he could see from the CIUE that the minimization of clutter would lead to better results. He would rather do marketing on less critical user journeys in an effort to optimise the usability of critical user journeys. He was also happy to see that there was still room for improvement. He recommended that the difference in business models should also be considered. For example, the marketing of rewards programs on sites will have an influence on the usability of that website.

When questioned whether he would invest into usability optimisation after seeing these results, he however said that they are currently busy with large projects that are priority and that they always apply common sense when designing information systems. He also said that he can see how an investment into usability optimisation may minimise the load on the call centre. He also said that he did not know much about A/B testing.

5.4.4 POSSIBLE CHANGES TO METHODOLOGY FROM THE INTERVIEW

From the interview conducted for the second evaluation case, it was found that another consideration that should be made in the selection of information systems to compare is the difference in business models that may influence the comparison. For example, some companies work with partners that facilitate processes on the company information system which may have an influence on the usability of an information system. Timing should also be a consideration in the planning phase of a CIUE to ensure that it fits into the organisations schedule. There may be a lack of priority on the enhancement of the usability of information systems when there are projects that are perceived to be more important.

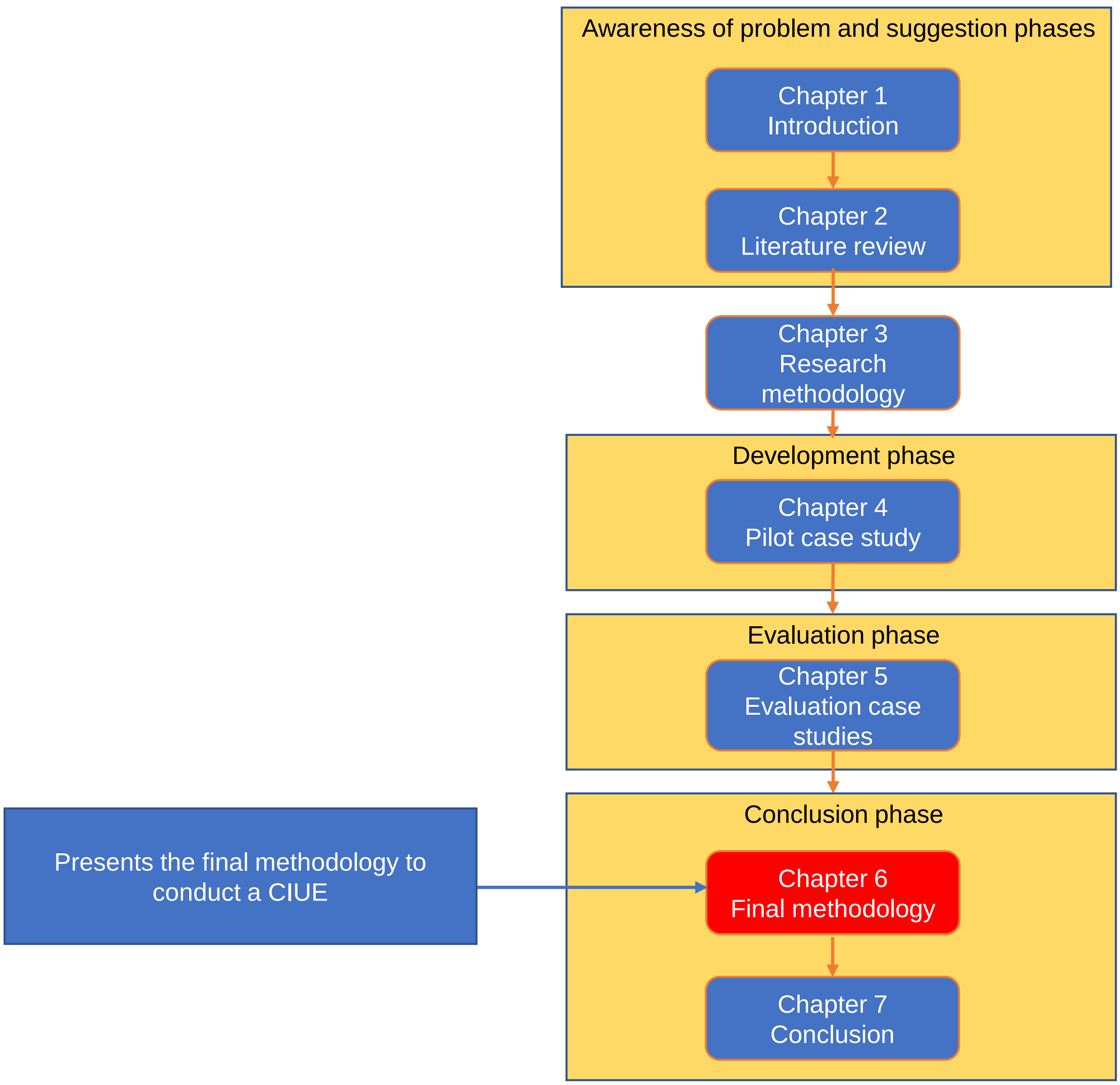
5.5 CONCLUSION

In this chapter, the methodology to conduct a CIUE was evaluated through its application in two cases and using interviews. One of the cases involved two medical insurance companies and the other involved three low cost aviation companies in South Africa. The CIUEs that were produced in the cases were also presented at interviews with relevant parties of companies evaluated.

After the cases and the interviews were conducted, changes that should be made to the methodology to conduct CIUEs were presented. The parties that were interviewed also found the CIUEs useful and learned from it. The final methodology to conduct CIUEs with the recommendations from this chapter will be presented in Chapter 6.

CHAPTER 6: FINAL METHODOLOGY TO CONDUCT A CIUE, THE MAIN RESEARCH CONTRIBUTION

The position of chapter 6 in the dissertation



6.1 INTRODUCTION

The main contribution of the research study outlined in this dissertation is a methodology to conduct Comparative Inter-organisational Usability Evaluations (CIUEs). This methodology was suggested in Chapter 2, developed with the use of a pilot case study in Chapter 4 and evaluated in Chapter 5 with the use of two evaluation cases and two interviews.

This chapter summarises the findings regarding the methodology by presenting the final methodology to conduct CIUEs. This is done by discussing the philosophy that is to be adopted, the approach to be followed, the time horizons and frames to be considered and the steps and procedures that are to be followed when conducting CIUEs. The goal of this chapter is to present the methodology to conduct CIUEs in a manner that it can be replicated. Considering this, there will be repetition from discussions in previous chapters.

6.2 PHILOSOPHY

The philosophy of User Centred Design (UCD) is regarded as a design process in which the end users influence how an information system is designed (Abras et al., 2004). To follow user centred design, the user of the design should be placed at the centre of the design process (Norman, 1988). UCD is the philosophy that underlies a CIUE, as actual users, or representatives of actual users, are involved in the usability evaluation of information systems when a CIUEs is done.

Furthermore, what makes UCD philosophy unique is that it tries to optimise the experience that a user will have when using a product by involving the user in the design process rather than expecting the user to change their behaviour to accommodate the product (Norman, 1988). This is done by considering how a user wants or needs to use the product by involving users in the design process. The ultimate goal of a CIUE is to improve the usability of an information system by showing that the information system's usability is not on the same standard as that of a competing information system.

The comparison done in CIUEs may serve as a motive for the improvement of the usability of information systems and for the further application of UCD (Venturi et al., 2006).

6.3 APPROACH

Approaches to research are applicable in CIUEs because the manner in which data is collected and analysed in CIUEs are also relevant in research. Quantitative and qualitative methods could be used in data collection and analysis processes (Bryman & Bell, 2015). Some examples of approaches that could be used in research are case studies, observation, experiments, surveys, grounded theory, ethnography or action research.

The approach that is recommended for CIUEs is controlled user observation. The observation is controlled to minimise the influence of the environment on the experiment. The variability in the data collected from the experiment should be focused on the differences between the interfaces compared in the experiment.

A further goal of a methodology to conduct CIUEs should be that it is done in an objective manner as doubt may exist if it were done subjectively. Expert recommendations are for example doubted due to the perceived subjective nature thereof (Law & van Schaik, 2012). It is further recommended that a quantitative approach is followed in the collection and analysis of the usability data collected in a CIUE. This quantitative approach will have the epistemological assumption that an objective reality exists in which the usability of information systems can be measured and analysed deductively (Bryman & Bell, 2015).

CIUEs will thus use the method of controlled user observation and will be done in a quantitative manner. In quantitative research, observations are made and data is analysed in a manner that makes deductions from the data collected from the observations. The inverse of deductive reasoning is inductive reasoning where general laws are inferred from instances of occurrence (Bryman & Bell, 2015). Inductive reasoning is typically associated with qualitative research methods which are not recommended for CIUEs.

6.4 TIME HORIZONS AND FRAMES

This research study only evaluated the use of cross-sectional CIUEs by conducting CIUEs at a particular time and not measuring trends over time. This is therefore the only recommendation that this study can make. However, it may be useful to conduct future research into longitudinal CIUEs as this could be used to examine the influence of a design change in an information system.

Furthermore, a CIUE should be done in a time that is suitable to the organisations involved. If the relevant piece of software is going to be replaced soon a CIUE will not be very useful. It may also not be useful to conduct a CIUE in a time where other projects are prioritised.

6.5 PRINCIPLES, STEPS AND PROCEDURES

This section discusses the steps and procedures that are to be followed when the methodology that is suggested by this dissertation is applied to conduct CIUEs. When following the steps, certain principles apply that are discussed in Section 6.5.1. The steps and procedures to conduct CIUEs are presented in Section 6.5.2.

6.5.1 PRINCIPLES FOR CONDUCTING CIUES

In addition to the principles of UCD discussed above, the following principles specific to CIUEs have been identified:

1. Information systems evaluated in a CIUE should not contain any system breaks, since any break in the system will influence the user's interaction with the system and the value of many metrics will be rendered meaningless for comparison purposes.
2. The environment where the CIUE is conducted should be as controlled as possible to minimise the influence it has on the outcome of the evaluation. It is the information systems being compared, not the influences of the environment on them.
3. It is important to gather data that will confirm other data so triangulation of data may be useful in this regard, as it will improve the reliability of the results.
4. The tasks that are to be compared should be similar in terms of the outcome for the user and in the goals that the organisations have with the task.
5. The number of participants should be the same for each organisation being evaluated.
6. The less homogenous the group of participants is, the larger the number of participants that should be used. This is an attempt to minimise the influence that characteristics of the participants have on the comparison.
7. The evaluations should be done in the same way for all the participants. The procedure followed, the evaluation environment and the tasks that participants are requested to complete should be the same for all participants.

8. The results of the CIUE should be presented in a manner that is easily understandable and unambiguous. This will make the results useful for more parties as a complex understanding of the subject will not be required.

