A comparative study on users' responses to graphics, text and language in a word processor interface

By

Tanya René Beelders

Submitted in fulfilment of the requirements for the degree

MAGISTER SCIENTIAE

In the Faculty of Natural and Agricultural Sciences
Department of Computer Science and Informatics
University of the Free State
Bloemfontein
South Africa

2006

Study leader:
Prof. P.J. Blignaut
Department of Computer Science and Informatics

Co-study leader:
Prof. T. McDonald
Department of Computer Science and Informatics

Declaration

I declare that the dissertation hereby submitted by me for the Magister Scientiae degree at the University of the Free State is my own independent work and has not been submitted by me at another university/faculty. I furthermore cede copyright of this dissertation in favour of the University of the Free State.

Tanya René Beelders

Date

Acknowledgements

I would like to express my sincerest thanks and gratitude to the following:

- Professor Blignaut and Professor McDonald for all your guidance, patience, assistance and insight throughout this research study.
- The personnel of the Department of Computer Science and Informatics at the University of the Free State for all your assistance. A special thank you to Mrs. Dednam of the Department of Computer Science and Informatics for your willingness and eagerness to help.
- To my family and friends, thank you for your encouragement and support.
- Thank you to the Telkom Centre of Excellence for the financial support.

Table of contents

CHAPTER 4: EXPERIMENTAL DESIGN AND METHODOLOGY

CHAPTER 5: PILOT STUDY

CHAPTER 6: TESTING NOVICE USERS

CHAPTER 7: TESTING TASK-KNOWLEDGEABLE USERS

CHAPTER 8: MENU AND ICON ERROR AND SELECTION ANALYSIS

CHAPTER 9: INTERPRETATION OF RESULTS

List of tables

List of figures and charts

Chapter 1

Introduction

1.1 Introduction

A word processor is a software application which allows for composition, editing and formatting of a printable document (Wikipedia, 2006). Text editors were the precursors of word processors and only allowed for composition and editing of text but provided no formatting capabilities (Wikipedia, 2006). The word processor has become a very popular tool in the everyday use of a computer (Roberts and Moran, 1983) and by 1984, 80-100% of users' time on a computer was spent using a word processor or other editor-based application (Rosson, 1984a).

The word processor application has evolved substantially since its initial inception. The original word processor - in the true sense of the word - was developed by IBM in 1969 and was known as the Magnetic Tape Selectric Typewriter or MT/ST (Eisenberg, 1992). In this model keystrokes were recorded on a 16mm magnetic tape and, while the MT/ST was capable of distinguishing between words, lines and paragraphs, the division of the full text into pages and the numbering of pages still had to be manually completed by a human operator (Eisenberg, 1992).

Since then the word processor has undergone a virtual metamorphosis to achieve the capabilities that are available in these applications today. The introduction of MS-DOS yielded great improvement in the capabilities of word processors with the inclusion of features such as endnotes, footnotes and the ability to edit more than one document by utilising the provision of increased memory and disk space (Eisenberg, 1992). The WYSIWYG (what you see is what you get) word processors, first released in the late 1980s, were popularised by applications such as Microsoft Word and MacWrite (Wikipedia, 2006). These applications took advantage of the potential of the graphical user interface and were capable of displaying multiple fonts (Wikipedia, 2006). Currently, Microsoft Word is the most popular word processor, with the number of users of the Microsoft Office suite being estimated at over 5 million worldwide (Wikipedia, 2006).

The word processor has become an integral part of everyday life for many people and as such it caters for a very diverse group of users. As is evident from the previous discussion, the word processor is constantly evolving to adapt to the needs of users and to exploit the increased capabilities offered by the newer technologies. Even so, it is highly unlikely that only one such complex application would be able to offer the best possible experience to all users (Sullivan, 1989). For these reasons, a word processor offers a unique environment and one rich in potential for improvement of the user experience. The word processor and the improvement of the usability thereof were the main focus areas of this research study.

1.2 Aim

The original aim of this research study was to determine whether the need existed for a localised word processor application for South Africa. However, no localised icons for South Africa could be obtained and attempts to design localised icons proved to be an almost insurmountable task. Thereafter, the aim shifted to determine whether or not it is necessary for a word processor application to be adapted for easier use by the South African market. This research study used a practical approach to actively examine whether user performance and usability of an application increased when using an interface that had been translated into the first language of the user, or when the icon type was changed, or when the icon set was adapted.

1.3 Motivation

By 2002 human-computer interaction (HCI) was still considered to be a relatively new field in South Africa (Hugo, 2002) and there is some concern that South Africa is not keeping pace with the rest of the world where HCI research and development is concerned (Miller, 2003).

South Africa is truly a multicultural nation whose cultural diversity is guaranteed protection and equal rights under the South African Constitution. However, despite these protective measures very little has been done in an effort to recognise cultural diversity in technology. Evidence of this is found in the fact that the majority of South

African websites are only available in English (Van Belle, Fellstad, Steele and Van Bakel, 2003), despite the fact that English is not the dominant language of South Africa (Chart 1.1).

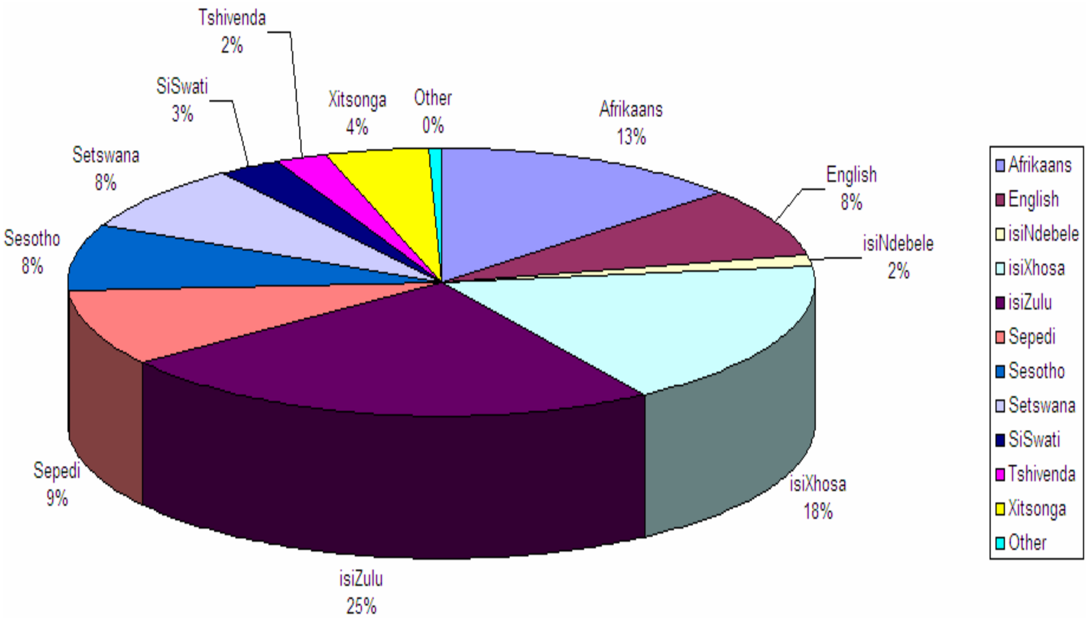


Chart 1.1: Distribution of South African home languages

Source: Digital census atlas (2001)

As South Africa attempts to establish equilibrium amongst its many cultures, it is important that technology does not get left behind. To prevent this from happening it is the responsibility of computer professionals to ensure that the cultural diversity of South Africa is fully represented in software applications.

Table 1.1 illustrates the lack of development in the technology field in South Africa, although the situation has improved somewhat since the 2005 publication date of the cited source. For example, the search engine Google is currently available in Afrikaans, Sesotho, isiZulu and isiXhosa and the open source software Open Office, which includes a word processor and spreadsheet, is now also available in all 11 official South African languages (Translate.org.za, nd).

Table 1.1: Current resources available in some South African languages

Source: Gee (2005)

As at January 2005	Browser	Word Processor	Spell Checker	Web Pages
Afrikaans	Yes	Yes	Yes	Yes
isiZulu	Yes		Yes	Few
Sesotho	Yes		Yes	
siSwati			Yes	Few
isiXhosa	Yes		Yes	
Setswana			Yes	Few
isiNdebele			Yes	
Sepedi			Yes	
Tshivenda			Yes	
Xitsonga			Yes	

This research study seized the opportunity to extensively investigate the possible ways in which usability of the word processor could be increased for South African users. In so doing it could be established whether the need existed for adaptation of the currently available applications to one more suitable for South Africans.

Iconic communication is the practice of using images to relate information in a nonverbal way (Lodding, 1983), and is a form of communication in human-computer interaction where icons have become a common interface component (Benbasat and Todd, 1993). The popularity of icon use in applications should allow for the development of a standard set of icons which can be used in all applications which include those functionalities (Hunt, 1996). Development of a tried and tested set of icons which are used in multiple applications will allow users to move seamlessly between applications without the added burden of having to learn a new set of icons (Richards, Barker, Banerji, Lamont and Manji, 1994). This is known as external consistency since it provides a measure of consistency for the users when they have to move between applications (Grudin, 1989). As it is, the icons used for common functions, such as Save, Cut and Copy, have become standardised through regular use, and the icons found in the Microsoft Office package have become the de facto standard for these functions. Although repeated use of icons confirms them as the established icons for the functions they depict, this does not mean that they should automatically be assumed to be the best images for depicting those functions (Hunt, 1996).

In addition, an intuitive or easy to use interface is not necessarily intuitive or easy to use for all users (Kang, 2003). Simply because the user responds immediately by clicking on an icon or menu option does not mean that they understand the meaning of that icon or menu option (Kang, 2003). In fact, it is quite possible that instead of the interfaces being intuitive to use through employment of a metaphor, users simply learn the meaning of the symbols and become accustomed to the terminology of graphical user interfaces (Hutchison, 1997). Users who are unsure of the meaning of icons and menus appear to resort to simply clicking by trial and error until they eventually happen upon the required icon or menu option (Kang, 2003) which counteracts the very reasons why icons and menus have become so popular. The reasoning behind the acceptance of standardised icons and the fact that interfaces are not suitable to all users highlighted the need to explore an alternative set of icons for use within a word processing application to determine whether a better, more intuitive set of icons could be provided to users.

Although menus and icons both motivate recognition above recall, the fact remains that the user is still forced to navigate through the menu in search of the correct menu option. For example, when wishing to italicise or bold, users will find no such menu options. Instead they have to recognise the fact that these are properties of the font style and as such they must invoke the font menu option. The danger also exists of creating submenus, or cascading menus, with increasing depth and breadth, which may confuse the user and lead to lower efficiency as the user gets lost in the myriad of ever-increasing paths leading to the next menu. Thus, although text processing is slower than picture processing, having to recall only a single word within the toolbar as opposed to having to navigate through a potentially complex menu structure to locate a menu option that may or may not have a name corresponding to the actual desired function, could increase the efficiency of the interface. Moreover, since it is easier to recognise an icon from a displayed set of icons as opposed to having to recall a command or function name from memory the question arises as to whether the same principle holds for the display of other interface components such as labels (Rogers, 1989b). In other words, is it the pictures that lessen the cognitive load or the presentation of the options to the user which increases the efficiency and effectiveness of the product? Therefore, it was important to test whether textual icons could increase the efficiency and effectiveness of a word processing package. No study was found that focused specifically on investigation of this issue within the context of a complex word processing application although similar studies have been undertaken using various other applications.

Finally, second language users encounter a number of problems when attempting to interact in a language which is not their primary language and processing of stimuli in a second language requires more effort and time than processing in an individual's first language. Translation of an interface remains a contentious issue and this research study approached this issue from the view of the bilingual user and the impact that bilingualism could have on the usability of a product.

1.4 Hypotheses

Three very broad general hypotheses were formulated for this study:

1. $H_{0,1}$: A set of alternatively designed icons does not influence the usability of a word processor.
2. $H_{0,2}$: A set of text buttons does not influence the usability of a word processor.
3. $H_{0,3}$: An interface in the first language of the user does not influence the usability of a word processor.

Within each hypothesis, usability incorporated effectiveness, efficiency and learnability of, as well as user satisfaction with a product. More specific hypotheses will be provided in each section and for each analysis.

1.5 Scope

The scope of the study was to focus on the effects of graphics, text and language on the usability of a word processor for a group of South African users. In this sense, graphics referred to the accepted standard set of icons which are currently used in a number of different applications, together with a set of icons which were designed as alternatives to these standard icons. The alternative icons were obtained either via previous studies or through brainstorming sessions where various alternatives were proposed and the most appropriate icon was chosen for each function.

Text was included in the evaluation of usability by providing a set of text buttons which replaced the pictorial icons on the toolbar. These text buttons displayed only the name of the function they represented and included no pictorial representation of the function. Menus and tooltips were also available in some interface configurations.

The effect of language on the usability of the interface was investigated by means of translation of the interface into Afrikaans and Sesotho. The original English interface was also still available and bilingual users were tested either on an interface in their first language or the English interface, where English was not considered to be their first language. The languages used in the study were determined by inclusion of only the dominant languages of the area in which the study was conducted.

Since it would be impractical to attempt to include all word processing functionality in a study of this size, only a small subset of the available word processing functions were investigated in this study. The functions included in the study were chosen because they were considered to be some of the most commonly used functions in a word processor environment.

1.6 Limitations of the study

Three limitations of the study were identified and are as follows:

1. Only a handful of the official languages of South Africa could be tested due to the fact that the area in which the study was conducted determined the availability of the languages present in test subjects.
2. The alternative icons were not designed by a graphic designer and may therefore have lacked the finer detail and refinement of the icons found in current applications.
3. Only a subset of the available functions in a word processor was included in the study since it would be impractical to attempt to include all available functionality in a study of this size.

1.7 Research methodology

The research problem attempted to determine whether or not usability is influenced by translation of the interface and whether text buttons or a set of icons different from the accepted industry standard could increase the usability of a word processor. Since user testing is considered to be the most robust form of usability evaluation (Nielsen, 1994), the decision was made to conduct user testing and to distribute user satisfaction questionnaires as a method of gauging user acceptance and satisfaction with the product.

A scaled-down word processor application was developed, which allowed for interchangeable interfaces. The interface configurations were compiled using pictorial icons and text buttons. Interface configurations which included menus and tooltips were also all provided in English, Afrikaans and Sesotho.