6.5.2 STEPS AND PROCEDURES WHEN CONDUCTING CIUES

This section discusses the steps and procedures to be followed when a CIUE is conducted. These steps are planning, setting up the environment, grouping of participants, conducting the usability evaluations, exporting and importing data, performing the analysis and presenting results. This CIUE process is depicted in Figure 6.5.1.

1. In the planning of a CIUE, consider the following:

   a. The timing of the CIUE: If the lifetime of the product that is being evaluated is not long it may not be useful to conduct the CIUE. Furthermore, if the organisation for which the evaluation is being done, is busy with projects that have a higher priority than the CIUE, the CIUE may not gain much interest.

   b. The data that should be collected: It should be determined which data will be collected; the cost of data collection and data analysis should also be considered with this selection. Furthermore, it should be considered how control data will be collected as well as what the criteria that will result in the elimination of data will be. Finally, it should be noted that if the outcome of the tasks that the participant is to complete differs in quality, this quality may be a good indicator to evaluate.

   c. With regard to the data that should be collected, it should be decided which candidate organisations are going to be involved in the study. The considerations that should be made here include that the amount of resources required for a CIUE increases with the number of candidates that are selected and that the CIUE becomes more useful when more candidates are selected. This is because the performance of a candidate becomes clearer as the number of candidates selected for the study increases. Furthermore, if the data is going to be used to determine benchmarks in the future, the number of candidates used should be maximised to increase the reliability of the benchmarks (Zhu, 2014).

d. The participants that are to be used for the study should be recruited. The number of participants depends on the diversity of the participants. If a group of similar participants are selected then the number of participants that can be used for the study is lower. If the group of participants selected is diverse then the number of participants that should be selected is bigger. The advantage of a more diverse group of participants is that the results of the study are valid for a larger part of the total population of possible users.

e. The location of the controlled environment usability evaluation required for the CIUE should be selected. It is recommended that the location should be easily accessible for the participants and easy to control. Typically, a usability laboratory should be used for this. The environment should be controlled in terms of light, sounds, smell, available hardware, setup of tools, cleanliness et cetera.

f. The selection of tasks to be completed for each of the information systems should be made. The tasks that are selected should have the same outcome for the user of that task. For example, if a participant is going to buy an airline ticket on one of the information systems then the same should be done on the rest of the information systems involved in the CIUE. The organisation that is requesting the CIUE to be done should decide which tasks should be evaluated.

g. It should be considered how the data gathered for the CIUE could be used to create usability reports for the organisations that were involved in the CIUE. There is always a large amount of usability data gathered from a CIUE, which may be useful for the evaluating of usability to improve the information systems.

h. The environment where the CIUE will be conducted and the recording devices that will be used should be set up.

2. After the planning, group the participants for each information system that will be evaluated. Ensure that participants are not requested to complete tasks on information systems that they have used before.

3. Conduct the usability evaluations for each information system. Ensure that the data is organised to avoid confusing the results of the CIUE.

4. Extract the data from the recording tools and import the data into the analysis tools. Again, it should be ensured that the data is kept organised to avoid confusing the results of the CIUE. For example, separate spreadsheets can be created for each organisation's evaluation. Another spreadsheet can be made to compare all the data. It is also advised that this data is stored in the cloud and backups are made to avoid data loss.

5. Perform the analysis of the data by comparing similar indicators. It would be interesting to evaluate why certain results occurred, but this analysis should not influence the results of the CIUE as it would be subjective. When the investigation is done into why certain results occurred it may be useful to analyse the popularity of the product analysed or the popularity of products associated with the information system that was evaluated.

6. Present the outcomes to the parties that requested the CIUE to be done. This can be done using reports, presentations et cetera.

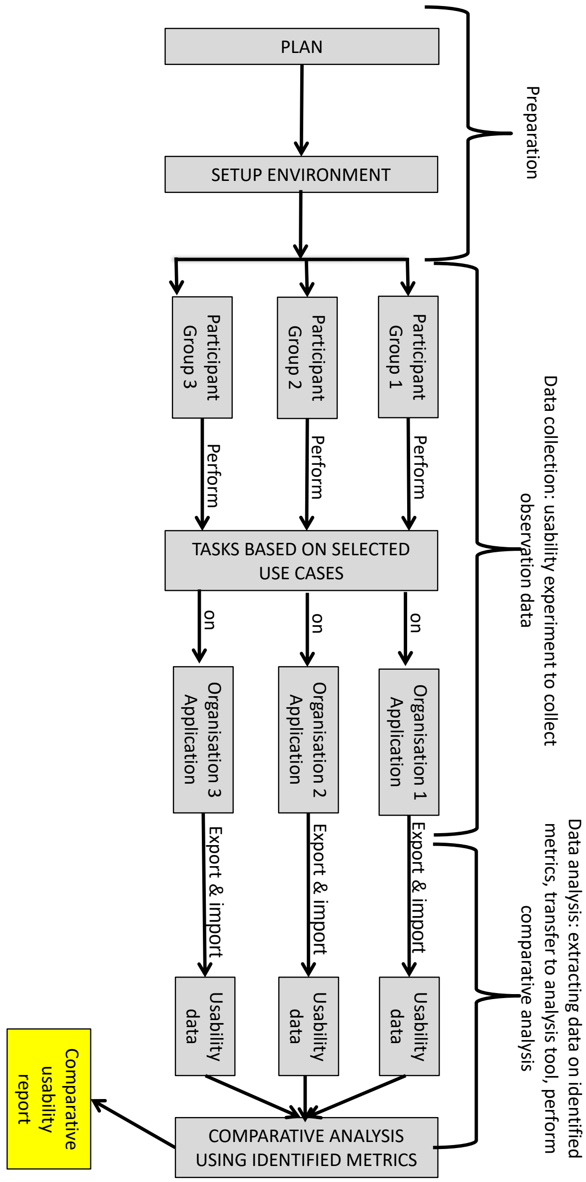


Figure 6.5.1: CIUE steps and procedures.

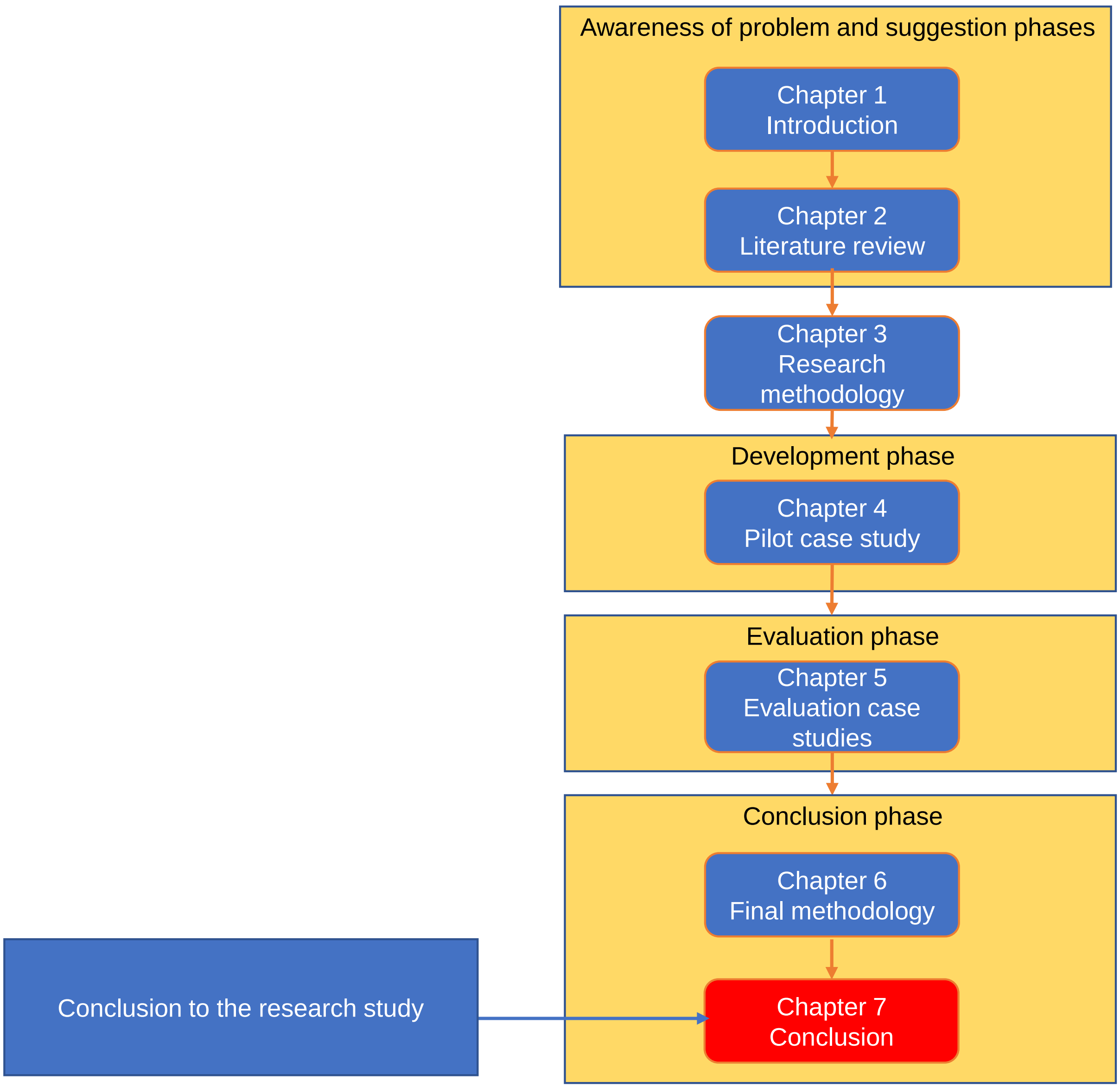
6.6 CONCLUSION

In this chapter, the final methodology to conduct CIUEs was presented. This is the main contribution of this dissertation. The UCD philosophy was discussed as being appropriate when conducting CIUEs as this philosophy puts the user at the centre of the design process. The recommended approach to be followed was a deductive approach to improve the perceived objectivity of a CIUE. It was also found, regarding time horizons, that future research could be done to investigate longitudinal CIUEs but the research currently done only investigated cross-sectional CIUEs.

It was also recommended that the environment be well controlled when a CIUE is conducted to limit the influences of factors other than the information systems on the results. Furthermore, the steps and procedures to a CIUE were suggested to be planning, setting up the environment, grouping participants, conducting the usability evaluations, exporting and importing data, performing analysis and presenting the results.

CHAPTER 7: CONCLUSION

The position of chapter 7 in the dissertation



7.1 INTRODUCTION

This chapter concludes this research dissertation by summarising the research findings, reflecting on the methodology that was used in this study, summarising the contributions made by this study and mentioning potential future research that was identified from this study.

The main contribution made by this study is a methodology to conduct Comparative Inter-organisational Usability Evaluations (CIUEs). This was done by applying Design Science Research (DSR) using a multiple-case study to develop and evaluate a methodology to conduct CIUEs. Experimental usability studies including participant interviews conducted in a usability lab equipped with eye tracking equipment as well as interviews with stakeholders were used for data collection. In the development of the methodology to conduct CIUEs, various secondary contributions were also made, especially to the organisations involved in the case studies.

7.2 SUMMARY OF RESEARCH FINDINGS

The main research question of this study was:

How can inter-organisational comparative usability evaluations be compiled into a useful methodology?

To answer this question, a methodology to conduct CIUEs was compiled using Design Science Research (DSR). It was shown to be useful through its application in various case studies, the presentation of a CIUE at a CONF-IRM conference in Cape Town in 2016 (Kruger et al., 2016), the paper that was presented is available in appendix G, and through interviews with potential users of a CIUE. The outcome was conceptualised by

- suggesting a methodology in Chapter 2,
- developing it further in Chapter 4 using a pilot case study,
- evaluating and refining it in Chapter 5 using two case studies and two semi-structured interviews, and
- presenting a final methodology in Chapter 6

The elements of the methodology to conduct CIUEs that were described in Chapter 6 were the philosophy, approach, time frames, steps and procedures. These elements are summarised in table 7.2.1.

Table 7.2.1 Elements of the CIUE methodology

Element of methodology	CIUE element selection
Philosophy	User Centred Design (UCD)
Approach	Deductive
Time frames and horizons	Cross-sectional when priority for organisation.
Steps and procedures	1. Planning 2. Setting up the environment 3. Grouping participants 4. Conducting the usability evaluations 5. Exporting and importing data 7. Performing analysis 8. Presenting the results

The UCD philosophy was discussed as being appropriate when conducting CIUEs as this philosophy puts the user at the centre of the design process.

The approach to be followed when conducting a CIUE was determined to be a deductive approach to improve the perceived objectivity of a CIUE. It was recommended that the evaluation environment be controlled when a CIUE is conducted to limit the influences of factors other than the information systems being compared on the results.

The steps and procedures to a CIUE were suggested to be planning, setting up the environment, grouping participants, conducting the usability evaluations, exporting and importing data, performing analysis and presenting the results. For an elaboration of this main contribution refer to Chapter 6.

The secondary research questions of this study were:

1. How can the information systems at organisations be compared in a manner that may convince organisations to invest into usability optimisation?

It was shown that the adoption of UCD can be encouraged by management at organisations ensuring that UCD is part of business strategy within organisations and supported at a higher level. This can be achieved by encouraging that competitive analysis is done on the usability of software products and by the outcomes of UCD phases being communicated effectively to relevant parties at organisations (Venturi et al., 2006).

It was also shown that when UCD is adopted, there is a strong emphasis on the usability of systems (Norman, 1988). A CIUE allows competitive analysis of the usability of software products. Therefore, it may encourage the adoption of the UCD philosophy within organisations, which in turn may convince organisations to invest in the usability optimisation of software products as usability is emphasized within UCD.

Furthermore, in Chapter 2, it was concluded that CIUEs should be done in an objective manner as the users of a CIUE may have greater doubt in subjective analysis (Law & van Schaik, 2012). Therefore, it is recommended that a quantitative approach is followed in the collection and analysis of the usability data collected in a CIUE.

2. Which indicators of usability are objectively comparable?

A strong emphasis was placed on the objective measurement of the usability of information system products in this study. This was done to minimise the scepticism that may occur in subjective measurement. The recommended approach to be followed when conducting a

CIUE is a deductive one. As such, indicators that could be measured without subjective influence are recommended for comparison. A list of objectively comparable indicators is available in Appendix A. This is the list of indicators that was used in this study. Future research can be done to enhance this list.

3. Are CIUEs useful in the selection of software products?

Jadhav and Sonar (2009) state that there is a need for the “framework comprising of software selection methodology, evaluation technique, evaluation criteria, and a system to assist decision makers in software selection”. A CIUE is an evaluation technique of software and is used for comparison. Theoretically, it may therefore be useful in the selection of software products. However, no research was done to validate the appropriateness of the use of CIUEs in the selection of software products in this dissertation as the emphasis of this research was on the comparison of the usability of systems at competing organisations.

Furthermore, in the interviews conducted with relevant parties at organisations involved in the evaluation CIUEs presented in Chapter 5, it was found that these parties found the CIUEs useful. Comments were made that the simplicity of the methodology to conduct CIUEs contributes to the practical feasibility of CIUEs and as such also the usefulness thereof. Future research may be done into whether CIUEs can be used in the selection of software products.

7.3 METHODOLOGICAL REFLECTIONS

The DSR methodology was applied in this research study. This was done by suggesting the methodology to conduct CIUEs from literature, developing the methodology to conduct CIUEs from a pilot case study and evaluating the methodology by applying it in two more cases and conducting interviews with parties that may use the CIUEs within their organisations.

DSR was useful to produce the final methodology to conduct CIUEs. However, expert consultation would have been useful to refine the methodology even further. The parties that were interviewed had a limited knowledge in the appropriate subject field even though they work to design information systems daily.

7.4 CONTRIBUTIONS

Through the application of the DSR methodology various practical and theoretical contributions were made. The main contribution of this research dissertation was a methodology to conduct CIUEs; this contribution is described in Chapter 6. The suggested methodology included the suggestion that the UCD philosophy be adopted in the conducting of a CIUE, the approach that should be followed in the conducting of a CIUE (which should be deductive to encourage the perceived objectivity of the CIUE) as well as various steps and procedures as to how a CIUE should be conducted.

Furthermore, the three CIUEs that were conducted in this research could be used by the eight organisations to encourage the adoption of UCD, to work towards the benchmarking of the usability of the information systems or to encourage the investment in improving the usability of the evaluated organisations information systems.

Several attempts to come up with a model to benchmark user interfaces and the usability thereof have been made (Presley & Fellows, 2013; Rohrer et al., 2016). These methods often depend on expert recommendation and the qualitative judgement of a range of usability indicators. However, the major drawback of expert recommendation is that it is often doubted (Law & van Schaik, 2012). A manner to objectively evaluate usability, like the benchmarking of usability, may be more suited than expert recommendation. A deductive approach is recommended for use in a CIUE and as such may contribute to the objective benchmarking of the usability of software products.

Additionally, in the second CIUE case, an indicator for optimal completion was used, since the outcome of the task that participants were requested to complete could be completed in ways that offered different results. This was a simple indicator that counted the number of participants that completed the task in the most optimal way. This indicator was not found in research and as such may be regarded as another theoretical contribution of the research described in this dissertation. A further list of indicators was suggested that can be used for CIUEs or other comparative evaluations of usability such as in the selection of third party software products.

A more practical contribution is the usability evaluation data of the information systems involved in this research study. The data could be analysed to give recommendations as to how these organisations can improve the usability of the information systems that were evaluated.

7.5 FUTURE RESEARCH

During this research study, various ways to use CIUEs were identified that may be elaborated using research. Firstly, research can be done into how a longitudinal CIUE can be conducted. This would entail doing the comparisons of the usability of systems at competing organisations over time. This could, for example, be used to measure the effectiveness of a usability adaption.

This research study emphasised the comparison of the usability of information systems at competing organisations. Research into how usability can be compared for different uses such as the selection of software products may also be useful. Furthermore, in the reflection of the research methodology that was used, it was found that more expert consultation may have been useful to refine the methodology to conduct CIUEs more. Research to refine the methodology to conduct CIUEs though expert consultation may thus also be useful.

It was also found in the literature review done for this study that the benchmarking of usability may be useful. However, no methodology to benchmark usability was found. A research study regarding the development of usability benchmarks using a similar approach as the one followed in this research study may also be useful.

7.6 CONCLUSION

This research study was concluded in this chapter. A methodology to conduct CIUEs was developed in this research study. DSR was applied to develop and evaluate the methodology to conduct CIUEs. This was done using a multiple-case study and interviews. It was further shown that a CIUE may be useful to encourage the adoption of UCD in organisations, to develop benchmarks and may also be useful in the selection of software products.

It was also found that it may be useful to study longitudinal CIUEs, validate the application of CIUEs in the selection of software products, research the application of CIUEs in the benchmarking of the usability of software products, research into finding more comparable indicators of usability and find other scenarios where CIUEs may be useful.

REFERENCES

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications, 37*(4), 445-456.

Acocella, I. (2012). The focus groups in social research: advantages and disadvantages. *Quality & Quantity, 46*(4), 1125-1136.

Alghamdi, A. S., Al-Badi, A., Alroobaea, R., & Mayhew, P. (2013). A comparative study of synchronous and asynchronous remote usability testing methods. *International Review of Basic and Applied Sciences, 1*(3).

AlRoobaea, R., & Mayhew, P. J. (2014). *How many participants are really enough for usability studies?* Paper presented at the Science and Information Conference (SAI), 2014.

Armstrong, G., Kotler, P., Harker, M., & Brennan, R. (2015). *Marketing: an introduction*: Pearson Education.

Arnott, D., Lizama, F., & Song, Y. (2017). Patterns of business intelligence systems use in organizations. *Decision support systems, 97*, 58-68.

Avison, D., & Fitzgerald, G. (2003). *Information systems development: methodologies, techniques and tools*: McGraw Hill.

Avison, D. E., & Fitzgerald, G. (1999). Information systems development. *Rethinking management information systems*, 250-278.

Ayala, C., Hauge, Ø., Conradi, R., Franch, X., & Li, J. (2011). Selection of third party software in Off-The-Shelf-based software development—An interview study with industrial practitioners. *Journal of Systems and Software, 84*(4), 620-637.

Barnum, C. M. (2010). *Usability testing essentials: ready, set... test!* : Elsevier.

Bastien, J. C. (2010). Usability testing: a review of some methodological and technical aspects of the method. *International journal of medical informatics, 79*(4), e18-e23.

Beck, E., Christiansen, M., Kjeldskov, J., Kolbe, N., & Stage, J. (2003). Experimental evaluation of techniques for usability testing of mobile systems in a laboratory setting.

Beukes, W., Gelderblom, H., & van der Merwe, A. (2016). Why designers responsible for websites of large organisations disregard basic web design principles.

Bevan, N. (2009a). Extending quality in use to provide a framework for usability measurement. *Human Centered Design*, 13-22.

Bevan, N. (2009b). Usability. In *Encyclopedia of Database Systems* (pp. 3247-3251): Springer.