A diverse group of users, that were representative of the end users, were tested on the different interfaces. In order to comprehensively test the interface, it was necessary to ensure that all four recognised user groups were tested. These user groups were (i) first-time users, (ii) novices, (iii) intermediate and (iv) expert users.

Users were then required to complete a number of tasks that were representative of common actions in a word processor environment. Usability measurements as per the available usability models were captured in order to determine which of the interfaces was the most usable. Statistical analysis of the captured data allowed for evaluation of whether the different user interfaces did indeed affect the usability of the product.

1.8 Outline of the dissertation

The dissertation will proceed according to the following outline. The next chapter, Chapter 2, will be a literature review which will summarise some of the available literature in order to provide a detailed and comprehensive view of the foundation on which the research was based. Following this, Chapter 3 will provide a detailed description of the research methodology on which the experimental design was based. The full experimental design will be discussed in Chapter 4.

Chapter 5 will cover the results of the pilot study conducted to test the methodology and which provided some interesting insights in the form of formative evaluation. Chapter 6 will describe the tests conducted with novice computer and word processor users and the results obtained from these tests. Novice users were tested to ascertain whether an interface in their home language would increase the learnability of the application. In order to do this, two groups were tested on interfaces using the same icon set but different languages. The users were also tested to see how well they could adapt to changes in the interface by testing them on a different interface which they had never encountered before.

Chapter 7 will cover the user tests conducted with the first-time, intermediate and expert users of a word processor. These users were divided into groups based on their home language, and they completed the test on an interface which was either in their home language or in English, where English was not their first language. The computer anxiety and attitude levels of these users were first tested using established anxiety and attitude questionnaires. These users were tested in order to determine whether the interface language affected their performance or their satisfaction when using the application. After establishment of the influence of interface language the tests were re-evaluated based on the icon set to highlight any performance difference between the icon sets. The subjective satisfaction experienced by these users was also measured by means of a questionnaire.

Chapter 8 will provide a detailed analysis of the errors made by the first-time and expert users during completion of the test. This analysis served to highlight the common errors and difficulties experienced by the users. Trends were identified which assisted in discovering usability problems or general difficulties experienced by the users during interaction with the application.

Chapter 9 will summarise the results of the tests and conclusions will be drawn from the results obtained. The conclusions include a recommendation as to whether the interface should be translated into the home language of users having the same profile as the participants in this study and a decision will be formulated as to whether the icons influence the performance of word processor users. A recommended interface which was compiled using the results of each individual task is also proposed. The final chapter, Chapter 10 will provide a final conclusion and summarises the entire dissertation.

1.9 Summary

Specifically, this research study was concerned with the effect of graphics, text and languages on a word processor application. The word processor was identified as the class of application most suited to investigate the identified hypotheses and aims of this study. The scope of the research study was contained within a certain set of predefined functions available in word processor applications and interfaces were only provided in the predominant languages of the area in which the research was conducted. The motivation for development of an alternative set of icons stemmed from the observation that the accepted standard set of icons currently being used cannot be acknowledged as being the best icons for depiction of their respective functions as a result of their repeated use. Text buttons, which contained no graphical depiction of the associated function, provided a means to determine whether the phenomenon of unequal processing of verbal and nonverbal stimuli can be extended to the user interface of a word processor application. Similarly, the unbalanced quality between the processing of first and second language stimuli provided strong motivation for the need for translation of the interface, motivation which still had to be confirmed through empirical testing.

User testing on a developed word processor application, which was capable of automatic logging of user actions and administration of a pre-defined set of representative tasks, was determined as being the best method of testing the usability of the interface configurations.

This chapter has provided a brief introduction of the research study discussed in this dissertation. The discussion presented a foretaste of the literature upon which this study is based and which provided the background and theoretical foundation for the study. The following chapter is therefore a detailed discussion on some of the literature available which provided the basis and sound motivation for the undertaking.

Chapter 2

Theoretical background

2.1 Introduction

In any research study it is essential that the research be founded in existing literature. This literature may detail previous studies that are similar to the current undertaking, and they may provide the background upon which the current study is based. The preceding chapter gave a broad view of the aim, motivation and scope of the current study and presented a brief preview of the literature which was consulted during the study. This chapter will discuss some of the available literature which formed the foundation of the study and which provided a starting point for the study. Important terminology and concepts will be discussed to establish the groundwork for further discussion.

2.2 Human-computer interaction and the user interface

Human-computer interaction is the "study, planning, and design of how people and computers work together so that a person's needs are satisfied in the most effective way" (Galitz, 2002). It has also been defined as the "set of processes, dialogs and actions through which a human user employs and interacts with a computer" (Baecker and Buxton, 1987).

The user interface of a system is defined as the part of the computer or application which the user is able to see, hear, understand or direct (Galitz, 2002) and provides a "communication tool between man-made machines and humans" (Kang, 2003). The user interface is both the text and imagery which is presented to the user via the computer screen and which facilitates the communication between the users and the computer (Zayas, 1996). This communication can be in terms of the display, feedback and the method by which the users inform the application what action must be carried out next via interaction with the display elements (Mayhew, 1999). Essentially, the user interface is that part of the system that provides a means for interaction between the user and the computer (Redmond-Pyle and Moore, 1995) and as far as the user is concerned, the user interface is the entire application (Mayhew, 1999).

A graphical user interface (GUI) consists of the icons, taskbars, and other elements which are used by the computer to display information to the user (Tuck, 2001). When making use of a graphical user interface, interaction with the system is achieved through use of a pointing device, such as a mouse, and users manipulate and carry out actions on objects (Galitz, 2002). This type of interaction is known as direct manipulation (Galitz, 2002). Specifically, direct manipulation refers to an interaction style whereby the user can point at a visual representation of the task, manipulate that representation and then immediately see the effects of the manipulation (Maguire, 1990).

An interface provides a boundary between user and computer, and as such is responsible for the smooth facilitation of the dialogue between the two components (Barker, Najah and Manji, 1987). Thus, the interface serves as a point of contact between the two components of human-computer interaction, namely the human and the computer. It is in this role that the interface is the component that must provide easy learning and use of the application (Jervell and Olsen, 1985). So important is the user interface, how it is perceived and how well it promotes usability, that it is considered to be the single most important factor determining the success or failure of a product (Baecker and Buxton, 1987). Regardless of how well the system is designed or how well it encapsulates all the required functionality, if the user interface does not facilitate ease of use and increase usability then the product is likely to fail (Baecker and Buxton, 1987). This is due to the fact that software that is difficult to use not only wastes the time of the user and causes frustration but also discourages continued use (Bevan and Macleod, 1994). The importance of creating usable and intuitive user interfaces is triggered by its central role in the acceptance of technology and the computer into everyday tasks and situations (Zayas, 1996). As such, the interface must make the system usable by providing usability (Redmond-Pyle and Moore, 1995).

Since the user interface plays such a central role in usability and acceptance of an application it is vital that it provide the best possible experience for the user. This study will be evaluating the graphical user interface of a word processor application in an effort to determine which combination of interface components provides the greatest usability for the targeted users and whether the current interface can be improved upon in any way to increase acceptance and to facilitate easier interaction. As such the research is conducted within the field of human-computer interaction.

2.3 Usability

"Software usability is no longer a luxury, but rather a basic determinant of productivity and of the acceptance of software applications." (Abran, Khelifi, Suryn and Seffah, 2003b).

Various definitions of usability have been proposed over the years, some of which will be mentioned here. Pearrow (2000) offers a comprehensive definition of usability as the "discipline of applying sound scientific observation, measurement, and design principles to the creation and maintenance of the Web sites in order to bring about the greatest ease of use, ease of learnability, amount of usefulness, and least amount of discomfort for the humans who have to use the system". Although this definition is aimed at websites in particular, it can easily be extended to include traditional software applications. It also manages to encompass testing methods and tools available to designers for ensuring the design and maintenance of usable products (Pearrow, 2000). A definition of usability that is pristine in its simplicity is that of Cato (2001) who said that "Usability is being able to do the things you want to, not the things you have to". This can be translated as the extent to which the product assists the user in reaching his/her goals, as opposed to becoming an additional obstacle to overcome in order to accomplish those goals (Levi and Conrad, 1998).

According to the International Standards Organisation (ISO) standard 9126-1 usability is "the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions". This definition is further expanded upon in ISO 9241-11 where usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO, 1998). Two important points which are very apparent in the extended definition of usability as defined by the ISO 9241-11 are, firstly, that usability is not concerned with interface features but rather with measurements of human-computer interaction (Dillon, 2001), and secondly, it is important to note that the definition is very specific about users performing tasks (Redmond-Pyle and Moore, 1995).

Since these definitions, particularly those of the ISO, lend themselves so appropriately to the study at hand, they form the basis of the accepted usability definition for this study, which was derived from a combination of the two ISO definitions. Therefore, for the purposes of this study, usability will be defined as follows:

The capability of the software product to be not only understood and learned but also to be used by specific users to achieve specified goals with effectiveness and efficiency. Furthermore, a usable system must incorporate the capacity to be attractive to the user and provide satisfaction during and after use.

Furthermore, for the purposes of this study this measurement of usability will be determined by applying sound scientific observation, measurement, and design principles by means of using available design guidelines and usability testing techniques.

Many usability definitions, including the derived definition for this study, contain the same keywords which allow for four distinct components of usability to be identified, namely effectiveness, efficiency, satisfaction and learnability (Abran, Suryn, Khelifi, Rilling and Seffah, 2003a; Brinck, Gergle and Wood, 2002). These components provide a means for measuring the usability of a product (Preece, Rogers, Sharp, Benyon, Holland and Carey, 1994) according to the recommendation contained within this study's derived definition, and are defined as follows:

- Effectiveness is how well the user is able to achieve that which must be done by using the system (Bevan and Macleod, 1994; ISO, 1998) and can be measured in terms of accuracy and completeness (Cato, 2001; ISO, 1998).
- Efficiency is the amount of resources required to complete the desired task (ISO, 1998), such as time, money or mental effort (Bevan and Macleod, 1994).

- Satisfaction is a subjective feeling relating to the attitude of the user towards the system (ISO, 1998).
- Learnability measures not only the time taken for a user to become familiarised with the system but also how well the user is able to remember system functionality (Cato, 2001).

Due to a rapid increase in the number of computer users more emphasis needs to be placed on the importance of the usability of the computer interface (Nielsen, 1992). Users may already have experienced computer applications that facilitate ease of learning and that are also a pleasure to use (Nielsen, 1992). Once they have used such systems, they are no longer willing to settle for applications that are cumbersome and require extensive learning time (Nielsen, 1992). These facts all contribute to the increased awareness and emphasis that is placed on creating a product which is usable (Nielsen, 1992). Usability is of paramount importance since it has been revealed that user preference and performance are not always complementary. This means that the interface that is most preferred by the user does not necessarily yield the best user performance (Galitz, 2002). Even for a graphical user interface usability is crucial due to the richness and complexity of the system which can easily lead to confusion for the user (Redmond-Pyle and Moore, 1995).

The primary aim of this study was to determine the effect and impact of various interface elements on the usability of the system. Specifically, icons chosen by a group of non-computer literate questionnaire respondents were tested to determine whether the preferred icons increased the usability of an interface. The set chosen by the non-computer literate respondents was completed by designing icons according to available guidelines. The set of standard icons, as currently found in the Microsoft Office packages, and a set of text buttons were also included in the study. In this way a number of different interface configurations were developed which could be compared and tested for usability. By measuring usability factors on these interfaces, this study aimed to determine which interface configuration is best suited to each participating user group. The interface which will be deemed the most suitable will be the one that shows the greatest usability in terms of the usability measurements identified and included within the dissertation's accepted definition. These usability measurements are efficiency, effectiveness, satisfaction and learnability and the metrics available to measure these components will be discussed in Section 3.4.1.3.5. The icons, interfaces and usability measurements used in the study will be discussed in greater detail in subsequent chapters.

2.4 Types of computer users

Users are those people who will eventually interact with the product or application (ISO, 1998). Shneiderman (1998) defines three distinct classes of users, namely first-time or novice users, knowledgeable intermittent users and expert frequent users.

First-time or novice users

The first class of users is subdivided into first-time users and novice users (Shneiderman, 1998). Novice users are defined as having little knowledge of either the task or the interface domain, whilst first-time users have knowledge of the task concepts but limited knowledge of the interface concepts (Shneiderman, 1998).

Knowledgeable intermittent users

These users have knowledge of both the task and interface concepts, but because of their staggered and infrequent use of the application, they do not retain the knowledge gained and may struggle to find the correct interface component to use in order to accomplish the desired task (Shneiderman, 1998).

Expert frequent users

This type of user is extremely competent in both the interface and task domain and strives to complete the task at hand as quickly as possible and with as little as possible effort and actions on their part (Shneiderman, 1998).

All of these user groups will be included in this study as it will be shown in a later section (Section 3.4.1.3.3) that these user groups are also applicable to user testing which formed the greater part of the study.

2.5 Computer anxiety and attitude

The terms "computer anxiety" and "computer attitude" have been used somewhat interchangeably in the literature and in the development of measurement scales (Kernan and Howard, 1990). These concepts are however, two different and separate constructs and should be treated as such (Kernan and Howard, 1990).

2.5.1 Computer anxiety

Anxiety is "a vague, unpleasant feeling accompanied by a premonition that something undesirable is about to happen" (Kagan and Havemann, 1980). A person who is in a state of anxiety is excitable, irritable or fearful (Landauer, 1972). Phobias are intense fear or anxiety reactions to particular objects and situations (Landauer, 1972). Anxiety can be subdivided into at least cognitive anxiety and somatic anxiety (Woodman and Hardy, 2001). Cognitive anxiety is a mental component of anxiety and constitutes worry and nervousness (Morris, Davis and Hutchings, 1981). It is associated with information processing (Eysenck and Calvo, 1992) and can be manifested in the form of feelings of helplessness, confusion, apprehension and negative thoughts (Peurifoy, 1992). The second component of anxiety – somatic anxiety – is more physiological in nature and is recognised by symptoms such as an increased heart rate, shaky hands (Morris et al., 1981), nausea and sweating (Landauer, 1972).

When anxiety is experienced by a person who is confronted with the possibility of working with computers (Johassen, 1985) or whilst they are using a computer (Martin, 1998), they are said to be computer anxious. The persistent fear that is present during interaction with a computer or even at the thought of possible interaction with a computer, is referred to as computer anxiety, technostress, cyberphobia (Martin, 1998) or computerphobia (Jay, 1981). Computer anxiety is commonly characterised as feelings of negativity towards computers which can be manifested as "(a) resistance to talking or even thinking about computer technology; (b) fear or anxiety, which may even create physiological consequences; and (c) hostile or aggressive thoughts and acts, indicative of some underlying frustrations" (Jay, 1981).

Anxiety may influence a person's performance in any given situation. The inverted U-hypothesis is a popular model which formalises the relationship between anxiety and performance (Rauh and Seccia, 2006). Increases in anxiety enhance performance when anxiety levels are low, but performance deteriorates with increased anxiety when the initial anxiety levels were already high (Rauh and Seccia, 2006). The processing efficiency theory concentrates on the relationship between cognitive anxiety and its effects on performance (Eysenck and Calvo, 1992). Anxiety increases worry and tension, resulting in cognitive resources being used by anxious thoughts as opposed to concentration on the task at hand, which hampers performance (Eysenck and Calvo, 1992). However, although anxiety can impact negatively on performance, it can also serve as a motivational factor, thereby increasing performance due to heightened effort (Eysenck and Calvo, 1992). When the anxiety levels are too high and they are detected to be interfering with the task, additional effort is applied to the task to counteract the negative impact of worry (Eysenck and Calvo, 1992).

Taking both the inverted U-hypothesis and the processing efficiency model into account provides the argumentation that performance will be affected by the anxiety levels present during completion of the task at hand. Anxiety will positively influence the performance of the individual should the initial anxiety levels be low, while anxiety will initially impact negatively on the performance of the individual when the initial anxiety levels are high. The theory behind the processing efficiency model then comes into play as eventually it will be detected that the anxiety is interfering with the task and more effort will be applied to the task.

Lack of knowledge about computers can induce psychological fear (Martin, 1998), thus increasing anxiety on the part of a user and resulting in an inverse relationship between computer anxiety and computer experience (Kernan and Howard, 1990). Computer experience has been found to have the most profound influence on computer anxiety (Maurer, 1994), and many studies have confirmed the correlation between computer experience and low anxiety (for example, Glass and Knight, 1988; Joncour, Sinclair and Bailey, 1994; Szajna, 1994).

Computer anxiety can affect the performance of computer users. People who are subject to a high degree of computer anxiety tend to expect poor performance on a computer task and report having more negative thoughts during execution of a task than users with low computer anxiety (Glass and Knight, 1988). High anxiety users required more time to complete computer-based tasks but did not make significantly more errors than low-anxiety users (Glass and Knight, 1988), a discovery that correlates to the fact that anxiety reduces the ability to concentrate on the task at hand since cognitive resources are exploited by nervous thoughts (Wine, 1971). Therefore, high anxiety users should take longer to complete a task without the accompaniment of a heightened error rate. The ability of the user to complete a task correctly or not has been shown to be a direct result of computer anxiety, as high computer anxious users had a lower correct response rate (Brosnan, 1998). A well designed system, amongst other possible methods, is capable of relieving computer anxiety (Gardner, Render, Ruth and Ross, 1985).

A number of questionnaires which provide a means of measuring computer anxiety within an individual have been developed and tested over the years, including:

- the Computer Anxiety Rating Scale (CARS) (as used in Glass and Knight, 1988);
- the Computer Anxiety Index (CAIN) (as cited in Martin, 1998);
- the Computer Anxiety and Learning Measure (CALM) (McInerney, Marsh and McInerney, 1999); and
- the Computer Anxiety Scale (CAS) (Marcoulides, 1989).

In this study, it was deemed important to gauge and measure the anxiety levels of the computer user accurately to ensure that it was in fact manipulation of the desired variable causing the observed effect and not some other influence such as anxiety of the computer user. The anxiety levels of the largest group of participants in this study were measured using one of the available questionnaires (Section 4.6). This study proposes that experience determines the anxiety levels of the user. Since anxiety is a feeling of impending doom or difficulty it was expected that anxiety levels would decrease as the degree of user experience increased. This supposition could be motivated by the reasoning that the more comfortable the user becomes when working with a computer the less they will fear the unexpected. Computer experience was not measured, although word processor experience was, which complicated the expectations of the anxiety results. It would not be unfounded to assume that a high degree of word processor experience would correspond to a high degree of computer experience but this does not necessarily work in reverse. This means that low word processor experience does not necessarily reflect low computer experience. Therefore, it was difficult to predict what the correlation, if any, would be between word processor experience and anxiety levels. It could be expected that high word processor experience would signal low anxiety but whether low word processor experience would result in high or low anxiety levels could not be accurately predicted. Nevertheless, is was imperative that anxiety levels be measured to ensure that a single user group did not exhibit abnormally high or low anxiety levels when compared to the other user groups.

2.5.2 Computer attitude

Attitude is defined as the way a person feels or thinks about another person, object or event (Landauer, 1972). More specifically it is “a mental and neural state of readiness exerting a directive influence upon the individual’s response to all objects and situations with which it is related” (Allport, 1935) which naturally predisposes a person to behave in a certain way (Kagan and Havemann, 1980).

Psychological characteristics, such as attitude, can also affect a user’s performance (Galitz, 2002) and it have in fact been found to do so (Jawahar and Elango, 2002). Computer experience has been found alternatively to influence attitudes towards computers (for example, Busch, 1995; Loyd and Gressard, 1984a; Orr, Allen and Poindexter, 2001) or to have no effect on attitudes towards computers (for example, Pope-Davis and Twing, 1991).

A number of questionnaires which provide reliable estimations of computer attitude have been developed over the years. Some of these questionnaires include:

- Attitudes Towards Computers (Reece and Gable, 1982);
- Computer Attitudes Scale, or CATT (as cited in Kernan and Howard, 1990);
- Computer Attitude Scale, or CAS (Loyd and Gressard, 1984b); and
- ACTUS: Attitudes-Toward-Computer Usage Scale (Popovich, Hyde and Zakrajsek, 1987)

The attitudes of the users in this study were measured using one of these established scales (Section 4.7). Attitude towards computers can be linked to the anxiety level (Loyd and Gressard, 1984b) of the user and computer experience has been shown to be the greatest influence on computer anxiety. Furthermore, there exists a definite correlation between computer experience and low anxiety and therefore it was expected that attitude of the users in this study would be directly related to the level of computer experience of the user. It was also proposed that attitude would influence performance, not only because anxiety is a component of attitude but also because psychological characteristics, like attitude, have been known to have an influence on the performance of the user. Therefore, attitude levels of the users were measured for the same reasons that anxiety was measured – to ensure that some users groups did not have a disproportionately elevated attitude level in comparison with the other user groups.

2.6 Software trends

Historically, the majority of software development occurred in the United States of America (Evers, 2001) – with 75% of all installed software packages worldwide in 1994 being produced in this country (Miller, 1994). Since then, Asian countries have emerged as strong competition to the American market (Evers, 2001), although more recently, the above figure has been estimated as having increased to the order of 80% of software packages being produced in the United States of America (O' Sullivan, 2003). These trends bring two phenomena to the fore, namely:

1. With development occurring predominantly in the United States of America, the graphical user interfaces (GUIs) developed for these products have a distinctive American look and feel to them (Evers, 2001); and
2. with the emergence of serious competition, usability of a product has become an important issue.

Recent years have shown a substantial increase in international growth in the information technology (IT) sector, in particular due to the increasing popularity of the Internet. This trend is evident in the fact that 50% of the sales revenue for most computer companies is generated from markets outside of the country of development

(Nielsen, 1996b), with international sales accounting for up to 50% of the total revenues for the top 100 US software companies (Miller, 1994). For example, international sales accounted for 55% of Microsoft's revenues for the 1992 fiscal year, and by 1994 Microsoft was shipping their flagship product – the Windows operating system – in 27 languages (Miller, 1994). The year 1995 marked a growth spurt for international markets with Asian software sales increasing by 90% and sales to India, New Zealand and China increasing by over 100% (Maner, 1997). Large companies, such as Apple and IBM, had international sales of 48% and 50% respectively for the 1995 business year (Maner, 1997).

This growth of the information sector into a truly global market forced developers to recognise the increasing diversity of the users and to take cognisance of the need to accommodate the diverse user base (Jackson, Biocca, Lim, Bradburn, Tang, Mou, Barbatsis, Von Eye, Zhao and Fitzgerald, 2003). Even so, in an investigative study of the corporate websites of 10 different IT companies, it was found that the same patterns of colour, symbols and images were used on the websites regardless of the country in which the company was based (Kang and Corbitt, 2003). Most of these websites were developed by American companies (Kang and Corbitt, 2003), thus the sites all had a distinctive American look and feel to them. Metaphors and values that are specific to the American culture are confusing to non-American users (Maner, 1997). If the users were to attempt to adapt to the interfaces in the original American format, it would require getting used to many different concepts, such as addresses, postal codes, measurements, punctuation and many more (Maner, 1997). These vast differences could account for the fact that users from other countries resent the distinctive Western feel and metaphoric use, to the extent that the users may even reject the use of the products (for example, Russo and Boor, 1993; Zahedi, Van Pelt and Song, 2001).

Heavy emphasis is currently being placed on achieving the goal of universal access to information and communication services and ensuring that the widest possible audience is reached by information technology (Shneiderman, 2003). The emergence and expansion of the so-called global market forces companies to develop the interface to be flexible enough for the overseas market (Abdelnour-Nocera, 2001). Corporations and designers who are unwilling to accommodate the diversity of the end users, can expect limited acceptance of their product and a flood of lost sales which will go to a competitor who is willing to consider user diversity (Bickford, 1997). For an application or website to be accepted by the broader population, it must reflect the needs of each community which it targets (Galitz, 2002), which in essence requires the culture of the target market to be considered during the design process.

2.7 Culture

Raymond Williams famously asserted that “culture is one of the two or three most complicated words in the English language” (Giles and Middleton, 1999). This is evident in the fact that most researchers are in agreement that the word “culture” is one of the most difficult terms to define and almost everyone studying culture has a different perspective on it (Gardiner and Kosmitzki, 1998). In fact, culture has been defined in hundreds of ways over the years, where each of the individual definitions highlight different aspects of culture and where some of the definitions are even in conflict with one another (Hall, 2002).

The anthropologist E.B. Tyler was the first to use the term in *Primitive Culture*, published in 1871. In this publication, culture was defined as “that complex whole which includes knowledge, belief, art, morals, laws, customs and any other capabilities and habits acquired by man as a member of society” (Gardiner and Kosmitzki, 1998).

In general, anthropologists agree that culture stands for “the way of life of a people, for the sum of their learned behavior patterns, attitudes, and material things” (Hall, 1959). However, although they agree on this general description most disagree on what exactly the substance of culture is, which has led to varying definitions of culture in the literature (Hall, 1959). Some examples of the definitions developed over the years follow:

- “A culture is learned by individuals as the result of belonging to some particular group, and it constitutes that part of learned behaviour which is shared with others” (Kluckhohn, 1949, p 26).

- Culture is “the collective programming of the mind which distinguishes the members of one group or category of people from one another” (Hofstede, 1991, p 5).
- “Culture is a historically shared system of symbolic resources through which we make our world meaningful” (Hall, 2002, p 4).

Language, music, public emotion (Trompenaars, 2004), symbols and values (Hofstede, 1991) are all part of the culture of a person.

2.7.1 Internationalisation and localisation

Internationalisation refers to the product’s readiness to be adapted to any specific locale by ensuring that it is not locale-specific, that is, there are no language or other locale-specific features of the product that are fully integrated and hard-coded (Ott, 1999). It involves the development of a generic end-product with full separation of culture-dependent and independent parts (Hars, 1996). Developing in such a way makes it possible to use a single interface on a world-wide basis (Nielsen, 1999).

At the other end of the scale is localisation of a product, which implies that the product has been adapted with a specific locale in mind (Nielsen, 1999; Ott, 1999). In other words, culture-dependant parts, which are applicable to the target culture, are integrated into the generic or internationalised product (Taylor, 1992).

2.7.2 Localisation of a product

The first and most important and, some would say, the only step in localisation, is translation of the interface (for example, Esselink, 1998; Fowler and Stanwick, 1995; Nielsen, 1999). However, it is important to note that localisation is not synonymous with translation; translation is merely the adaptation of a product into the language of the target market (Collins, 2001; Gribbons, 1997; Hars, 1996; Keniston, 1997; Ott, 1999). Users have exhibited distinguishable preferences for interface components such as language, navigation, symbols and colour (Cyr and Trevor-Smith, 2004). These facts motivate the need for careful consideration of translation and icon development, amongst other things, in user interfaces (Johns, 1997). The need for localisation and cultural usability studies is far from over and researchers have only begun to scratch the surface regarding the influence culture has on the five components of interfaces (Marcus, 2006), i.e. mental models, metaphors, interaction, navigation and appearance (Marcus, 2001).

The initial aim of this study was to attempt to develop a localised version of a word processor and then to determine the effect of that localised product on usability. A limited number of studies have focused on localised icons for South Africans (Taylor and de Villers, 2005), but none could be found that localised word processing icons. After several exhaustive attempts to design localised versions of the icons used in a word processor as well as discussions with anthropologists on the subject, it was eventually decided that such a task was virtually impossible. Therefore, the focus of the study instead shifted to determining whether the icons found in popular word processors could be improved upon. Language was still considered but due to the unique quality of the language situation in Africa and South Africa (see following section), the study concentrated on bilingualism and how it impacts the usability of a word processor.

2.8 Bilingualism

Originally a bilingual individual was considered to be someone who was capable of and had mastered communication in two languages, but over the years the literature has extended the use of the term bilingual to encompass individuals who are capable of language communication in more than one language (Adler, 1977). Language communication also no longer only refers to equal proficiency in the spoken languages but includes “passive-knowledge” of a language in the written form and the ability to use the language to “complete meaningful utterances” in the environment of that language (Mackey, 2005). As such, bilingualism now encompasses all individuals having knowledge and some, possibly limited, proficiency in more than one language (Mackey, 2005). Bilingual individuals are capable of using the spoken languages interchangeably (Wei, 2005).

One third of the world’s population is able to communicate in more than one language (Wei, 2005). The situation becomes a little more complex on the African continent which has already been described as a "linguistically distinct continent" due to the complexity of the language situation found there (Batibo, 2005). Africa not only has the highest concentration of languages within a single continent, but most of the residents of Africa are capable of speaking more than one language (Batibo, 2005).