Bhattacherjee, A. (2012). Social science research: principles, methods, and practices.

Brown, I., & Molla, A. (2015). Determinants of Internet and cell phone banking adoption in South Africa. *The Journal of Internet Banking and Commerce, 2005*.

Bryman, A. (2015). *Social research methods*: Oxford university press.

Bryman, A., & Bell, E. (2015). *Business research methods*: Oxford University Press, USA.

Buenstorf, G., & Fornahl, D. (2009). B2C—bubble to cluster: the dot-com boom, spin-off entrepreneurship, and regional agglomeration. *Journal of evolutionary Economics, 19*(3), 349-378.

Carvallo, J. P., & Franch, X. (2009). *On the use of Requirements for Driving Call-for-tender Processes for procuring Coarse-grained OTS Components.* Paper presented at the Requirements Engineering Conference, 2009. RE'09. 17th IEEE International.

Checkland, P. (1981). Systems thinking, systems practice.

Chin, K., Keese, G. L., Apple, A. J., & Folkert Herlyn, I. (1998). Benchmarking information systems. *Chemical Engineering, 105*(6), 149.

Co-operation, O. f. E., & Dev. (1993). Glossary of Industrial Organisation Economics and Competition Law. In.

Coad, P., & Yourdon, E. (1991). Object oriented analysis.

Cragg, P. B. (2002). Benchmarking information technology practices in small firms. *European Journal of Information Systems, 11*(4), 267-282.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Mis Quarterly*, 319-340.

Davis Jr, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results.* Massachusetts Institute of Technology,

de Queiroz Pierre, R. d. S. (2015). Heuristics in Design: A Literature Review. *Procedia Manufacturing, 3*, 6571-6578.

Dix, A. (2004). *Human-computer interaction*: Pearson Education.

Djamasbi, S. (2014). Eye tracking and web experience. *AIS Transactions on Human-Computer Interaction, 6*(2), 37-54.

Djamasbi, S., Siegel, M., & Tullis, T. (2010). Generation Y, web design, and eye tracking. *International Journal of Human-Computer Studies, 68*(5), 307-323.

Donald, A., & Draper, S. W. N. (1986). User Centered System Design: New perspectives on human-computer interaction.

Drnevich, P. L., & Croson, D. C. (2013). Information technology and business-level strategy: Toward an integrated theoretical perspective. *Mis Quarterly, 37*(2).

Dumas, J. S., & Redish, J. (1999). *A practical guide to usability testing*: Intellect books.

Easterby-Smith, M., Thorpe, R., & Jackson, P. R. (2012). *Management research*: Sage.

Fenton, N., & Bieman, J. (2014). *Software metrics: a rigorous and practical approach*: CRC Press.

Francis, M. M., & Willard, M. (2016). UNLOCKING THE POTENTIAL OF INFORMATION AND COMMUNICATION TECHNOLOGY FOR BUSINESS SUSTAINABILITY BY SMALL, MICRO AND MEDIUM ENTERPRISES IN VHEMBE DISTRICT, SOUTH AFRICA. *IADIS International Journal on Computer Science & Information Systems, 11*(2).

Gane, C. P., & Sarson, T. (1979). *Structured systems analysis: tools and techniques*: Prentice Hall Professional Technical Reference.

Gelderblom, H., Adebesin, F., Brosens, J., & Kruger, R. (2017). Using Eye Tracking as a Tool to Teach Informatics Students the Importance of User Centered Design. *International Journal of Information and Communication Technology Education (IJICTE), 13*(4), 111-125.

Glackin, C. (2013). Entrepreneurship: Starting and Operating a Small Business, 3e.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *Mis Quarterly, 37*(2), 337-355.

Hasan, L., Morris, A., & Probets, S. (2013). E-commerce websites for developing countries-A usability evaluation framework. *Online Information Review, 37*(2), 231-251.

Helander, M. G. (2014). *Handbook of human-computer interaction*: Elsevier.

Hertzum, M. (2016). Usability testing: too early? too much talking? too many problems? *Journal of Usability Studies, 11*(3), 83-88.

Hevner, A., & Chatterjee, S. (2010). *Design science research in information systems*: Springer.

Hoehle, H., & Venkatesh, V. (2015). Mobile Application Usability: Conceptualization and Instrument Development. *Mis Quarterly, 39*(2).

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*: OUP Oxford.

Hovorka, D. (2009). *Design Science Research: A call for a pragmatic perspective.* Paper presented at the Proceedings of SIGPrag Workshop. Sprouts Working Papers on Information Systems.

Jadhav, A. S., & Sonar, R. M. (2009). Evaluating and selecting software packages: A review. *Information and Software Technology, 51*(3), 555-563.

Jurca, G., Hellmann, T. D., & Maurer, F. (2014). *Integrating Agile and user-centered design: a systematic mapping and review of evaluation and validation studies of Agile-UX.* Paper presented at the Agile Conference (AGILE), 2014.

Krueger, R. A., & Casey, M. A. (2014). *Focus groups: A practical guide for applied research*: Sage publications.

Kruger, R. M., Gelderblom, H., & Beukes, W. (2016). The value of comparative usability and UX evaluation for e-commerce organisations.

Kuechler, B., & Vaishnavi, V. (2008). On theory development in design science research: anatomy of a research project. *European Journal of Information Systems, 17*(5), 489-504.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International journal of information management, 34*(3), 387-394.

Laudon, K. C., & Laudon, J. P. (2013). Management Information Systems 13e.

Law, E. L.-C., & van Schaik, P. (2012). *To Measure or not to Measure UX: An Interview Study.* Paper presented at the I-UxSED.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*: Morgan Kaufmann.

Lee, S. Y., Kim, B. K., & Choi, Y. K. (2016). Organizational Factors of the Successful Adoption of User-Centered Design. *한국HCI학회 학술대회 논문집*87-89.

Li, H., & Sun, J. (2013). Predicting business failure using an RSF-based case-based reasoning ensemble forecasting method. *Journal of Forecasting, 32*(2), 180-192.

Liengme, B. (2015). *A Guide to Microsoft Excel 2013 for Scientists and Engineers*: Academic Press.

Lilholt, P. H., Jensen, M. H., & Hejlesen, O. K. (2015). Heuristic evaluation of a telehealth system from the Danish TeleCare North Trial. *International journal of medical informatics, 84*(5), 319-326.

Low, C., & Chen, Y. H. (2012). Criteria for the evaluation of a cloud-based hospital information system outsourcing provider. *Journal of medical systems, 36*(6), 3543-3553.

Lundeberg, M. (1982). *The ISAC approach to specification of information systems and its application to the organization of an IFIP working conference.* Paper presented at the Information systems design methodologies: A comparative review.

Majchrzak, A., Markus, M. L., & Wareham, J. D. (2016). Designing for digital transformation: Lessons for information systems research from the study of ICT and societal challenges. *Management Information Systems Quarterly, 40*(2), 267-277.

Mamoghli, S., Goepp, V., & Botta-Genoulaz, V. (2017). Aligning ERP systems with companies' real needs: an 'Operational Model Based'method. *Enterprise Information Systems, 11*(2), 185-222.

Mann, L. (2017). Africa's Information Revolution: technical regimes and production networks in South Africa and Tanzania by James T. Murphy and Pádraig Carmody. *Africa: The Journal of the International African Institute, 87*(2), 435-437.

Melville, N., Kraemer, K., & Gurbaxani, V. (2004). MISQ REVIEW - Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *MIS quarterly : management information systems., 28*(2), 283.

Mingers, J. (2004). Real-izing information systems: critical realism as an underpinning philosophy for information systems. *Information and organization, 14*(2), 87-103.

Molich, R., Ede, M. R., Kaasgaard, K., & Karyukin, B. (2004). Comparative usability evaluation. *Behaviour & Information Technology, 23*(1), 65-74.

Morgan, D. L. (2014). Pragmatism as a Paradigm for Social Research. *Qualitative Inquiry, 20*(8), 1045-1053.

Nielsen, J. (1995). *Usability inspection methods.* Paper presented at the Conference companion on Human factors in computing systems.

Nielsen, J. (2003). Usability 101: Introduction to usability. In.

Nielsen, J., & Gilutz, S. (2013). Return on investment (ROI) for usability. *Retrieved September*.

Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces.* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Nielsen, J., & Norman, D. (2014). The definition of user experience. nngroup. In.

Norman, D. (1988). The psychology of everyday things.(The design of everyday things). In: Basic Books.

O'Donnell, K. (2017). Local Government RIM Benchmarking Study 2016. *IQ: The RIM Quarterly, 33*(1), 27.

Ochara, N. M., & Mawela, T. (2015). Enabling social sustainability of e-participation through mobile technology. *Information Technology for Development, 21*(2), 205-228.
Offermann, P., Blom, S., Schönherr, M., & Bub, U. (2010). *Artifact types in information systems design science–a literature review.* Paper presented at the International Conference on Design Science Research in Information Systems.
Otaiza, R., Rusu, C., & Roncagliolo, S. (2010). *Evaluating the usability of transactional web sites.* Paper presented at the Advances in Computer-Human Interactions, 2010. ACHI'10. Third International Conference on.
Pallot, M., Trousse, B., Senach, B., & Scapin, D. (2010). *Living lab research landscape: From user centred design and user experience towards user cocreation.* Paper presented at the First European Summer School" Living Labs".
Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K., & Newman, J. K. (2004). *The determinants of web page viewing behavior: an eye-tracking study.* Paper presented at the Proceedings of the 2004 symposium on Eye tracking research & applications.
Patton, M. Q. (2005). *Qualitative research*: Wiley Online Library.
Pearlson, K. E., & Saunders, C. S. (2009). Strategic management of information systems. In: Wiley Hoboken, NJ.
Pearlson, K. E., Saunders, C. S., & Galletta, D. F. (2016). *Managing and Using Information Systems, Binder Ready Version: A Strategic Approach*: John Wiley & Sons.
Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45-77.
Peppers, D., & Rogers, M. (2016). *Managing Customer Experience and Relationships: A Strategic Framework*: John Wiley & Sons.
Petuskiene, E., & Glinskiene, R. (2015). Entrepreneurship as the basic element for the successful employment of benchmarking and business innovations. *Engineering Economics, 22*(1), 69-77.
Porter, M. E. (2008). *On competition*: Harvard Business Press.
Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction design : beyond human-computer interaction* (Fourth edition. ed.). Chichester, West Sussex :: John Wiley & Sons Ltd.
Presley, A., & Fellows, P. (2013). An analytic hierarchy process model for evaluating and comparing website usability. *International Journal of Business Information Systems, 12*(2), 123-139.
Pretorius, M., Hobbs, J., & Fenn, T. (2015). *The User Experience Landscape of South Africa.* Paper presented at the Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists.
Rohrer, C. P., Wendt, J., Sauro, J., Boyle, F., & Cole, S. (2016). *Practical Usability Rating by Experts (PURE): A Pragmatic Approach for Scoring Product Usability.* Paper presented at the Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.
Saunders, M. L., & Lewis, P. (2009). P. & Thornhill, A.(2009). *Research methods for business students, 4*.
Sayed, Y., Badroodien, A., Hanaya, A., & Rodríguez, D. (2017). Social Cohesion, Violence, and Education in South Africa. In *Enlarging the Scope of Peace Psychology* (pp. 239-254): Springer.
Schneidermeier, T. (2015). *Lessons Learned in Usability Consulting.* Paper presented at the International Conference of Design, User Experience, and Usability.
Schreiber, G., Wielinga, B., & Breuker, J. (1993). *KADS: A principled approach to knowledge-based system development* (Vol. 11): Academic Press.
Shamoo, A. E., & Resnik, D. B. (2009). *Responsible conduct of research*: Oxford University Press.
Shneiderman, B. (2010). *Designing the user interface: strategies for effective human-computer interaction*: Pearson Education India.
Slack, N. (2015). *Operations strategy*: Wiley Online Library.