There are different degrees of bilingualism, based on the proficiency of the individual and his/her capability of communicating fluently in the spoken languages (Mackey, 2005). It is very rare to find someone who is capable of speaking more than one language with equal proficiency (Adler, 1977). Another factor in bilingualism is that bilingual individuals tend to use their different languages in different contexts (Batibo, 2005) and for different purposes (Wei, 2005).

This study specifically concentrated on usability of a word processor application for bilingual individuals. Therefore, it was important to first establish how the mind of a bilingual individual processes and stores the spoken languages to determine whether a bilingual individual is disadvantaged by use of an English interface.

2.8.1 Language storage and processing

Early on in bilingual studies, debates raged on whether the spoken languages were all stored in a single repository or whether each language had its own repository (French and Jacquet, 2004). A further subject of debate in bilingual matters was whether or not the second language (L2) was mediated through the first language (L1), a concept which is referred to as lexical mediation, or whether L2 words were connected directly to their meanings, which is known as concept mediation (Potter, So, Von Eckardt and Feldman., 1984).

2.8.1.1 Single and dual-code theory

In answer to the first question posed above, two models of linguistic representation were proposed, namely a single- and a dual-code model (Durgunoğlu and Roediger, 1987). The single-code model proposed that the bilingual individual represented words in a manner that transcended language, that is, the word representation was independent of language (Durgunoğlu and Roediger, 1987). Conversely, the dual-code model suggested that the two languages were stored separately and associative links were established between matching concepts in the two language bases (Durgunoğlu and Roediger, 1987). There was supporting evidence for both models (for example, Chen and Ng, 1989; Goggin and Wickens, 1971) and early research was unable to provide unequivocal evidence proving only one theory (French and Jacquet, 2004).

Thereafter it was established that both models were accurate, but at different representational levels. This was evident in a study undertaken by Durgunoğlu and Roediger (1987), who predicted and proved that results of data-driven tasks support the dual-code model and results of conceptually-driven tasks support the single-code model. Therefore, they concluded that whether there was a single- or dual-model of language representation was indeterminable, since evidence for each model was obtained by varying the retrieval demands of tasks (Durgunoğlu and Roediger, 1987).

2.8.1.2 Hierarchical memory models

It eventually became clear that bilingualism had to be interpreted as being task specific which led to a tentative agreement that the bilingual memory had to at least be separated between a conceptual and lexical level (French and Jacquet, 2004). The conceptual level is shared between the spoken languages while the lexical stores are specific to each language (French and Jacquet, 2004). Support for both the separation between conceptual and lexical stores and lexical stores for each language can be found in the literature (French and Jacquet, 2004).

By building on these findings the available hierarchical memory models were developed which take the division of conceptual and lexical stores into account as well as the supposition that each spoken language has a separate lexical store (French and Jacquet, 2004). Links which exist between the conceptual store and a lexical store are known as conceptual links and provide access to meaning for the language in the lexical store. Links that are established between two lexical stores are referred to as lexical links. The available hierarchical models are the word association model, the conceptual mediation model, the mixed model and the Revised Hierarchical Model (RHM).

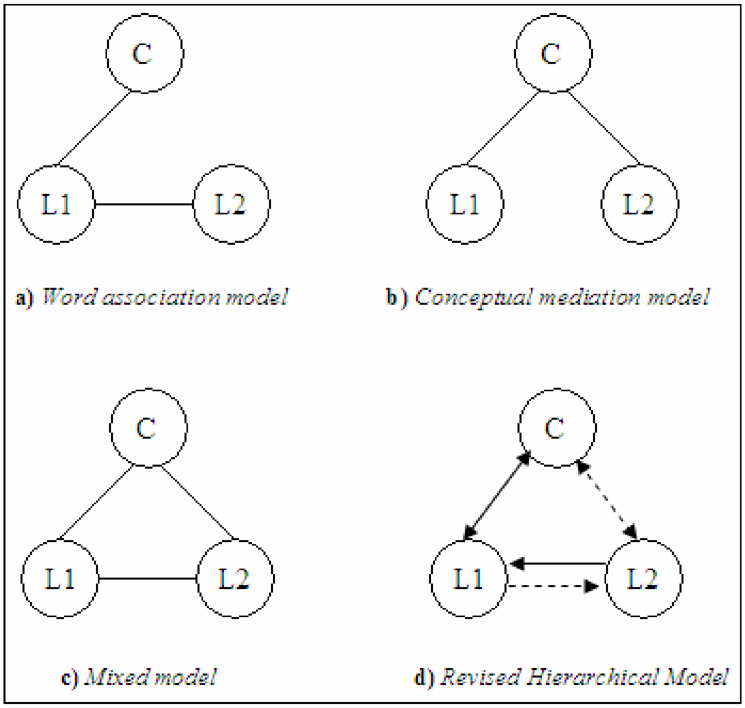


Figure 2.1: Hierarchical models for language representation

Source: French and Jacquet (2004)

The word association model (Figure 2.1a) has a conceptual link only for L1 and a lexical link between the L1 and the L2 stores (Potter et al., 1984; French and Jacquet, 2004). Therefore, this model proposes that L2 is only capable of accessing concepts via use of L1 (Potter et al., 1984).

However, since bilingualism occurs in varying degrees of fluency (Mackey, 2005), the conceptual mediation model (Figure 2.1b) is more fitting for an individual who has gained increased proficiency in L2 (French and Jacquet, 2004). This model provides for a conceptual link between L1 as well as L2 and the conceptual store (Potter et al., 1984). There is no lexical link between L1 and L2 (French and Jacquet, 2004), allowing for direct access to semantic meaning from L2 without mediation through L1 (Potter et al., 1984).

Subsequently it was established that the lexical links between the two languages do not disappear with increased proficiency of L2, hence the need for the mixed model (Figure 2.1c) (French and Jacquet, 2004). This model takes the residual effect of the lexical links into account, as well as providing for conceptual links between both L1 and L2 with the conceptual store (French and Jacquet, 2004).

However, performance from L1 to L2 is not the same as from L2 to L1, therefore the lexical links between L1 and L2 must be weighted (Dufour and Kroll, 1995; French and Jacquet, 2004). The initial dependence of L2 on L1 during L2 language acquisition means that L2 words are related to L1 words by the individual during the learning process, but there is a lack of L1 dependence on L2 (Kroll and Stewart, 1994). Therefore the RHM (Figure 2.1d) was developed which essentially combined the word association and conceptual mediation models into a single model, while still making provision for the initial reliance of L2 on L1. The conceptual links of L1 are also much stronger than those for L2 since initially the bilingual relies on L1 to gain access to meaning of L2 – therefore L1 initially serves as a lexical intermediary between L2 and the conceptual meaning (Dufour and Kroll, 1995; Kroll and Stewart, 1994). As L2 proficiency increases the conceptual set which can be activated by L2 grows larger, resulting in a reduction of reliance on L1 (Figure 2.2). This allows the bilingual individual to access meaning through direct use of L2 (Dufour and Kroll, 1995). There does, however, remain an unbalanced quality to these connections (Dufour and Kroll, 1995), which leads to the conclusion that semantic processing of an L2 message requires more cognition than that of an L1 message. The L1 lexical store is also larger than the L2 store, even for fluent bilinguals, since research indicates that there is a residual imbalance between the L1 and L2 lexical storage (Dufour and Kroll, 1995). Empirical studies have been conducted which support the theory behind the RHM (for example, Kroll and Stewart, 1994; Mitchel, 2005; Sholl, Sankaranarayanan and Kroll, 1995). Acceptance of the RHM leads to the conclusion that semantic processing of an L2 message requires more cognition than that of an L1 message. This suggestion has been supported by empirical testing which has found that processing of L1 stimuli is easier than processing of equivalent L2 stimuli (Sholl et al., 1995).

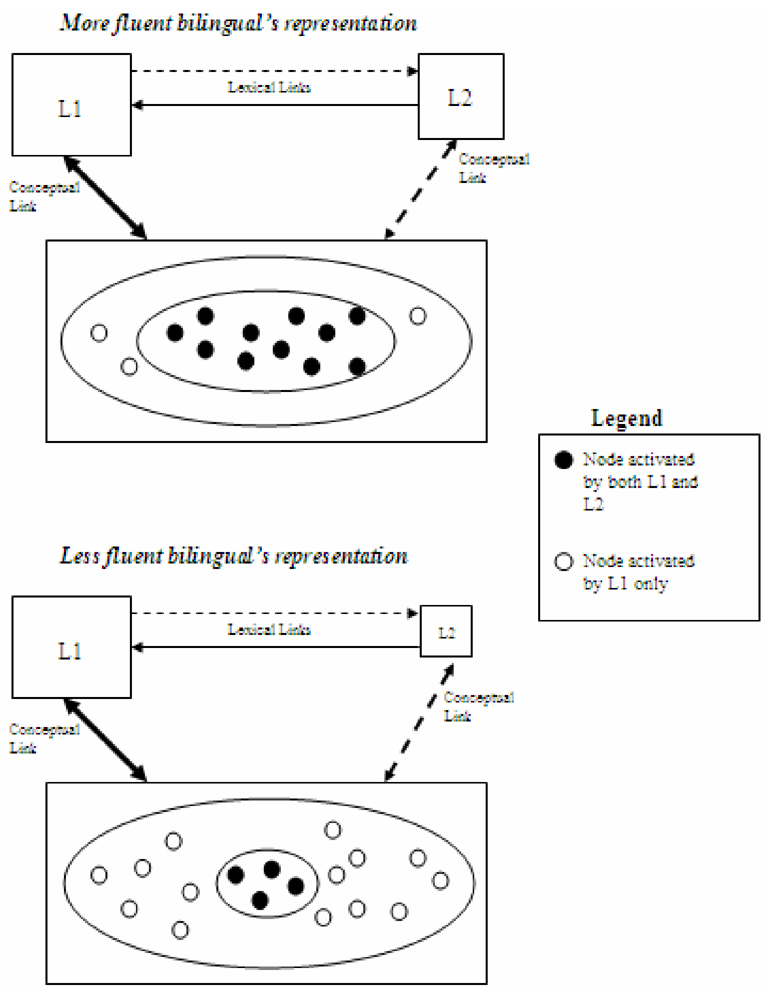


Figure 2.2: Conceptual representation for bilinguals at different levels of L2 proficiency (Source: Dufour and Kroll, 1995)

The hierarchical models all indicate that processing of L2 stimuli requires more cognition and effort than processing of L1 stimuli. A certain amount of interference is also present when using one language or accessing knowledge thereof, while simultaneously speaking or writing in another language (Mackey, 2005), which could cause difficulties for non-English speaking users when having to make use of an application that is entirely in English. From these two aforementioned points it could be deduced that L2 users of a computer interface could experience difficulty during interaction with an application. This deduction led to one of the key elements of this research study, which was to determine whether or not there was a need for translation of the interface into the L1 of bilingual word processor users. Examination of the available literature on processing and storage seems to provide enough motivation for translation of the interface into the L1 of all end users. However, the presence of different degrees of bilingualism, the lack of available terminology in some languages and the fact that prolonged exposure to L2 could negate the drawbacks associated with use of an interface in the L2 of the user, all provided ample motivation for the need to investigate the issue further. Language, therefore, played an important role in this study in terms of whether an English interface should also be provided in the L1 of a bilingual user and whether this translation would increase the usability of a word processor application.

2.9 Interface translation

At the forefront of catering for other markets is the question of whether or not to translate interfaces into the language of the end user. As discussed in the previous section, second language users may have a distinct disadvantage from the outset when using a product that is in English due to the nature of the processing of the second language stimuli. Apart from that, users of an interface that is not in their first language do encounter a number of inherent problems, one of which is verbal context – where surrounding words serve to place a word in context, allowing users to identify the actual meaning of the word rather than the potential meaning thereof (Kukulska-Hulme, 2000). Many interfaces do not include verbal context in menus, toolbars or buttons, which is clearly disadvantageous to second language users (Kukulska-Hulme, 2000) and could also lead to difficulty for novice or first-time users who are unfamiliar with the domain terminology and concepts.

Consider for example, the difference between American and British English where differences such as the use of a 'z' instead of an 's' (summari**z**e in contrast to summari**s**e) exist between the two languages. Greater problems can be experienced with the use of different words referring to the same thing, such as gas, petrol and fuel, all of which could potentially refer to the same thing but are used differently in different countries. Another example is the difference between what is meant by a public school in America and in Britain. Differences such as these could cause misunderstandings and resentment towards the application or website and promote the case for translation into locale-specific English. However, if you are going to make provision for different English dialects, how can the large group of users still be expected to interact in a language which is not their L1 and in which they may even not be fluent?

However, translation of an interface into the first language of the target market remains a fairly contentious issue, with many proponents for (for example, Battle and Degler, 2001; Bodley, 1993; Esselink, 1998; Fowler and Stanwick, 1995; Nielsen, 1999), and detractors from the issue who advise caution when considering translation as not all users show a preference for carrying out tasks in their mother tongue (for example, Blignaut and McDonald, 2006; De Wet, Blignaut and Burger, 2002) and performance is often hampered by translation (De Wet et al., 2002).

Apart from this, translation of an interface also presents a number of challenges. Firstly, translation may not be feasible due to technical (Kukulska-Hulme, 2000; Stander, 1998) or economic reasons (Kukulska-Hulme, 2000), as it remains a time-consuming and costly venture (Hars, 1996). Another consideration that must always be taken into account is to ensure that the translation of the interface is done correctly and that the intended meaning is indeed conveyed to the user (Ott, 1999). Not all translators are proficient in technological terminology and if the translation is undertaken by a language professional who is unfamiliar with the terminology and the underlying meaning of the words, the translation could be incorrect and not succeed in conveying the correct message to the user (Ott, 1999).

2.9.1 International translation studies

Since interaction with a computer in a person's mother tongue facilitates easier and faster learning (Griffiths, Heppel, Millwood and Mladenova, 1994), language can have an enormous effect on the usability of a website (Singh, 2002) and, by extension, on a software application. Evidence exists for user preference for translation of websites into the language of the target market, with 75% of Chinese and Korean users preferring to use websites that have been translated into their language (Ferranti, 1999), and a large number of French and Spanish users also prefer local language websites (Lynch, Kent and Srinivasan, 2001).

The impact of language on search behaviour appears to be hardest felt for L2 users who are unfamiliar with the domain knowledge, while L2 users who have high domain knowledge and L1 users behave in a similar fashion (Kralisch and Berendt, 2005). Collective findings from four different usability tests indicated that second language users made more mistakes than first language users during completion of preset tasks but that there was no significant difference in task completion time detected between first and second language speakers (Kang, 2003). Possible motivation for these findings was corroborated by the fact that second language users appeared to select items and menu options at random rather than reading through all available options like the first language users did (Kang, 2003). Therefore, the efficiency of L2 users appears to suffer through an increased error rate and random selection of options in an attempt to achieve the desired effect. User performance appeared to be dependent on the degree of domain knowledge of the L2 users, therefore it could also be inferred that the random selection of options by L2 users would decrease as the domain knowledge of the L2 users increases.

From an international perspective it would seem as though there is a great need for translation to accommodate users of different language groups, particularly for users with a low degree of bilingualism or low domain knowledge. The South African perspective is, however, not necessarily the same as the international one.

2.9.2 South African translation studies

Multilingualism within a country presents a problem because it requires a decision to made as to which language must be used (Keniston, 1997). With 11 official languages in South Africa, no single dominant culture exists (Hugo, 2002) and developers may be hard-pressed to find a single language acceptable to all citizens. Besides, the solution of a single application to suit all users will not be sufficient in the diverse South African climate and will result in the exclusion of some sectors from the benefits and exposure to information technology (Hugo, 2002). Therefore, it is imperative that we look beyond tradition, convention and imitation in the development of software interfaces (Hugo, 2000).

2.9.2.1 Website translation

In a study conducted at a South African university the responses of 132 students to a questionnaire were analysed in an attempt to determine whether or not there is a need for localisation of South African websites (Masoeu and De Villiers, 2001). The overwhelming (82.82%) language preference was for English websites although respondents did feel that it would be considerate and respectful to other language groups to translate websites (Masoeu and De Villiers, 2001). One drawback to this study is the fact that more than half (65%) of the respondents in this study were Afrikaans speaking, with English speaking respondents being the second largest group at 12% of the total respondents. This unequal language representation amongst the users could influence the results, which will then not be truly representative of all languages. Further hampering the interpretation of the results is the fact they were not presented according to home language, which makes it difficult to determine precisely what the impact of the study was and whether the preference for English websites was found in all language groups.

A more representative study provided a contradiction to the findings of Masoeu and De Villiers (2001) by finding that there is a strong preference by non-English first language speakers for websites to be provided in their home language, particularly by Afrikaans speaking users (Van Belle et al., 2003). In a partial confirmation of these findings Blignaut and McDonald (2006) found that Afrikaans speaking users prefer having a website available in their first language while users of other non-English languages, including Sesotho, Setswana and isiXhosa, preferred using a website in English to a website in their own first language. This phenomenon was found in three diverse groups of users which included high school students, second year university students and a group of labourers (Blignaut and McDonald, 2006). Although these studies indicate contrasting views on whether or not websites should be made available in languages other than English for non-English speaking users, it is important to bear in mind that the studies were undertaken in different regions in South Africa, and that this could have had an influence on the results. To illustrate this point, the largest language group in the former study, i.e. Van Belle et al. (2003), was that of isiXhosa first language speakers, while the latter study, i.e. Blignaut and McDonald (2006), had a larger representation of Sesotho speakers. It is important to bear in mind that user preference does not always reflect user performance. The fact that users may prefer having websites in English does not mean that their performance will not be enhanced by the translation of the website text into their first language.

A number of studies have concentrated on user performance as opposed to simply gauging user preference. In one such study a website was designed specifically for the study and translated into Afrikaans and Sesotho. Users were then required to answer a 26-question comprehension test based on the content of the web site (De Wet et al., 2002). Afrikaans speaking users fared better on the Afrikaans website, while Sesotho speaking users not only preferred to interact with the website in English, they also had increased performance on the English website as opposed to the one in their first language (De Wet et al., 2002). Several reasons were provided for the poorer performance of Sesotho speaking users on the Sesotho website. Firstly, formal education had been received in English up to that stage. Secondly, much of the technical terminology used on the website had no Sesotho equivalent and Sesotho speakers had a hard time understanding the website, saying that it was not the same Sesotho dialect that they spoke (De Wet et al., 2002). A second study provided contradictory results to the study of De Wet et al. (2002). When comparing user interaction with a website, where users were required to complete a number of tasks that had been identified as representative tasks by means of analysis of the website log, no difference in performance was detected between users of different languages (McDonald and Blignaut, 2005). The website used in this study was only available in Afrikaans and English.

The difference between the studies of De Wet et al. (2002) and McDonald and Blignaut (2005) could be due to the fact that most of the participants in the second study had previous exposure to the website used in this study and as such could have already learned the structure of the website. In contrast, in De Wet et al. (2002), no participants had previously used the website in the study as it had been designed specifically with that study in mind. These studies indicate that Afrikaans speaking users prefer using a system in Afrikaans and that they also perform better on an Afrikaans system, while translation into languages such as Sesotho should be carefully considered due to user preference and performance. However, in the long run, it seems as though translation is not necessary since after prolonged use of the system, different language groups were able to use the system with equal effectiveness and efficiency, which supports the conclusions drawn from comparison of the international translation studies. The lack of available terminology in some languages also means that translated interfaces should also be verified by potential users to ensure that the dialect and translation are understandable to these users.

The listed South African studies did, however, focus on the translation of websites which generally contain more text than traditional interfaces and much more so than the single commands found in a word processing environment. They do nevertheless provide insight into the preferences and performance of different language users. Studies which concentrated on translation of software applications will be examined next.

2.9.2.2 Software application translation

Bodley (1993) undertook a study in a car manufacturing plant in order to determine whether what were defined as third-world users required an interface in their home language. The profile of the users in the study was very specific and third-world users were defined as being isiXhosa-speaking individuals who had only a Grade 8 education level or lower. Many of the users were forced to interact with an English interface even though they were not literate in English. The resultant interface developed in isiXhosa increased the effectiveness of the users. Therefore, Bodley (1993) recommends the translation of the user interface into the home language of the user. However, the limitations presented by the specific definition of the third-world users and the fact that many of the users were not fluent in English does cause concern for the applicability of the study to the wider community of South African computer users. Considering the fact that the majority of the participants had no working knowledge of the English language, it comes as no surprise that the respondents indicated a preference for interfaces in their L1 and that the interface enhanced the effectiveness of the user. Therefore, this study cannot be generalised to a greater audience and is certainly not indicative of user performance or user preference for bilingual users.

Stander (1997) investigated the effect of culture on the usability of a simple text-based interface. The first study was conducted with a sample of 277 students who completed a simple task on what was referred to as a "simulated interface" (Stander, 1997). A significant difference was found between the performances of the different culture groups for the simulated study (Stander, 1997). A second, "real-life", study was conducted using over 2000 students from a South African university who registered for the academic year by means of computer terminals positioned on the grounds of the university campus (Stander, 1997). The time taken to complete the registration was measured and no significant difference was detected between the performances of members of this group (Stander, 1997). These contradictory findings were explained by the fact that the real-life study used students who were fluent in the English language and also by the fact that some of the language groups consisted of only a small number of students (Stander, 1997). The fact that differences were detected in the simulated study where users encountered the system for the first time indicates that culture affects initial interaction with a system. However, bilingual users did not struggle to use an interface in English and this is the result that impacts on the current study as bilingual users will also be tested. Therefore, this study could prove useful as a comparison to the current study. Even then, only the "real-life" study and not the "simulated" study should be examined as it is the "real-life" study which was conducted with actual bilingual students who completed set tasks on a university registration application.

More recently, in a study undertaken in rural South Africa, in the province of KwaZulu-Natal, a number of interesting findings were made regarding the language of an interface. Users were exposed to English interfaces for products such as Microsoft Word and Excel, while an open source accounting package was adapted to use an isiZulu (the majority language of the area) interface (Heukelman, 2006). The greater majority (80%) of the users agreed that having an interface in their home language increased their understanding and the memorability of the interface (Heukelman, 2006). Nearly 70% of the users felt that they would be able to transfer their knowledge to another interface but more than 60% of the respondents indicated that they would have preferred having the interface of the accounting package in English (Heukelman, 2006). When asked what they would like to change on the Excel interface only two respondents, who required an interpreter, answered that the interface should be in isiZulu (Heukelman, 2006). The researcher points out that all other software packages used in the study were in English and users became accustomed to the terminology of these packages and might have resented having to adjust to the isiZulu terminology (Heukelman, 2006). Also, a general complaint from the test subjects was that the isiZulu translations were longer than the English originals which made the menu items longer and more cumbersome (Heukelman, 2006). Both of these points could have caused the apparent disinterest in an isiZulu interface and they should be kept in mind when introducing new languages into an interface.

Most of the users in this study were also bilingual and when viewed in this context it would indicate that bilingual users see no need for translated interfaces. However, a great cause for concern in this study is the fact that the users were exposed to a number of English interfaces and only a single isiZulu interface. When exposing users to a number of common applications, all only available in English and then making a single application available in isiZulu, it comes as no surprise that users will resent having to switch between the two. Since the majority of the applications were in English it is understandable that users would become accustomed to the English interface and have mastered the English terminology, then when suddenly confronted with an interface in isiZulu. It is to be expected that they would not respond well to the added effort that adjustment to the new terminology demanded. In such cases, translation would do more harm than good. Long-term exposure to a website allowed users of different languages to perform equally well with a number of set tasks. Therefore, it could be expected that long-term exposure to an application in a certain language may cloud the preference of the user simply because they have become accustomed to using the product in a foreign language and do not desire an adjustment period to an application in a different language, even if that different language is their first language.

Further afield, in responding to a questionnaire, users in Botswana indicated high desirability for a localised interface, an opinion which was contradicted by the finding that the same users responded poorly to localised icons (Onibere, Morgan, Busan and Mpoeleng, 2001). However, only 25% of these respondents wanted commands to be in Setswana, which, together with English, is the official language of Botswana

(Onibere et al., 2001). These findings corroborate those of Blignaut and McDonald. (2006), De Wet et al. (2002) and Heukelman (2006) who all found no inclination by bilingual users to have an interface translated into their home language. They also show that in such a case preference may reflect the measured performance of users, since bilingual users have indicated that there is no need for a translated interface (Onibere et al., 2006) and that the performance of these users is also not hampered by use of a strictly English interface (Stander, 1997).

2.9.2.3 Applicability to study

In summary, many of the studies made use of websites which required large amounts of reading text and when viewing the results specifically in terms of bilingual users, results were mixed. Results seem to indicate improved performance for Afrikaans users when using an Afrikaans website, while other non-English speaking users fared better on an English website when the content and structure of the website was unknown. Should the website be one which is known and used on a regular basis by users, then performance was not affected by the language of the website. Prolonged use and exposure to the website reduced the initial negative impact of interaction in the L2 of the user. Regarding preference of non-English speaking users, the overwhelming evidence seems to be in favour of provision of websites only in English.

When viewed strictly in terms of bilingualism, user performance was neither positively nor negatively impacted by translation of the software application interface, while continued usage of website in the L2 of the user lessened the negative influence of an interface in English for a non-English website user. Users of a Sesotho website expressed concern that the Sesotho translation appeared to be in a different dialect than the one in which they generally conversed, while isiZulu users of software applications were unhappy at the length of isiZulu menu items as this increased the amount of reading that was required of them.

Since translation is a contentious issue with studies offering different results concerning both preference and performance, this offers an ideal topic for further investigation, i.e. whether software applications require translation to achieve further saturation in the global market. Very few studies were found within the South African context which could serve as comparison studies for this study. Many of the studies made use of a website requiring a fair amount of reading. Even so, these studies did indicate that the translation of an interface into the L1 of a bilingual user did not always add to the usability of the website. The only project which studied the translation of software for bilingual users using a methodology that could be considered similar to the study at hand (Stander, 1997), indicated no need for translation into the L1 of a bilingual user. These findings were contrary to what would be expected due to a bilingual individual's unequal processing capabilities of L1 and L2 stimuli. The aforementioned study was undertaken in the Western Cape province of South Africa, which has a very different language demographic to that found in the Free State province, where the current study was based. Therefore, it is important to determine whether the same results could be replicated in a study with users of a different language so that the results could be generalised even further.

The expectations of this study were that translation would influence usability differently depending on the degree of bilingualism of the users and on their degree of exposure to the application terminology. Users who were fluent bilinguals should not be influenced by the language of the interface. Similarly, users who had had extended periods of interaction and exposure to the application in a language would have become accustomed to the terminology of the application in that language and therefore their performance should not have been affected by the translation of the interface into their L1. However, users with limited experience and in particular, those who had had no previous experience with the interface and task domain, should have benefited from translation into their first language. Translation of the word processor application was achieved by making the menus and tooltips available in English, Afrikaans and Sesotho, which were the dominant languages of the region in which the study was completed.

2.10 Verbal versus nonverbal processing

The span of immediate memory limits the amount of information that an individual is able to receive, process and remember (Miller, 1956). Information is received, processed and then stored in short-term memory, after which it is transferred to long-term memory (Norman and Bobrow, 1976). When information needs to be recalled, it is retrieved from long-term memory and placed in short-term memory (Kent, 1981). The number of chunks that can be held in short-term memory for immediate recall was estimated at 7±2 (Miller, 1956). The way in which this information is stored and processed differs between verbal and nonverbal information (Paivio, 1971). Much research has been done into the difference between how images and text are processed, and it has been found that images are processed at a higher speed than text, even though it appears as though the processing of images and text draw on a common knowledge base (Hogaboam and Pelligrino, 1978).

Dual-coding theory offers an explanation for this unequal processing of pictures and text. The theory is concerned with cognition and was originally aimed at describing the differences between verbal and nonverbal memory assimilation and processing (Sadoski and Paivio, 2004). Since then it has been extended to include other areas of cognition (Sadoski and Paivio, 2004). Dual-coding theory is based on the idea that information is stored in memory based on the format and qualities of the external experience or stimulus from which they were obtained (Sadoski and Paivio, 2004). These stimuli can be either verbal or nonverbal (Sadoski and Paivio, 2004). The difference between the verbal and nonverbal stimuli results in separate mental systems for the processing and representation of the verbal and nonverbal events (Sadoski and Paivio, 2004). The nonverbal system is often referred to as the imagery code due to the fact that it is primarily focused on the generation, analysis and transformation of mental images (Sadoski and Paivio, 2004).