Spaull, N. (2013). South Africa's education crisis: The quality of education in South Africa 1994-2011. *Johannesburg: Centre for Development and Enterprise*.
Srinivasan, S. S., Anderson, R., & Ponnavolu, K. (2002). Customer loyalty in e-commerce: an exploration of its antecedents and consequences. *Journal of retailing, 78*(1), 41-50.
Stair, R., & Reynolds, G. (2013). *Principles of information systems*: Cengage Learning.
Stein, M.-K., Jensen, T., & Hekkala, R. (2015). Comfortably 'Betwixt and Between'? Delimiting and Blending Space, Time, Tasks and Technology at Work.
Swinyard, W. R. (1993). The effects of mood, involvement, and quality of store experience on shopping intentions. *Journal of Consumer Research*, 271-280.
Teece, D. J. (2010). Business models, business strategy and innovation. *Long range planning, 43*(2), 172-194.
Tobii. (2016a). Tobii Pro Studio. Retrieved from http://www.tobiipro.com/product-listing/tobii-pro-studio/
Tobii. (2016b). Tobii Pro X3-120. Retrieved from http://www.tobiipro.com/product-listing/tobii-pro-x3-120/#Specifications
Van Grembergen, W., & De Haes, S. (2017). *Introduction to IT Governance and Its Mechanisms Minitrack.* Paper presented at the Proceedings of the 50th Hawaii International Conference on System Sciences.
Venkatesh, V., Hoehle, H., & Aljafari, R. (2017). A usability study of the obamacare website: Evaluation and recommendations. *Government Information Quarterly*.
Venturi, G., Troost, J., & Jokela, T. (2006). People, organizations, and processes: An inquiry into the adoption of user-centered design in industry. *International Journal of Human-Computer Interaction, 21*(2), 219-238.
Vogt, W. P., & Johnson, R. B. (2011). *Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences: A Nontechnical Guide for the Social Sciences*: Sage.
Weaver, P., Lambrou, N., & Walkley, M. (1998). *Practical SSADM Version 4+: a complete tutorial guide*: Financial Times Pitman.
Whish, R., & Bailey, D. (2015). *Competition law*: Oxford University Press, USA.
Wilkinson, C. R., & De Angeli, A. (2014). Applying user centred and participatory design approaches to commercial product development. *Design Studies, 35*(6), 614-631.
Winter, R. (2008). Design science research in Europe. *European Journal of Information Systems, 17*(5), 470.
Xu, Y., Chen, N., Fernandez, A., Sinno, O., & Bhasin, A. (2015). *From infrastructure to culture: A/b testing challenges in large scale social networks.* Paper presented at the Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
Yin, R. K. (2013). *Case study research: Design and methods*: Sage publications.
Zaimi, A., Ampatzoglou, A., Triantafyllidou, N., Chatzigeorgiou, A., Mavridis, A., Chaikalis, T., . . . Stamelos, I. (2015). *An empirical study on the reuse of third-party libraries in open-source software development.* Paper presented at the Proceedings of the 7th Balkan Conference on Informatics Conference.
Zapata, B. C., Niñirola, A. H., Idri, A., Fernández-Alemán, J. L., & Toval, A. (2014). Mobile PHRs compliance with Android and iOS usability guidelines. *Journal of medical systems, 38*(8), 81.
Zhu, J. (2014). *Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets* (Vol. 213): Springer.

APPENDICES

APPENDIX A: DESCRIPTION OF INDICATORS

Indicator	Description	Case studies used
Age	Contains the participant's age.	All
Gender	Contains the participant's gender.	All
Educational level	Contains the participant's educational level. There are 5 possibilities in the interview.	Evaluation case study 1 and evaluation case study 2.
Devices used	Contains the devices that participant typically uses. There are 4 possibilities in the interview.	Evaluation case study 1 and evaluation case study 2.
Self-rated proficiency	Contains the participants rating of their overall proficiency in terms of the devices they selected in interview question 4.	Evaluation case study 1 and evaluation case study 2.
Qualification interview question	This contains the participant's highest qualification.	Development of methodology case study.
Participant No	This is a number to identify the participant throughout the workbook and in this paper.	All
Time to complete	This is the total time that the user spent to complete the task. This time is measured from when the user presses the space bar to when the facilitator presses the escape key.	All
Number of fixations	This is the total number of fixations that the user had to complete the task. This number of fixations is measured from when the user presses the space bar to when the facilitator presses the escape key.	All

Time until first click	This is the total time that the user spent before clicking the first time. This time is measured from when the user presses the space bar to when the user clicks the first time.	All
Fixations until First click	This is the total number of fixations that the user had until he or she clicked the first time. This number of fixations is measured from when the user presses the space bar to when the participant clicks the first time.	All1
No pages visited	This is the total number of pages visited by the user to complete the task.	All
Number of breaks	This is the amount of times the system broke while the user was in the process of completing the task.	Evaluation case study 1 and evaluation case study 2.
Average time to complete task	This is the average of the amount of time that all the participants took to complete the task as measured from when the user presses the space bar and when the facilitator presses the escape key.	All
Median time to complete task	This is the median of the amount of time that all the participants took to complete the task as measured from when the user presses the space bar and when the facilitator presses the escape key.	All
Minimum time to complete task	This is the minimum of the amount of time that all the participants took to complete the task as measured from when the user presses the space	All

	bar and when the facilitator presses the escape key.	
Maximum time to complete task	This is the maximum of the amount of time that all the participants took to complete the task as measured from when the user presses the space bar and when the facilitator presses the escape key.	All
Average fixations to complete task	This is the average of the number of fixations that all the participants took to complete the task as measured from when the user presses the space bar and when the facilitator presses the escape key.	All
Median fixations to complete task	This is the median of the number of fixations that all the participants took to complete the task as measured from when the user presses the space bar and when the facilitator presses the escape key.	All
Minimum fixations to complete task	This is the minimum of the number of fixations that all the participants took to complete the task as measured from when the user presses the space bar and when the facilitator presses the escape key.	All
Maximum fixations to complete task	This is the maximum of the number of fixations that all the participants took to complete the task as measured from when the user presses the space bar and when the facilitator presses the escape key.	All

Average time until first click	This is the average time that the user spent before clicking the first time on each information system. This is the average of the amount of time that is measured from when the user presses the space bar to when the user clicks the first time.	All
Median time until first click	This is the median time that the user spent before clicking the first time on each information system. This is the median of the amount of time that is measured from when the user presses the space bar to when the user clicks the first time.	All
Minimum time until first click	This is the minimum time that the user spent before clicking the first time on each information system. This is the minimum of the amount of time that is measured from when the user presses the space bar to when the user clicks the first time.	All
Maximum time until first click	This is the maximum time that the user spent before clicking the first time on each information system. This is the maximum of the amount of time that is measured from when the user presses the space bar to when the user clicks the first time.	All
Average fixations until first click	This is the average number of fixations that were measured before the user clicked the first time. This is the average of the number of fixations that is measured from when the user presses	All

	the space bar to when the user clicks the first time.	
Median fixations until first click	This is the median number of fixations that were measured before the user clicked the first time. This is the median of the number of fixations that is measured from when the user presses the space bar to when the user clicks the first time.	All
Minimum fixations until first click	This is the minimum number of fixations that were measured before the user clicked the first time. This is the minimum of the number of fixations that is measured from when the user presses the space bar to when the user clicks the first time.	All
Maximum fixations until first click	This is the maximum number of fixations that were measured before the user clicked the first time. This is the maximum of the number of fixations that is measured from when the user presses the space bar to when the user clicks the first time.	All
Average number of pages visited	This is the average of the number of pages that were visited by all the participants before completing the task.	All
Median number of pages visited	This is the median of the number of pages that were visited by all the participants before completing the task.	All
Minimum number of pages visited	This is the minimum of the number of pages that were visited by all the participants before completing the task.	All

Maximum number of pages visited	This is the maximum of the number of pages that were visited by all the participants before completing the task.	All
Optimal completion	This is an indication of whether a participant completed the prescribed task exactly as prescribed	Evaluation case study 1 and evaluation case study 2.
Optimal completion Total	This is the total number of participants that completed the tasks in the manner that was prescribed exactly.	Evaluation case study 1 and evaluation case study 2.
Optimal completion Percentage	This is the percentage of all the participants that completed the tasks in the manner that was prescribed exactly.	Evaluation case study 1 and evaluation case study 2.

APPENDIX B: ETHICAL CLEARANCE CERTIFICATE

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Engineering,
Built Environment and Information Technology

1956 – 2016
60
years of
Engineering Education

Reference number: EBIT/83/2016 29 November 2016

Mr RM Kruger
Department Informatics
University of Pretoria
Pretoria
0028

Dear Mr Kruger,

FACULTY COMMITTEE FOR RESEARCH ETHICS AND INTEGRITY

Your recent application to the EBIT Research Ethics Committee refers.

Conditional approval is granted.

1. This means that the research project entitled "Towards a methodology to conduct inter-organisational comparative usability evaluations" is approved under the strict conditions indicated below. If these conditions are not met, approval is withdrawn automatically.

Conditions for approval

a) The applicant is not required to submit an updated application, but needs to explain how eye tracking will be done (which device will be used, and what is the potential risk to the participant)? This should be addressed in a letter to the EBIT Ethics Committee.
b) All data stored on a personal computer or other electronic platform should have password protection.
c) All participating organisations need to stay anonymous.