Generally, the difference can be explained in broad terms as the difference between processing in the left and right hemispheres. While the left hemisphere is responsible for the processing of language, which is done in a linear fashion, the right hemisphere uses parallel processing to absorb visual information (Paivio, 1971). The verbal system is specialised for processing in a sequential manner, while visual imagery, or the nonverbal system, is geared towards processing in parallel (Paivio, 1971). The receptors of the visual system are capable of processing, transmitting and receiving information simultaneously in a spatial array, thus it is said to be capable of parallel processing (Paivio, 1971). In contrast, since the verbal system is linked to the auditory system and the auditory system is only capable of dealing with temporally organised stimuli patterns, the verbal system must be classified as a sequential processing system (Paivio, 1971). Thus, images can be processed at a much higher speed than what text or language can.

More specifically the difference in processing can be traced back to the building blocks of verbal and nonverbal stimuli. The building blocks of the verbal and nonverbal systems are known as logogens and imagens respectively (Sadoski and Paivio, 2004). A logogen is “anything learned as a unit of language”, while an imagen corresponds to a visual stimulus and tends to be perceived as nested sets (Sadoski and Paivio, 2004). For example, a baseball bat can be visualised on its own or as it is held in the hand of a baseball player, for instance (Sadoski and Paivio, 2004). Logogens and consequently the verbal system are restricted to sequential representation (Paivio, 1971; Sadoski and Paivio, 2004), the units of which can be combined into sequences at all levels (Sadoski and Paivio, 2004). In other words, certain letters can be reorganised to form certain words, or two words can be placed next to one another to form another concept, for example, baseball bat (Sadoski and Paivio, 2004). The verbal system forms a hierarchy such that smaller units can be combined into larger units and larger units can sequentially be deconstructed into smaller units (Sadoski and Paivio, 2004). On the other hand, the nonverbal system is represented in a more continuous and integrated manner (Paivio, 1971; Sadoski and Paivio, 2004) and cannot easily be decomposed into smaller isolated elements (Sadoski and Paivio, 2004). A hierarchy still exists for the nonverbal system in terms of the nested sets described earlier which can be built and synthesised into ever larger sets (Sadoski and Paivio, 2004). In turn, these larger sets can also be broken down into smaller sets, but not to the extent that the resulting units are comparable to letters, words and phonemes that are the basic units of the verbal system (Sadoski and Paivio, 2004).

Verbal and nonverbal stimuli are processed differently once they are perceived by the sensory systems (Figure 2.3) (Sadoski and Paivio, 2004). Essentially the verbal system is structured according to a strict hierarchical, sequenced arrangement of logogens, while the nonverbal system is represented by imagens which can be nested within other imagens, or overlap with other imagens, or exist as stand-alone imagens which are as yet not associated with any other set (Sadoski and Paivio, 2004). The referential connections between the verbal and nonverbal illustrate the concept that verbal processing can stimulate and activate imagens, which can then in turn, refer back to the verbal system (Sadoski and Paivio, 2004).

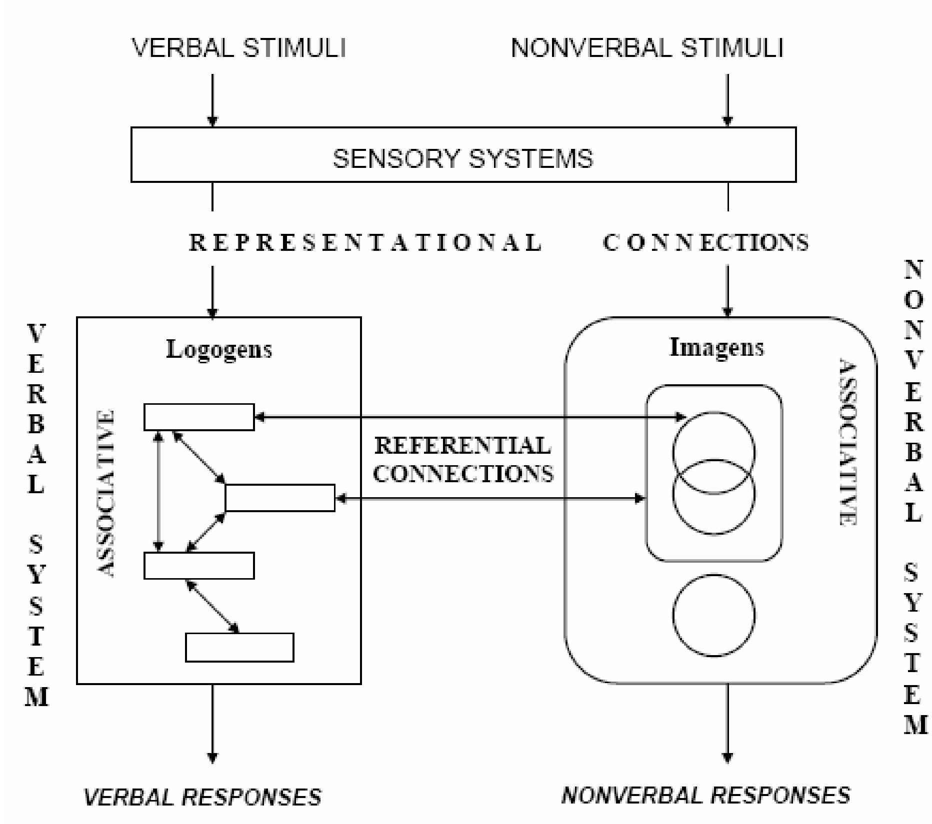


Figure 2.3: Dual-coding theory

Source: Sadoski and Paivio (2004)

There are of course, detractors of the dual-code theory (for example, Kintsch, 1974; Norman and Rumelhart, 1975) who are in favour of a memory representation where all information, whether images or words, is stored abstractly and not linked to any type of modality. However, Kintsch (1977) states that there is overwhelming evidence for the dual-code theory as proposed by Paivio (1971), but that it may in fact have to be extended to a multi-code theory to include such modalities as acoustical memory codes.

A number of studies have highlighted the obvious superiority of picture processing over verbal processing in a number of different aspects. Where subjects were required only to remember items, but not the order in which they appeared, pictures proved to be superior memory aids (Paivio, 1971). Nonverbal imagery or a combination of nonverbal and verbal increases the memorability of items, and pictures also have a higher recognition rate than words, thus images have greater use as mnemonic aids (Paivio, 1971). Recall of verbal stimulus decayed after a short while, which was not the case with image reporting on a read text (Sadoski, Goetz, Olivarez, Lee, and Roberts, 1990). Studies using school-going children (Gambrell and Jawitz, 1993; Pressley, 1976) and high school students (Anderson and Kulhavy, 1972) have shown superiority of image processing. When instructed to read text, some subjects were instructed to form images and others not. In all of the studies, the group which was explicitly instructed to form mental images of the text outperformed the group which was instructed to remember the text by any means possible. Recall (Gambrell and Jawitz, 1993; Pressley, 1976), factual accuracy (Gambrell, 1982) and comprehension (Anderson and Kulhavy, 1972) were all increased with the use of images.

These findings are by no means a complete foundation or support for the dual-code theory but they do, together with the salient point made by Kintsch (1977) regarding the existence of overwhelming evidence for support of the dual-code theory, and the fact that image processing, recall and recognition is superior to that of text processing, lead to the acceptance of the dual-code theory for the purposes of this study.

The difference between the processing of images and texts is pertinent to the current research study since one of the designed icon sets consisted only of a text representation of the function and contained no graphical component. Since images are processed at a higher speed than text, it could be assumed that the pictorial icons would promote a higher degree of efficiency than the text buttons. However, the advantages associated with the use of icons are all just as applicable to the use of text buttons and could provide sufficient countermeasures to neutralise the negative effects of verbal processing when comparing with nonverbal processing. Furthermore, the text buttons would not require any interpretation of meaning of an image, such as would be required with the pictorial icons. Therefore, there are both positives and negatives to replacing the pictorial icons with text buttons and the effect on usability could not be predicted and subsequently had to be tested to provide empirical evidence of the superiority of one icon type over the other.

2.11 Menus and icons

Humans are better equipped for and better at recognition than recall (Galitz, 2002; Redmond-Pyle and Moore, 1995). Over the years the design of interfaces has shifted focus to exploit the superiority of recognition over recall by including menus and icons in the interface design (Preece et al., 1994). Because of this, users simply have to scan through the available options, recognise the required option and then select it, which eliminates the need for memorisation of complex syntax of a command-based system (Jones, 1989; Marcus, Smilonich and Thompson, 1995). This reduces the cognitive load on the memory of the user, since no recall is required by the user (Rogers, 1989b) and it reduces the number of syntactic errors made by the user (Gittins, 1986).

Available commands are always visible to the user (Galitz, 2002; Lodding, 1983) by means of a toolbar which is normally situated along the top of the window (Fowler and Stanwick, 1995). Toolbar icons are immediately visible to the user, where the correct menu option is embedded in a menu structure and is hidden from view (Ellis, Tran, Roo and Shneiderman, 1995). Although menus also provide a means for users to select a command from a set of listed commands (Arend, Muthig and Wandmacher, 1987), menus still have to be read and image processing is considered to be faster than textual processing (Paivio, 1971). However, in order to use icons, users must remove their hands from the keyboard to grasp the mouse, whereas menus can be accessed and navigated by using the keyboard (Ellis et al., 1995). When faced with the decision of whether to use a menu or an icon, users tend to use the icon rather than attempt to navigate through the menu structure in search of the correct menu item (Bickford, 1997).

Menus and icons were two of the interface components whose influence on the overall usability of a word processor application was investigated during this study.

2.11.1 Menus

Menus list commands, attributes or states for the user to choose from (Apple, 2006). They provide an easy way for users to browse through all the available options

(Galitz, 2002) and as mentioned previously provide a way of lightening the cognitive load of the user (Jones, 1989; Marcus et al., 1995; Rogers, 1989b).

Menus do, however, have a number of disadvantages. Firstly, another menu must often first be searched for and selected before the required menu is displayed (Galitz, 2002). Therefore, care must be taken when designing menus to ensure that the resultant menu is not too deeply nested and therefore difficult to use because users have to navigate through successive layers of menu options in order to find the desired menu option (Fowler and Stanwick, 1995). Secondly, users must look away from the work area to use the menu and thirdly, menu items are smaller than full-size buttons, which significantly slows selection time (Galitz, 2002). Even though menus also provide an itemised listing of all available functions, applications using text menus require some memorisation of terminology, whereas iconic options require users to only recognise an image (Lodding, 1983). Lastly, menus have a tendency to slow down expert users (Fowler and Stanwick, 1995).

Menus may be graphical, textual or both (Marcus et al., 1995). A graphical menu is one which uses images or icons to display options (Marcus et al., 1995).

Menus were included in some of the interface configurations to determine if their inclusion could increase the effectiveness and efficiency of the word processor. Menus were provided in English, Afrikaans and Sesotho and thus provided a means to determine whether usability is influenced by translation of the interface.

2.11.2 Icons

A sign is "something which stands to somebody for something in some respect or capacity (*sic*)" (Peirce, 1985). A sign is either a symbol, an icon or an index (Peirce, 1985).

The advent of the graphical user interface (GUI) resulted in an increase in the use of icons within computer applications (Kacmar and Carey, 1991).

An icon is defined as being (a) an image, picture or representation or (b) a symbol which can be associated with the real-world object that it represents (Collins English dictionary, 2000) simply by means of its own characteristics (Peirce, 985). In other words, an icon is a representation of a concept (Fitrianie and Rothkrantz, 2005). In terms of computer usage, icons are common interface components that employ images (Preece et al., 1994) to represent an object or an action that can be carried out by the user (Benbasat and Todd, 1993; Jones, 1989). They represent features that are available in the software (Barr, Noble and Biddle, 2002b) in a form which corresponds to the real world (Rogers, 1989a). Therefore, icons in a computer user interface are used to visually display information (Blattner, Sumikawa and Greenberg, 1989). Icons are used as a type of "visual shorthand" that allows designers to convey a complex idea or concept as a single image (Bickford, 1997). Icons have replaced commands and menus as the format by which communication is achieved with the user (Gittins, 1986). Icons are usually displayed in a toolbar which is a group of buttons that allow for easy and quick access to a number of frequently used options from the menu (Marcus et al., 1995).

Semiotics is the theory and study of signs and symbols, particularly when employed within some form of communication, such as language (Pelc, Sebeok, Stankiewicz and Winner, 1984). It is the study of what constitutes signs and the laws that govern them (de Saussure, 1985) and it is therefore considered to be the formal doctrine of signs (Peirce, 1985). In the computer-rich world of today, semiotic analysis can be extended to include the interpretation of GUIs which tend to make liberal use of icons.

The fact that images are present all around us in everyday life indicates that the modern day user has already become proficient in semiotic analysis since these images need to be interpreted in order to function in modern society (Hodge and Kress, 1988; Luke, 1996). This fact serves as motivation to create an image-rich user interface for computer applications. Since the user is accustomed to interpretation of images it should simplify the use of a graphical interface, with the added advantage that the use of metaphor brings by simply extending the simplicity of the interpretation of the icon (Gittins, 1986).

Strictly speaking icons do not employ the use of metaphor (Barr, Biddle and Noble, 2002a), and it is generally the entire interface which is metaphoric, for example, the desktop metaphor. Rather, icons make use of metonymy (Barr et al., 2002a) which is the "use of one entity to refer to another that is related to it" (Lakoff and Johnson, 1980). Icons thus refer to objects within the realm of the interface, which is usually metaphoric; therefore icons refer to a metaphor (Barr et al., 2002a). For example, the document icon refers to the metaphoric document which is used to store data (Barr et al., 2002a).

Since pictorial communication predates communication via textual means, the naturalness associated with pictorial communication has motivated the need for iconic interfaces, particularly in view of the increase in the number of computer users (Lodding, 1983). User performance may be influenced by the icons used in the interface (Piamonte, Ohlsson and Abeysekera, 1999). Apart from this fact, icons have a number of advantages and disadvantages that come into play in human-computer interaction.

2.11.3 Advantages of icons

The popularity of the icon stems from the belief that they reduce the complexity of the system while increasing the ease of use and learnability (Preece et al., 1994; Schild, Power and Karnaugh, 1980) and reducing the number of errors made by the user (Lodding, 1983) thereby increasing the overall productivity of the user (Kacmar and Carey, 1991). This is primarily due to the fact that they can successfully be used to communicate more informatively than other interaction methods while using the same amount or, in some cases, less screen space (Gittins, 1986). Even when used in combination with text within a menu structure there is a marked improvement in user accuracy (Muter and Mayson, 1986) due to the fact that images are more readily recognised and can be recalled faster than text (Shepard, 1967). Even though the user must still recognise the required image, this processing is easier since the user is still able to pick the correct icon even when they are unable to fully reproduce the image from memory (Lodding, 1983).

A number of individual icons can also be placed and grouped in such a manner that the group becomes a cohesive whole, and collectively corresponds to an object found in the real world (Gittins, 1986), thus allowing for easier interpretation by association of the individual icons to the whole set. Grouping of similar icons also allows the icon group to be processed as a single unit, therefore increasing the total number of icons which can be placed on the toolbar to stay within the 7±2 memory limit (Bickford, 1997).

One of the greatest possible advantages of images and a significant motivator for why iconic communication is such a great asset is the potential of images to communicate across language barriers (Lodding, 1983; Rogers, 1989b). For instance, a document remains a document regardless of the communication language (Rogers, 1989b), such as Afrikaans, Sesotho or French etc. This means that icons present the possibility of becoming a truly universal communication tool (Rogers and Oborne, 1987). However, this assertion does not hold in all cases (Lodding, 1983). Consider, for example, the image of a mailbox which is typically used in the United States of America but which is virtually unrecognisable in most other parts of the world (Figure 2.4A). As a further indication of the paramount importance of testing and adapting signs and icons for use in other countries is the discovery that illiterate isiZulu speaking adults interpret the common “no smoking” sign (Figure 2.4B) as meaning they are allowed to smoke half a cigarette (Andrews as cited in Murrell, nd). Furthermore, to them, the emergency exit sign (Figure 2.4C) means “do not run that way or you will get your hands and feet chopped off” (Andrews as cited in Murrell, nd) and in a South African context, a brown snake would be a better sign to indicate danger than the commonly used skull-and-crossbones image (Figure 2.4D) (Murrell, nd).

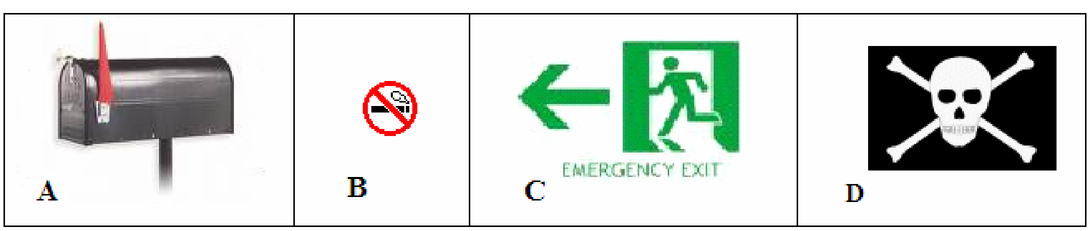


Figure 2.4: Signs and symbols open to possible misinterpretation

Repeated use of icons can also lead to standardisation of icons across applications (Hunt, 1996; Richards et al., 1994) which will lead to fewer user errors, while simultaneously increasing the speed with which a user is able to make a selection (Kacmar and Carey, 1991; Richards et al., 1994).

2.11.4 Disadvantages of icons

Apart from the obvious lack of efficiency when compared to other methods, the icon toolbar also has the disadvantage that the size display influences the distance that must be covered in order to select the relevant icon on the toolbar (Lane, Peres, Sándor, and Napier, 2005b). This causes an increase in completion time for larger display sizes which have become more popular over the years (Lane et al., 2005b).

Icons can be difficult to use and may require that much of the work be done by the user since interpretation could take a long time (Potosnak, 1988). Furthermore, although icons eliminate syntactic errors, image equivalents may not exist for all functions present in a system (Gittins, 1986; Lodding, 1983), which may force developers to design an icon from scratch, resulting in unknown icons or images being used in an interface and forcing users to have to learn the meaning of the icon (Galitz, 2002), which may severely reduce the usability of the product (Sen and Chen, 2002). Since icons are open to interpretation via semiotic analysis, this analysis is subject to success or failure (Barr et al., 2002b) which could lead to users making semantic errors during the selection process should the user select an incorrect icon (Gittins, 1986). Apart from this, unknown or badly designed icons create possibly the greatest disadvantage associated with iconic use – misinterpretation.

An image which seems obvious and self-explanatory to the designer is not always obvious to the user (Potosnak, 1988; Rogers, 1989b), resulting in possible misinterpretation by users if the chosen image invokes unintended or multiple associations (Benbasat and Todd, 1993; Lodding, 1983; Rogers, 1989b) – the picture that "speaks a thousand words may say a thousand different words to different viewers" (Zammit, 2000). Ambiguity is more prevalent in images than in text (Zammit, 2000), and would inevitably decrease the usability of the interface as users would be forced to resort to the documentation in order to interpret the intended meaning of the icons (Rogers, 1989b). In cases where multiple interpretations are possible, users may resort to a trial and error method, by selecting an icon and then determining whether the invoked action was the desired action (Rogers, 1989b). The misinterpretation of icons has been revealed in numerous studies (for example, Ferreira, Noble and Biddle, 2006). Misinterpretation is, however, unlikely to occur with words when used in a menu or text-based interface. The only misinterpretation that could possibly occur was if the word was interpreted out of context, a problem which could be solved by provision of more verbal context.

2.11.5 Iconic interfaces

Judging by the rapid growth and adoption of the graphical user interface, the advantages of icons seem to outweigh the disadvantages as continued use of icons has been attributed to the fact that they are easier for users to learn (Schild et al., 1980) and to use (Reingold, 1989).

Icon usage has been heralded as creating a more learnable and usable system by reducing the cognitive load of the user since images do not need to be memorised, they are easier to recognise and they can be interpreted easily by users (Section 2.11.3). However, there is a contrasting view to this standpoint. Early research indicated that it is not the interaction style, be it menus, icons or commands, which affects the usability of the application but rather the care with which the interface has been designed (Potosnak, 1988). Furthermore, it was found that the system which proved to be the best for experienced users also proved the best for first-time users (Potosnak, 1988). From these observations, it was then concluded that icons do not necessarily increase the usability or learnability of an interface (Potosnak, 1988).

Despite this fact, use of icons as an interaction method remains the most popular amongst all users, whether experienced or not, even though keyboard shortcuts are the more efficient method of interaction (Lane, Napier, Peres and Sándor, 2005a), especially when compared with icon toolbars (Lane et al., 2005b).

Instead of icons being intuitive, it has been found that very often the icons themselves need further explanation; a fact which is evident by the repeated use of tooltips to explain the icons (Raskin, 2001). Icons themselves should only contain a label should there be any doubt as to what the icon represents, although addition of tooltips for all icons is highly recommended (Barr et al., 2002b; Bickford, 1997).

In this study, a set of alternative icons was developed in order to determine whether or not the icons which have been accepted as the de facto standard icons are truly the best icons for use within applications. This resulted in two sets of pictorial icons, referred to as the standard and the alternative set, being available to investigate their relevant impact on the usability of a word processor interface. A third set of icons was also developed which contained no images but instead displayed the name of the function represented by the icon. Although verbal and nonverbal stimuli are processed and recalled at different speeds, the advantages of icons were still applicable to the text buttons, while the use of only single words on the text button presented the prospect of less processing than that required for a large piece of text or a phrase. As has been discussed in a previous section, the use of icons creates additional difficulties, since icons generally use metonymy to refer to the larger metaphor in which the icon is found and are therefore subject to interpretation through semiotic analysis. The need for interpretation of an icon and the likelihood that an icon may be misinterpreted by the users are two of the disadvantages associated with icon use. It was expected that together with the advantages associated with the use of icons, the text buttons would simultaneously manage to eliminate the need to extract meaning from an image, thereby increasing their usability. The fact that the text buttons only included a single word, promoted recognition above recall and eliminated the need for interpretation of an image to extract meaning, all indicated that the text buttons could very well enhance the usability of the word processor application. The three icon sets were then included in a word processor application to determine which icon set afforded the best usability to the participants of the study.

2.11.6 Tooltips

Labels can be used to assist the user in determining the function linked to the icon (Richards et al., 1994). These labels can be in two forms, permanent or dynamic (Richards et al., 1994). Permanent labels are always visible and are usually found at the bottom of the icon (Richards et al., 1994), such as, for instance, the icons used on the Windows desktop. Dynamic labels are labels that appear as a pop-up window only when the mouse is moved or positioned over the icon (Galitz, 2002). These labels contain text describing the associated control (Galitz, 2002) and should be included in the interface to expand the meaning of the icon by giving additional leading text to explain the meaning of that icon (Microsoft, nd). These pop-up labels are also referred to as tooltips.

The major advantage associated with tooltips is that they assist in identifying otherwise unidentifiable objects (Galitz, 2002).

Some disadvantages of tooltips are (Galitz, 2002):

- They are not immediately visible and must be discovered by the user.
- Their appearance can be distracting for the user.

Tooltips should always only appear after a short time, ideally three-quarters of a second after the mouse pointer has been stationary on an icon or other interface component (Galitz, 2002). This will avoid distracting users should they inadvertently move the mouse over an icon. Tooltips are particularly useful for novice and beginner users, but the option should be provided to prevent the appearance of the tooltips as their inadvertent appearance could frustrate more experienced users (Maguire, 1990). The question remains, however, as to whether tooltips actually increase the usability of the interface.

It has been found that tooltips do not always serve to assist the user in determining what the icon represents (Zammit, 2000). Different users use visual cues provided by the icon and the textual cues provided by the tooltips differently depending on their ability to interpret these cues (McDougall, Forsythe and Stares, 2005). For instance, second language speakers were able to use labels showing the function name only when this function name was a frequently-used English word (McDougall et al., 2005).

When using icons to create a language for communication between professionals in the medical field, labels used to identify the object depicted in the icon did not overly influence icon preference, and only assisted users in interpreting iconic sentences when the meaning was not clearly communicated by the icons (De Carolis, De Rosis and Errore, 1995). However, this study only used tooltips in the cases where the meaning of the icon was not clearly depicted anyway (De Carolis et al., 1995).

The reason for inclusion of tooltips in this study was twofold. Firstly, it was investigated whether or not tooltips increased the usability of the interface. To investigate this, tooltips were included in some interface configurations and excluded from others. Secondly, tooltips provided a second interface component with which the translation of the interface could be tested. Therefore, all interface configurations which included tooltips were available with tooltips in English, Afrikaans and Sesotho.

2.11.7 Classification of icons

Classification of icons originated during the Xerox project that was responsible for the first commercial product that used icons (Smith, Irby, Kimball and Verplank, 1982). During this project, icons were simply categorised as being data or function icons (Smith et al., 1982). Since then a number of icon classification systems have been developed. For example, Gittins (1986) classified icons according to their form, type and colour. Lodding (1983) used the image functions as defined by Arnheim (as cited in Lodding, 1983) and paired each of these functions with a certain icon design. The image functions were classified as either physically displaying that which they represented (pictograph), or being analogous to that which they represented (symbol), or simply being an image that has been learnt and is an accepted representation of an object or function (sign). The picture function is then paired with representational design of an icon, the symbol function with an abstract icon design and the sign function with an arbitrary icon design (Lodding, 1983). However, none of these classifications make provision for images which incorporate text in any way, or which consist entirely of text and thus none were ideally suited to the current study.

Zammit (2000) provided a more appropriate taxonomy by classifying the icon according to the visual representation of the icon (Figure 2.5).

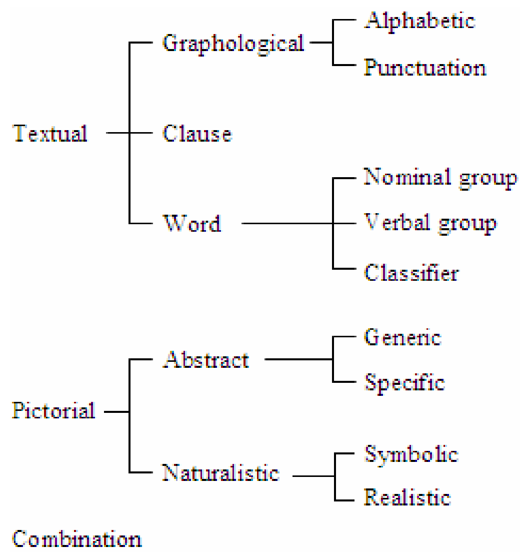


Figure 2.5: Icon classification

Source: Zammit (2000)

According to this taxonomy, icons fall into a textual, pictorial or combination category. The textual icons use no images, instead they either use letters of the alphabet (alphabetic graphological), punctuation marks (punctuation graphological), or a complete word as the icon (word), where the word used is a noun (nominal group), a verb (verbal group) or an adjective (classifier) (Zammit, 2000). For example, the icons contained in Figure 2.6, obtained from the current version of Microsoft Office, are all alphabetic icons, and Figure 2.7 shows a punctuation icon, also from Microsoft Office.

Figure 2.6: Microsoft Office textual alphabetic icons

Figure 2.7: Microsoft Office textual punctuation icons

Pictorial icons are classified according to the style of the image that is represented on the icon (Zammit, 2000). Abstract pictorial icons (Figure 2.8 A & B) have minimal colour use and are 2-dimensional in nature, whilst naturalistic icons (Figure 2.8 C & D) are more colourful and are 3-dimensional in nature (Zammit, 2000). Generic icons are commonly used images and can be found in settings other than the application in which they being used (for example, the icons used for play and stop on a video recorder), while specific icons are unique to the application in which they are found.

Symbolic icons do not reflect the actual function of the icon – they are symbolic of the represented function, while realistic icons are concrete representations of the function represented by the icon (Zammit, 2000).

Figure 2.8: Microsoft Office pictorial icons

Using the taxonomy of Zammit (2000) to classify icons, this study used two sets of abstract pictorial icons and a set of textual icons which could be further classified as belonging to the word group of textual icons. The textual icons included in this study will be referred to as text buttons.

2.11.8 Development of icons

Correct interpretation of an icon depends on the actual image used on the icon, the tooltip used to expand on the icon and the frame of reference provided by the context in which the icon is used (Lodding, 1983).