2. This approval does not imply that the researcher, student or lecturer is relieved of any accountability in terms of the Code of Ethics for Scholarly Activities of the University of Pretoria, or the Policy and Procedures for Responsible Research of the University of Pretoria. These documents are available on the website of the EBIT Ethics Committee.

3. If action is taken beyond the approved application, approval is withdrawn automatically.

4. According to the regulations, any relevant problem arising from the study or research methodology as well as any amendments or changes, must be brought to the attention of the EBIT Research Ethics Office.

5. The Committee must be notified on completion of the project.

The Committee wishes you every success with the research project.

Prof JJ Hanekom
Chair: Faculty Committee for Research Ethics and Integrity
FACULTY OF ENGINEERING, BUILT ENVIRONMENT AND INFORMATION TECHNOLOGY

Fakulteit Ingenieurswese, Bou-omgewing en Inligtingtegnologie
Lefapha la Boetšenere, Tikologo ya Kago le Theknolotši ya Tshedimošo

APPENDIX C: CONSENT FORMS

Informed consent form

1 Title of research project: Towards a methodology to conduct inter-organisational comparative usability evaluations.

2 I …………………………………………… hereby voluntarily grant my permission for participation in the project as explained to me by

…………………………………………………………………………………..

3 The nature, objective, possible safety and health implications have been explained to me and I understand them.

4 I understand that an interview will be conducted with me, and the information obtained from this interview will be used to evaluate the perceived value of an inter-organisational comparative usability study.

5 I also understand that a usability study of a website may be conducted with the use of an eye tracker and the observations made from my use of the website may be used as content for a usability report.

6 I understand that my involvement is anonymous and confidential. I am aware that the results of the investigation may be used for the purposes of publication.

7 I hereby also acknowledge that my interview and possible usability evaluations will be recorded and the recordings will be observed and interpreted for the purposes of this study.

8 Upon signature of this form, I will be provided with a copy.

Signed: ________________________ Date: ______________

Witness: ________________________ Date: ______________

Researcher: ________________________ Date: ______________

APPENDIX D: INTERVIEW QUESTIONS

Questions for post pilot CIUE case interviews

1. How old are you?
2. What is your gender?
3. What qualifications do you have?

Questions for post evaluation CIUE case interviews

1) Age? ___________ 2) Gender? ____________

3) Education level? (Please select one)
☐ No degree ☐High School ☐Attended university ☐Undergrad degree ☐Post grad degree

4) Please select the device(s) you make use of regularly (Can choose more than one)
☐ Smart phone ☐Laptop/PC ☐Video game console ☐ Tablet (e.g. iPad)

5) How would you rate your overall proficiency in terms of using the above devices?

Questions for post evaluation interviews

Present the comparative inter-organisational usability studies done in this dissertation and ask the following questions:

1. What is your opinion of the comparative inter-organisational usability studies?
2. What did you learn about your organisation from the comparative inter-organisational usability study that involved your organisation?
3. What changes would you make to the comparative inter-organisational usability studies done in this dissertation?
4. Depending on the results of the comparative inter-organisational usability studies' results, comment on your willingness to invest into usability evaluations.
5. Do you know about AB testing?
6. If yes, how does a comparative inter-organisational usability study compare to AB testing?

7. Do you think comparative inter-organisational usability studies can be used to facilitate AB testing?

APPENDIX E: EXCEL WORKBOOKS

Please refer to the following link for the excel workbooks:

https://drive.google.com/a/up.ac.za/file/d/0B5QM7wzMJOvNYnk1TzFpeDRLRm8/view?usp=sharing

Please note that sensitive data has been omitted from the workbooks made available.

APPENDIX F: RAW DATA FLOW ANALYSIS

Please refer to the following link for images containing the raw data flow analysis:

https://drive.google.com/drive/folders/0B5QM7wzMJOvNOThGVHNZaFB2VEU?usp=sharing

Please note that sensitive data has been omitted from the workbooks made available.

APPENDIX G: PREVIOUS PUBLISHED CONFERENCE PAPER

Please page over for a paper that originated from this research and that was presented at the CONF-IRM conference in Cape Town in 2016.

Association for Information Systems
AIS Electronic Library (AISeL)

CONF-IRM 2016 Proceedings | International Conference on Information Resources Management (CONF-IRM)

2016

The value of comparative usability and UX evaluation for e-commerce organisations

Rendani Maarten Kruger
University of Pretoria, rendi619@gmail.com

Helene Gelderblom
University of Pretoria, helene.gelderblom@up.ac.za

Wynand Beukes
University of Pretoria, wynand.beukes@up.ac.za

Follow this and additional works at: http://aisel.aisnet.org/confirm2016

Recommended Citation
Kruger, Rendani Maarten; Gelderblom, Helene; and Beukes, Wynand, "The value of comparative usability and UX evaluation for e-commerce organisations" (2016). *CONF-IRM 2016 Proceedings*. 9.
http://aisel.aisnet.org/confirm2016/9

This material is brought to you by the International Conference on Information Resources Management (CONF-IRM) at AIS Electronic Library (AISeL). It has been accepted for inclusion in CONF-IRM 2016 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

6. The value of comparative usability and UX evaluation for e-commerce organisations

Rendani Maarten Kruger[1]
University of Pretoria
rendi619@gmail.com

Helene Gelderblom[1]
University of Pretoria
helene.gelderblom@up.ac.za

Wynand Beukes[1]
University of Pretoria
wynand.beukes@up.ac.za

Abstract
The objective of this paper is to investigate the possible value of comparative user experience (UX) or usability evaluations for e-commerce organisations. Poor website usability has been identified as one of the main reasons why users abandon potential transactions. Appropriate evaluation of these sites is therefore essential. A problem with usability and UX evaluation is the lack of trust designers have in the evaluator's recommendations due to the subjective nature thereof. This paper investigates the possible enhancement of the objectivity of such evaluations through cross-company comparative evaluations, so that designers can assess their design success against that of direct competitors in the market. We conducted an empirical, comparative evaluation of three similar organisations' e-commerce websites using eye tracking as the primary data collection mechanism, and then demonstrated the potential value and usefulness of the outcomes.

Keywords
Web Design, User Experience, Usability, Comparative Evaluation, Eye Tracking, Telecommunications Websites

1. Introduction
In an investigation into why a large telecommunications organisation failed to invest in appropriate usability and user experience evaluation of their website despite the website's obvious importance for successful business, Beukes (2015) established that there is a general disregard for traditional good web design and evaluation practice amongst designers in the organisation. Very little attention was given to proper user experience (UX) and usability evaluation. The broader problem that we investigate here is how such organisations can be convinced of the value that UX evaluation and how this could contribute to their business success.

[1] Department of Informatics

Flawed website design is still a major barrier for successful e-commerce (Sivaji et al. 2011). A survey by Sivaji et al. amongst 10,000 online shoppers revealed that 30% of the respondents regarded poor website organisation as the main reason for abandoning potential transactions.

The value of evaluation is often obscured by negative influences of doubt that are rooted in the potentially subjective nature of evaluator recommendations (Law, van Schaik & Roto 2012). Also, UX is a theoretical concept that is difficult to operationalise in a way that prescribes what exactly should be measured when evaluating it (Law et al. 2012). An evaluation method that relies clearly on factual information may thus be of great value. UX and usability evaluation through eye tracking is one approach that endeavours to provide objective data about user interaction behaviour that, if analysed accurately and appropriately, could provide convincing evidence and design recommendations (Djamasbi 2014). Designers can however still discard these results as inaccurate or prone to evaluator bias if they are strongly committed to their design ideas.

This paper investigates a method that uses factual user behaviour data that are collected from interaction with the website to be evaluated, as well as data from user interaction with closely competing websites that offer directly comparable services. In this study we evaluated the websites of three similar telecommunications organisations using eye tracking as the main data collection method. Our aim was to determine if comparison of the evaluation results can overcome some of the reservations about the evaluation process and outcomes.

Although we believe our findings apply to UX and usability evaluation in general, our emphasis in this empirical case study was on the usability aspects of web design rather than on the more affective UX aspects. We do, however, see these concepts as inseparable.

2. Background

2.1 The value of usability and user experience evaluation

The ISO defines usability as the extent to which users can use a product to achieve their goals effectively, efficiently and with satisfaction within a specific context of use (Travis 2013). User experience (UX) refers to the experience(s) that result from encountering technology systems (Roto et al. 2010). UX includes encounters with systems – not only active, personal use, but also in a more passive way, such as observing someone else using a system. UX is unique to an individual, it is influenced by prior experiences and expectations, and is rooted in a social and cultural context.

Whereas traditional usability factors focus on performance and seamless interaction, UX factors are more concerned with affect, interpretation and meaning (Roto et al. 2010). Usability and UX are however very closely tied – if an online shopper cannot complete a transaction successfully because of usability issues, the UX will not be positive. This is confirmed by Rogers, Hutchinson and Fu (2010) who include 'task successes' as one of four crucial metrics in their UX evaluation framework.

It is difficult to quantify the value of UX and usability evaluation in comparison with its cost in terms of time and money, but researchers have for many years been trying to put a figure on the return on investment in proper user evaluation (Landauer, 1995; Heppner et al. 2005;

Weinschenk 2005). The results of launching flawed products often manifest in reduced usage which in turn can significantly impact on expected business. Studies have estimated that 80% of the total system maintenance costs incurred have been related to users having problems with the systems and not with a system's technical flaws; of this, 64% was directly related to system usability problems (Landauer 1995; Mentis & Gay 2003; Weinschenk 2005). Weinschenk quotes real world examples to explain the value of good usability and UX. One example shows that once a system is in development, it will cost 10 times more to fix the problem than when the same problem was solved during design. If the system has been released, this cost rises to 100 times more. In another example, a large computer company spent $20,700 on usability work to improve the sign-on procedure in a system used by several thousand people. The improvement in productivity was calculated to save the company $41,700 only on the first day the system was used. The benefit within the first year was $6,800,000.

According to Lee (2012), the user's virtual experience when using a company's technology will influence the strength of brand influence and the overall perception that the person has of that organisation. Clients can be attracted and retained by improving the performance of a website. E-commerce sites commonly drive away nearly half of repeat traffic by not making it easy for visitors to find the information they need (Weinschenk, 2005). Weinschenk reports on a study of an e-commerce site where first time users spent an average of $127 per purchase, while repeat users spent nearly twice that. Usable e-commerce sites build goodwill.

2.2 The purpose of comparative studies for organisations

Oxforddictionaries.com (2015) defines comparative as "measured or judged by estimating the similarity or dissimilarity between one thing and another". In reports on comparative usability studies available in the literature, different entities are used as the objects of comparison. Some studies compare different evaluation methods (Molich & Dumas 2008 and Ssemugabi & De Villiers 2007), some compare different implementations of the same system or prototype (Zhang & Moore 2014) and others compare different user groups using the same system (Arianezhad et al. 2013). Reports on studies that compare e-commerce applications of competing organisations are scarce. When initiating a business, a goal is to identify business potential in the current relevant markets. This requires a comparative study whereby the new business owners will compare the macro and micro environments of markets at the present time, to the potential macro and micro environments of the markets in the future (Pîndiche & Ionita 2013). A cross-company comparative study can reveal where the planned business stands in terms of its offering and how this overlaps with, or extends, the business of potential competitors (Czepiel & Kerin 2012).

From the above, when e-commerce is involved, comparison of competing organisations' online strategy and presence is essential (Czepiel & Kerin 2012; Weinschenk, 2005). The focus in our study is on how a cross-company comparative UX and usability evaluation of e-commerce websites can benefit a business. We define a cross-company comparative UX or usability study as the act of examining, in detail, the similarity or dissimilarity between the scientifically observed facts about user behaviour when using the technological interfaces of different organisations that have comparable purposes.

Any study that compares entities should follow some basic principles to ensure that the comparison is valid. Translating De Zepetnek's (1998) principles of a comparative literature study to a comparative UX or usability evaluation yields the following guidelines:

Comparable entities should be included in the study; this means the entities must serve a similar purpose, operate within similar environments and be based on similar backgrounds.

The tests for the various entities should be conducted in a similar way. This entails using the same tools, procedures and variables.

The results from the tests must be comparable in nature.

Against the background described above, we can now formulate the research objective of the study reported in this paper and articulate it in terms of a research question.

2.3 Research objective and question

Based on the purpose of comparative studies for organisations and the potential value of usability and user experience evaluations, the objectives of this study were:

1. To demonstrate how a cross-company comparative UX or usability evaluation (as opposed to distinct evaluations) could be conducted.
2. To investigate the results of such a study to determine what value (if any) is added through the comparison.

This was done by comparing the websites of three large organisations that conduct their business through their respective company websites, and showing how this can increase the perceived value of the evaluation results for the designers of the respective websites. First, we had to identify a suitable business sector, then choose at least three organisations whose websites provide similar services to customers with a similar profile. Next, we did separate evaluations focusing on one specific service offered by all three websites. Finally, we investigated whether a comparative analysis of the evaluation results provides more value than considering the three evaluations in isolation.

The question that we asked is: What value can e-commerce organisations derive from cross-company comparative UX and usability evaluation studies?

3. Methodology

This was an empirical evaluation study conducted in a usability laboratory. We chose a case study for our research design. We start this section by describing the case study and then the participants before explaining how data was collected and analysed.

3.1 The case study

We used a comparative case study (Yin 2003). Three prominent telecommunications organisations, all based in South Africa, with their respective client bases dispersed throughout Africa were selected for the study. We refer to them anonymously as TelecomA, TelecomB and TelecomC as their specific identities are irrelevant in the context of this research. Being a mobile service provider is one of the core business functions of all three organisations. Their business success relies greatly on their respective websites where their clients can perform similar tasks.

One of the core functions of a mobile service provider – mobile data top-up – was selected as the focus in this comparative evaluation.

Following exactly the same procedure in each sub-case, we conducted an evaluation of each website to determine how well they have designed their top-up functionality respectively, in terms of usability and UX.

3.2 Participants

For each organisation, fifteen to twenty users were recruited to participate in a user experience evaluation experiment. A combination of convenience and snowball sampling was used. The Tobii Studio software used to record their eye tracking data indicates the accuracy of the captured data and the data of participants with accuracy lower than 60% were not included in the data set. Table 1 describes the participant groups whose data were included. Two studies included 15 users and one 14. A samples size of 16±4 is accepted as adequate for usability studies (AlRoobaea & Mayhew 2014).

The participant profiles were similar. There was an even distribution between males and females, with slightly more females in each study. The age distribution and average ages were also comparable across companies. All participants use computers on a daily basis and none of them had exposure to the particular web site they were allocated to use during the evaluation (i.e. they had never been clients of that specific organisation).

	TelecomA	**TelecomB**	**TelecomC**
Number of participants	15	14	15
Gender distribution	8=F, 7=M	9=F, 5=M	9=F, 6=M
Age range	14-60	14-42	20-47
Average age	33	26	30

Table 1: Demographic information about participants

3.3 Data collection

Demographic data about users were collected prior to the evaluation task using a simple questionnaire that was either self-administered or evaluator-administered depending on the participant's preference.

User experience and usability data were gathered through eye tracking and informal post-test interviews. The interviews were only conducted when the evaluators observed specific behaviours during the interaction that needed clarification. Eye tracking is a method to record people's eye movements while they are looking at a stimulus. In UX and usability evaluation, it provides an objective measure of the users' attention on interface elements throughout the interaction period (Duchowski 2007). Capturing users' gaze patterns (i.e. saccades and fixation points) provides accurate information on what the areas of focus were and which parts of the interface they ignored.

The experiments were conducted at the University of Pretoria. We used a Tobii T120 desktop eye tracker to record users' eye movements while they completed the same task on each of the three websites. The eye tracker was calibrated for each participant's eyes to ensure accurate data recording. In this study a five-point calibration was used. Once calibration was completed the instructions were displayed to the participants, followed by the task they were required to complete. They performed a task known as “mobile data top-up” or “purchasing of a data bundle” after the following scenario was explained to them: “You just bought a new mobile SIM card. After arriving home you insert the SIM card into your phone and realise that you are not able to browse the Internet because you do not have any data available. On your computer, you open the TelecomX website to top-up your SIM card with mobile data”. When they started the task we had already logged into the website so that they did not need to go through the login process. The task ended when they indicated that they have located the top-up function.

After the eye-tracking recording, data was exported with the Tobii Studio software. The data included static gaze plots indicating users’ eye movements across selected pages of the respective websites (see Figure 1 for an example); time taken to complete the task, time and number of fixations until first click, time spent on the final screen until the top-up function was selected, and the number of pages visited during the task. These are standard metrics used in usability and UX evaluation.

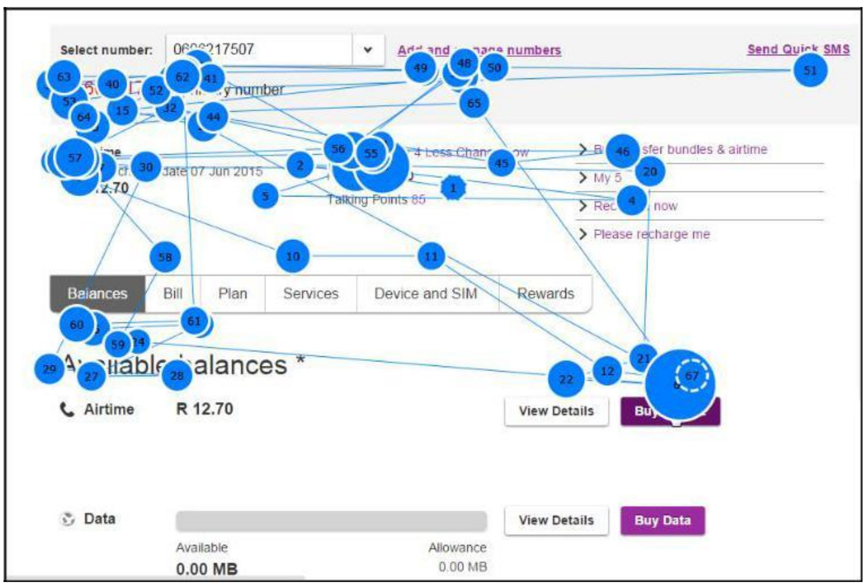


Figure 1: Example of a gaze plot

3.4 Data analysis

Eye tracking data were exported and analysed using the Tobii Studio software. This included quantitative metrics such as task completion time, number of fixations, time to first click, time spent on the home screen and time spent on the top-up screen. We transferred this data to an Excel spread sheet where descriptive statistics including minimum, maximum, average and median values were calculated for each set of metrics, for each of the three organisations. The results for the three organisations were then compared and the comparative data summarised in tables and graphs (see section 4 below).

4. Results of the comparative evaluation

The results are discussed with reference to time to complete the task, time and number of fixations until first click, time on destination screen until top-up is selected and the number of pages visited.

4.1 Total time to complete the task

Table 2 gives the time to complete the task on all three web sites in terms of the slowest user, the quickest user and the average and median times. Although the slowest respective users on the three web sites took equally long, there is a notable difference between the median and average completion times across the websites (see Figure 2). It took users of the TelecomC web site an average of 44 seconds to complete the task, while the average completion time for TelecomA was 3 minutes 14 seconds. This indicates a problem with TelecomA's site, especially if the competition allows users to complete the task in less than one minute.

Partcipant	TelecomA	TelecomB	TelecomC
Slowest	05:51	05:30	04:20
Average	03:14	01:35	00:44
Median	03:28	01:28	00:25
Quickest	00:50	00:21	00:14

Table 2: Time to complete the task

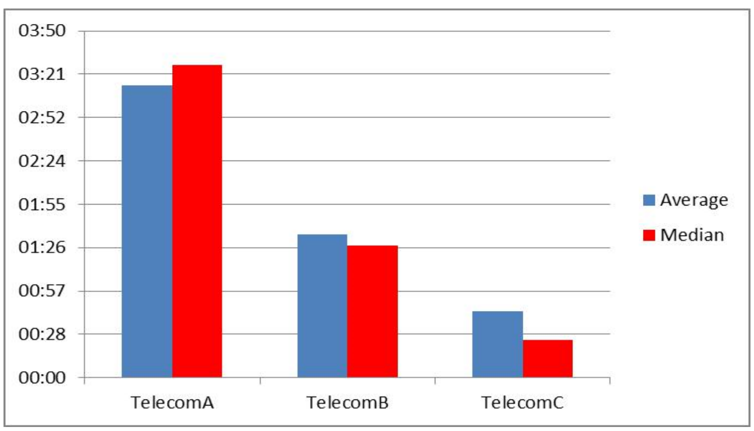


Figure 2: Average and Median time to complete the task

4.2 Time and number of fixations until first click

Table 3 gives the time it took users to make a decision to click on the home page, as well as the number of fixations they had on the screen up to the first click. Figure 3 compares the number of fixations. This data do not demonstrate significant differences in terms of the median and average times until first click, but there were a large number of fixations before first click (more than the competition) on TelecomB's website. An examination of the gaze plots of TelecomB's home page, confirmed that the users looked at a large range of elements before deciding on a route to take to complete the task. The number of elements on the TelecomB home screen may therefore be one of the reasons why there were so many fixations and so much time spent on the home screen before first click.

This data can however not be analysed in isolation because the first click may not have been on target, and this could lead users astray in the remainder of the task. In section 4.4 we look at the navigation paths which are more meaningful in this context.

	TelecomA	TelecomB	TelecomC
	Time (Fixations)	Time (Fixations)	Time (Fixations)
Minimum	00:00:06 (117)	00:00:02 (7)	00:00:02 (5)
Maximum	00:00:33 (48)	00:01:23 (244)	00:00:20 (67)
Median	00:00:10 (36)	00:00:19 (57)	00:00:11 (36)
Average	00:00:14 (12)	00:00:24 (71)	00:00:11 (34)

Table 3: Time (hour:min:sec) and fixations until first click

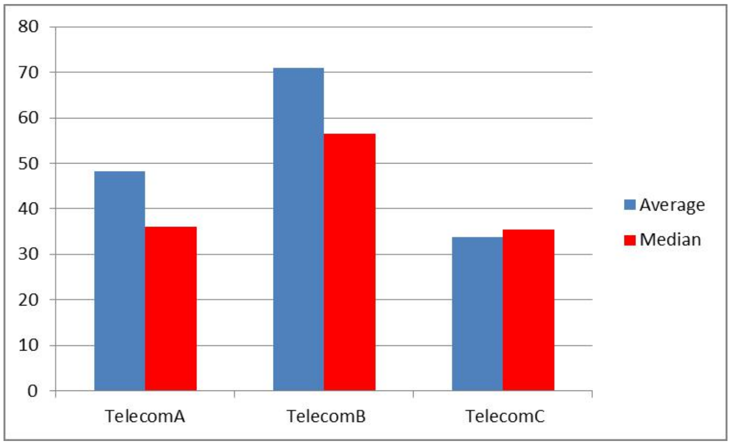


Figure 3: Fixations until first click

4.3 Time on target screen until top-up is selected

Here we compare how long it took users to locate the target on the final screen. As can be seen in Table 4, the slowest user on TelecomA's web site took just over three minutes, while the slowest users on TelecomB and TelecomC took 26 and 12 seconds respectively. This indicates a usability problem on TelecomA's website.

The quickest user on the TelecomA's top-up page took 26 seconds and the median time on that page was 41 seconds. This may set the designers at ease with regard to the slowest user, but comparing this time to the average and median times on the TelecomB and TelecomC web sites (see Figure 4) proves that there are probably usability problems on the TelecomA top-up page.

Partcipant	TelecomA	TelecomB	TelecomC
Slowest	03:06	00:26	00:12
Median	00:41	00:09	00:07
Quickest	00:26	00:04	00:04

Table 4: Time on last screen until top-up

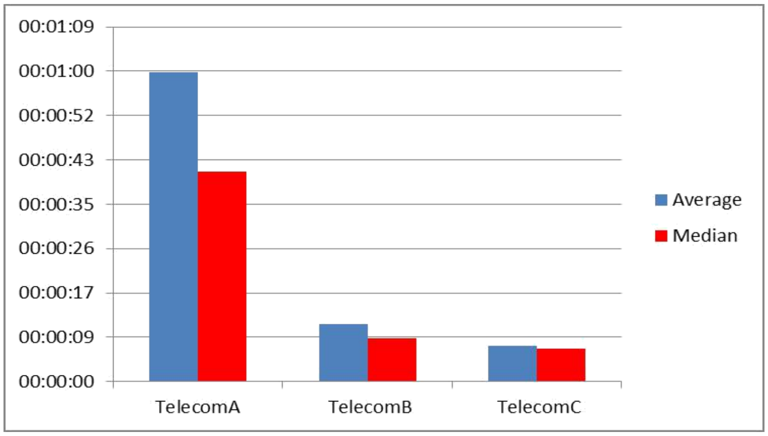


Figure 4: Time on last screen until top-up

The average number of pages visited by the TelecomA users is 8.87, while the averages for TelecomB and TelecomC are 3.36 and 2.57 respectively. This may be an indication of a problem with navigation on the TelecomA web site.

There is a correlation between the number of pages visited and the time spent completing the tasks. The more pages visited, the more the time spent completing the task. Assuming that time can be seen as an indication of the amount of effort it takes to complete a task then the average user put in the biggest effort to complete the same task on TelecomA's website. A supporting indication of this is the route taken to complete the task and the number of pages revisited.

Partcipant	TelecomA	TelecomB	TelecomC
Most	16 (1)	10 (1)	10 (1)
Average	8.87	3.36	2.57
Optimum	4 (3)	2 (5)	2 (13)

Table 5: Number of pages visited during the task (the number of users involved appear in brackets)

The results of the evaluation provide clear evidence that the TelecomA web site may have severe usability problems with their data top-up functionality on the website when compared to the TelecomB, and especially, the TelecomC websites.

5. Value added by the comparative evaluation

When considering the results of the evaluation of each of the individual websites in isolation, an experienced UX and usability evaluator would be able to identify problem areas in the design. For example, the number of screens that users visited on the TelecomA website to reach the top-up screen clearly indicates navigational problems, even without comparing it to the other websites. Also, the time taken to complete the task on the TelecomA website is an indication that there are usability issues. It would however be much easier for the evaluator to convince the TelecomA web designers that there are serious problems when the comparative data is presented together with the data about their own website. When they have the factual evidence that users of competing organisations find it easier to perform a function, they would be less likely to discard the evaluation results based on suspected evaluator bias or personal opinion.

The added value of the comparative results is further illustrated when considering user performance on the TelecomB website. Comparing their results with that of TelecomA will boost the organisation's confidence in the success of their web design. However, when compared to the results of TelecomC, the results appear less positive for TelecomB.

For TelecomC the outcomes were very positive. Their obvious superior results compared to the other two websites provide objective evidence that they have made design decisions that support usability – in particular with regard to the top-up functionality.

We recommend that comparative UX and usability studies include at least three competing organisations. Although we only tested one function – mobile data top-up – in our comparative evaluation and still obtained very useful comparative data, we would recommend that more than one function is tested for a more complete comparison.

6. Conclusions

This paper reported on an empirical study to demonstrate the value of a comparative usability study through a case study. The case involved three telecommunications organisations with similar lines of business and the same functionality was tested on all three websites with users with comparable demographic profiles.

Using eye-tracking technology, similar user behaviour data was collected for each website, including time spent to complete the task, time until first click, the number of fixations until first click and the number of pages visited. The data were then summarised to facilitate comparison across the organisations and analysed.

TelecomA's results revealed clear usability and navigational issues, especially in comparison to the results of TelecomB and TelecomC. TelecomB users had more fixations on parts of the website indicating that there are more screen elements than on the other two sites. Although TelecomB fared well when compared to TelecomA, its results show that TelecomC supports the top-up functionality much more effectively.

Although we did not report on the specific usability problems identified in the poorly designed websites, the next step from an organisation's point of view would be to identify and address the specific design issues. The added value of comparing UX and usability evaluation results across organisations has been successfully illustrated. The results from studies like these could be further used to set up benchmarks and improve the general ease of use of e-commerce websites.

References

AlRoobaea, R. and P.J. Mayhew (2014) "How many participants are really enough for usability studies?", In *Science and Information Conference (SAI)*, IEEE 2014, pp. 48-56.

Beukes, W. (2015) *Why Website designers disregard design guidelines that support usability*. M.IT Dissertation, Dept of Informatics, University of Pretoria, South Africa.

Czepiel, J. A., and R. A. Kerin (2012) *Competitor analysis*, Handbook of Marketing Strategy, vol. 41.
De Zepetnek, S. T. (1998) Comparative literature: Theory, method, application. Rodopi.
Djamasbi, S. (2014) "Eye tracking and web experience", *AIS Transactions on Human-Computer Interaction*, 6(2), pp. 16–31.
Heppner, C.C., Kates, J., Lynch, J.S., and R.R. Moritz (2005) "At Sprint, Understanding the Language of Business Gives Usability a Positive Net Present Value", in *Bias*, R.G and Mayhew, D.J., *Cost-Justifying Usability, An Update for an Internet Age*, 2nd edition, pp. 601–611.
Landauer, T. K. (1995) *The Trouble with Computers*. MIT Press.
Law, E. L. C., van Schaik, P. and V. Roto (2012) "To Measure or not to Measure UX: An Interview Study", *I-UxSED*, pp. 58-63.
Lee, K. (2012) "Consumer processing of virtual experience in e-commerce: A test of an integrated framework", *Computers in Human Behavior*, (28)6, pp. 2134-2142.
Mentis, H. and G. Gay. (2003) "User recalled occurrences of usability errors: implications on the user experience", *Extended Abstracts of the 2003 Conference on Human Factors in Computing Systems (CHI)*. ACM Press, pp. 736–737.
Molich, R. and J.S. Dumas (2008) "Comparative usability evaluation (CUE-4)", *Behaviour & Information Technology*, (27)3, pp. 263-281.
Oxforddictionaries.com (2015) *Oxford Dictionaries - Dictionary, Thesaurus, & Grammar*. [online] Available at: http://www.oxforddictionaries.com/ [Accessed 30 Oct. 2015].
Pîndiche, E. and R. Ionita (2013) "The Influence of Micro and Macro Environment Components on Trade Companies in Romania", *Journal of Knowledge Management, Economics and Information Technology,* (1)1.
Rodden, K., Hutchinson, H. and X. Fu (2010) "Measuring the User Experience on a Large Scale: User-Centered Metrics for Web Applications", *Proceedings of CHI 2010*, pp. 2395-2395.
Roto, L, Law, E., Vermeeren, A., and J. Hoornhout (2010) "User experience white paper. Bringing clarity to the concept of user experience", Result from Dagstuhl Seminar on Demarcating User Experience, September 15-18, 2010.
Sivaji, A., Downe, A.G., Mazlan, M.F., Soo, S.T. and A. Abdullah (2011). "Importance of incorporating fundamental usability with social and trust elements for e-commerce website", *International Conference on Business, Engineering and Industrial Applications (ICBEIA)*, IEEE, pp. 221-226.
Ssemugabi, S. and R. De Villiers (2007) "A comparative study of two usability evaluation methods using a web-based e-learning application", *Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists SAICSIT*, ACM, pp. 132-142.
Travis, D. (2013) *ISO 9241 for Beginners*. Ninth. London: Userfocus Ltd.
Weinschenk, S. (2005) *Usability a business case.* White Paper, Human Factors International.
Yin, R.K. (2003) *Case study research: Design and methods*. SAGE.
Zhang, S. and A.B. Moore (2014) "The Usability of Online Geographic Virtual Reality for Urban Planning", In *Innovations in 3D Geo-Information Sciences,* Springer International Publishing, pp. 225-242.