To ensure correct interpretation of an icon, icon development guidelines should be closely adhered to. When designing the image used in the icon the following guidelines played an important part in the development process for this study (Section 4.4.1):

- Represent the object in a familiar and recognisable way (Apple, 2006; Shneiderman, 1998). If possible icons should resemble their real-world counterparts (Marcus et al., 1995). Users assume that the icon depicts the underlying function and therefore tend to interpret the icon based on that assumption (Ferreira et al., 2006). Therefore, the most successful icons are those that closely resemble the function that they represent (Ferreira et al., 2006).
- Make each icon distinctive from every other icon (Apple, 2006; Shneiderman 1998) to avoid confusion between similar icons.
- Ensure the harmoniousness of each icon as a member of a family of icons (Shneiderman, 1998). Interpretation of the icon within context limits the potential for iconic ambiguities (Rogers, 1989b). The context in which the

icon is interpreted plays an important role in the understanding of the icon (Barr et al., 2002b; Preece et al., 1994). Context can be provided by surrounding icons, labels, text or windows (Barr et al., 2002b; Piamonte, 2000). For example, replacing the well-known male and female figures generally used to indicate the toilet facilities with a man's shoe and a woman's high-heeled shoe inside a restaurant should not cause any confusion as the context within which these images are found serve to assist patrons in interpreting their meaning (Rogers, 1989b). However, should these signs have been used in a supermarket their meaning could very well be interpreted as the area in the supermarket where men's and women's shoes are sold (Rogers, 1989b). The prior knowledge that the user brings to the system also assists the users in interpreting the meaning of the icons (Rogers, 1989b). Thus, cognisance should be taken not only of the single icon which is being designed but also the effect it will have on the interpretation of the adjoining icons and whether that impact will be a positive one. Given this fact it is obvious that novice or first-time users would experience the most difficulty in interpreting the icons correctly since they have the least knowledge of the system, while users who are experienced in the task domain should be able to use the application easily (Rogers, 1989b).

- Icons are only useful within an application if they successfully present a meaningful association between the icon and that which it represents. This will speed up the learning process, increase recall and provide a more effective system (Galitz, 2002). A meaningful icon is one in which the association between the icon image and the functionality it represents is noticeable (Lansdale, 1988). The extent to which something can be remembered is directly related to its meaningfulness (Preece et al., 1994). The more meaningful an object is, the greater the depth at which it is processed, thereby increasing the likelihood of remembering the object (Preece et al., 1994). Designers should therefore ensure that icons are meaningful (Preece et al., 1994). Although user performance is not affected by the meaningfulness of the icon, meaningful icons can be recalled with greater ease by the users (Lansdale, 1988).

Some of these guidelines for the development of the icons were used to design the icons required to complete the alternative set of icons. This development process will be discussed in greater detail in Chapter 4 (Section 4.4.1). These guidelines also assisted in the evaluation of the usability of some of the icons when inspecting the results of the user interaction. This evaluation of the icons will be discussed in Chapters 7, 8 and 9.

2.11.9 Evaluation of icons

Intuitive icons are icons which can be recognised by at least 95% of users the first time they see them (Ellis et al., 1995). A "good" icon is one which successfully conveys the same message to all users, regardless of their background, education or nationality (Jervell and Olsen, 1985). Hersh (as cited in Lodding, 1983) lists three criteria by which an icon can be assessed to be a successful icon:

1. Users must be able to interpret the meaning of the icon and find that meaning to be appropriate to the functionality of the icon.
2. There should be only a single icon which can be identified as appropriate when having to identify the correct icon for a task.
3. The icon must not be able to be associated with any negative connotations.

Chapter 5 includes a discussion on the results of icon recognition where users were unable to establish the meaning of some newly-designed icons. The icons eventually had to be discarded as inappropriate due to the lack of recognition and the inability of the users to correctly interpret their meaning. Chapter 7 contains a discussion of the effect of violation of point number 2, where two icons that were similar to one another were included in the same interface and created difficulty for users when attempting to complete a task.

The ISO methodology for evaluating icons is a three-step process (as cited in De Carolis et al., 1995):

1. The first step is called a production test and requires users to depict the icon function in a manner that they think best represents the function.
2. The second step is known as the appropriateness ranking test. This step involves users ranking the icons in order of appropriateness.

3. The final step is called the comprehension-recognition test where the chosen icons are tested with a group of potential representative users.

The testing of the icons can be performed as a recognition test or a timed test (Rogers 1989a). The recognition test involves showing a group of representative users the icons and then requiring them to provide a short explanation of what they think the icons represent (Rogers 1989a). The timed test is the capturing of the time required by users to identify the icon (Rogers 1989a). Although these tests are useful for evaluation and testing of the icons, context still plays a vital role in the interpretation of icons, as discussed in the previous section. Therefore, it may still be necessary to test the icons within the context of the application for which they are intended in order to establish their effectiveness and the effect that they have on the performance of the user (Rogers 1989a).

A few of the steps of the mentioned icon testing methods were implemented in this study. A production test was used to design a number of different alternative icons during a brainstorming session. An appropriateness ranking test was then done by means of distributing the designed alternatives amongst a group of non-computer literate users and requiring them to select the most appropriate icon from all the listed icons for each function. Similarly, some of the icons used in the study had been identified in a previous undertaking (Teklebrhan and Blignaut, 2005) by means of an appropriateness ranking test. An adapted form of the production test was then used during a second brainstorming session to complete the set of alternative icons. This production test included using the icon design guidelines mentioned in a previous section. Chapter 5 also describes an adapted appropriateness ranking test using only the different icons tested in this study (Section 5.7). A word processor application was then developed to allow the icons to be tested within the context in which they would eventually be used. A comprehension-recognition test was thereafter conducted, requiring users to complete a number of set tasks using the developed word processor application. Chapters 6 and 7 discuss the results of the empirical testing of the icons as they were implemented in different interfaces.

2.11.10 Icons, menus and text

Working on the acceptance of Paivio's (1971; Section 2.10) dual-code theory and accepting the fact that images are processed faster than text, and also that images offer better recognition than text, would seem to suggest that an iconic interface would be much easier and faster to use than one using only text-based interaction methods. However, even under the assumption that images and words are processed and assimilated using different techniques, it still cannot simply be assumed that the performance of the users will influenced (Lansdale, Simpson and Stroud 1990). This is evident from the conclusions previously drawn on the possible advantages associated with using text in place of images as icons (Section 2.11.5). The GUI is a definite improvement on the older style of textual interfaces where users were required to interact with the system via key presses and not the mouse. The GUI has increased effectiveness, efficiency and satisfaction (Staggers, 2000) and both novice and expert users performed slightly better on a graphical interface than on a text interface (Dillon and Song, 1997).

Any attention that is devoted to the interface detracts from the concentration of the user and constitutes interference with the primary task (Benbasat and Todd, 1993). Shneiderman (1998) notes that word processor applications generally make use of textual menus, whilst painting programs usually incorporate the use of icons, although nowadays most applications make use of both menus and icons. This distinctive design style is motivated by the fact that users are assisted by textual menus when concentrating on textual tasks and the more visual icons are better suited to a visually sensitive task, such as painting (Shneiderman, 1998). This prevents the user from having to switch between differing cognitive styles (Ellis et al., 1998). Conversely, it has been proposed that text and pictures require different amounts of inference in order to process the information (Paivio, 1971). In support of this, it has been suggested that since processing of text is considered to be a cognitive task, and the user is typically focused on some cognitive task when using a computer, more interference will be caused when using a text-based interface as it draws on the same cognitive resources as those required during completion of the task (Benbasat and Todd, 1993). The less the interface interferes with the completion of the task, the more successfully the task can be completed, thus an iconic interface should be more usable than a text-based interface (Benbasat and Todd, 1993). A number of studies have focused on the differences in user response between traditional pictorial icons and those using textual icons or menus in an attempt to verify either one of these theories.

2.11.10.1 User performance

Users do not always prefer icons to text due to possible ambiguity of icons (Kang, 2003). Textual links were preferred to icons even though textual links were more inconvenient and required more effort to use because users had to scroll through the page to locate the correct text link, while the iconic links were all situated in a navigation bar situated on the left hand side of the screen (Kang, 2003).

Performance does not always reflect preference and no visible advantages were detected between pictorial and textual icons (Benbasat and Todd, 1993). Direct manipulation methods appeared to increase the rate at which users completed tasks when compared with a strictly menu-based interface, a phenomenon which disappeared after repeated tests, showing that users quickly became accustomed to their particular interface style (Benbasat and Todd, 1993). No difference was detected between the error rates of the users of these interfaces (Benbasat and Todd, 1993). Numerous other studies have been conducted testing different aspects of iconic interfaces and comparing textual interfaces with iconic interfaces. Benbasat and Todd (1993) summarised these available studies in icon comparison and found that none indicated an advantage of the iconic interface.

Neither pictorial nor text icons were always immediately recognisable to a group of 11 and 12-year-olds (Zammit, 2000). Thus, it could be concluded that the pictorial icons, including naturalistic pictorial icons, were not necessarily better or easier to read than textual icons (Zammit, 2000). It was, however, discovered that some icons tended to be better than others because of the differing performance results obtained during use of these icons (Zammit, 2000).

Comparison of the interaction styles of command line, menu and icons indicated that no particular interaction style increased user performance significantly more than any other interaction style (Whiteside, Jones Levy and Wixon, 1985). Menu systems were responsible for the worst performance for all types of users tested in the study and iconic systems did not assist beginner users any more than the other systems did, contrary to what is suggested by the advantages listed for menus and icons (Whiteside et al., 1985). Therefore, it was concluded that it is not the interaction styles that influenced user performance, but rather the care with which the interface was developed that had an influence on the ease of use and ease of learning of the interface (Whiteside et al., 1985).

Important to note is that the same icons and the same types of icons do not necessarily offer the same usability to different users therefore, interfaces should be tested for compatibility with target audiences. This fact was highlighted by the differences detected between American and Chinese users; where the former users performed better with alphanumeric and the latter with the pictorial icons (Choong and Salvendy, 1998). An attempt to replicate this study was made but with some subtle differences. Mixed modality icons were defined as those that had an image incorporating both a textual item, either in English or in Chinese, as well as a non-textual item which was hypothesised as adding an additional verbal cue to the original pictorial icon (Kurniawan, Goonetilleke and Shih, 2001). Textual icons were found to be more meaningful and appropriate than pictorial icons, but the difference between Chinese and English meaningfulness and appropriateness was not significant (Kurniawan et al., 2001). This could be explained by the fact that the test subjects were all fluent bilinguals, thus whether the text appeared in English or in Chinese made no difference to them (Kurniawan et al., 2001). Mixed modality icons did not significantly increase user performance (Kurniawan et al., 2001).

2.11.10.2 Accuracy, understandability and recall

Rogers (as cited in Rogers, 1989b) conducted a study comparing the learnability and memorability of icon-based versus name-based menus for a set of word processing functions and system operating commands. Results indicated that user performance was not affected by the type of menus used but that icons were more easily recalled than the name-based menus (Rogers, 1989b). Even though users were able to perform on an even par, regardless of the type of menu system used, there was clearly a difference in memorability between the different types (Rogers, 1989b). In a similar study comparing the recall qualities of icons versus text, Lansdale et al. (1990) failed to confirm these findings when using an information retrieval task as their base and detected no significant difference between the recall abilities of the participants.

Muter and Mayson (1986) found that the best combination for selection from a menu was when a graphic was added to the menu name. Egido and Patterson (1988) confirmed these findings with their comparison of pictorial, textual and a combination of the two. Users of a catalogue-browsing system indicated a higher preference for inclusion of pictures to aid the search process. This was confirmed by enhanced user performance with browsing aids which were a combination of pictures and labels (Egido and Patterson, 1988). Further confirmation of this was found when user accuracy was shown to be highest when selecting from a mixed format of text and graphics in a menu structure as opposed to a graphics only or text only menu structure (Kacmar and Carey, 1991). However, no discernable difference in the time taken to make a selection was detected between the three formats (Kacmar and Carey, 1991).

Many studies which have compared the use of icons to labels have found that the labels were easier to understand than the icons – a finding which contradicts one of the supposed greatest advantages of icons due to their purported intuitiveness and understandability (Raskin, 2001).

In summarising the findings of the previous two sections it would seem that text promotes higher understandability than icons, while icons increase the recall capabilities of users. When comparing the objective performance measures of time to complete a task and the number of errors made during completion of the task, no performance differences were detected between the users of text icons and users of pictorial icons, although accuracy was increased in selection from a menu structure when the menu items combined text and pictures. These findings were, however, not replicated when set in a word processing environment. Icons increased memorability when used in a word processing application but not when used in an information retrieval task. All of these seemingly contradictory findings point to the fact that different aspects of text and icons are applicable in different degrees based on the environment, the user group and the tasks which must be completed. The tasks and environment would most likely go hand in hand and based on the results of the discussed studies it seems probable that different tasks influence the usability of the text and icons in different ways. These differences could be attributed to the level of interference present between attention focused on the interface and that focused on the task. Users of a certain application may share some traits and characteristics and since users tend to use the provided cues differently and to different extents, these users will respond differently to the use of icons and text within an interface. Due to these conflicting results between different applications and tasks, care should be taken when deciding whether to use text or pictures, how to use them and to what extent they should be included in the interface. Therefore, it is important that the effects of pictorial and textual icons be tested with each distinctive group within the target market for each application type to determine which icon type best suits that group. This study aimed to investigate the effect of pictorial and textual icons on users from a selected region in South Africa when interacting with a word processor.

2.11.10.3 Word processing studies

A previous section mentioned the study undertaken by Rogers (as cited in Rogers, 1989b) to investigate the difference between understandability and memorability of name-based and icon-based menus, which found no difference in the performance of users of these interaction styles.

In a study similar to the current undertaking a word processing environment was chosen to test the hypotheses that (i) giving users a choice of working either with icons or menus would be the fastest interaction method and (ii) by giving users only a menu to work with they would make fewer errors than those with only icons or a choice between menus and icons (Ellis et al., 1995). 15 participants were divided into three groups of 5 and each group was assigned to a single interface (Ellis et al., 1995). One interface had only icons, the second interface had only menus and the third had both icons and menus (Ellis et al., 1995). Users were required to complete 15 word processing tasks (Ellis et al., 1995). The time to complete a task and the number of errors were measured and analysed as measurements of efficiency and effectiveness, and user satisfaction was also determined (Ellis et al., 1995). The majority of the tasks were performed faster by the users of the iconic interface but only those tasks which required users to change the font style of the text indicated a significant difference between the task times (Ellis et al., 1995). The tasks required users to bold, underline, italicise and change the character spacing and in all of these tasks users of the iconic interface performed the tasks faster than the users of the menu interface (Ellis et al., 1995). Using the menu to complete these tasks required users to navigate through the menu and to interact with complex dialog boxes; using the icons, however, was more effective and succeeded in conveying the correct function to the users (Ellis et al., 1995). The icons (the same as those found in current word processing applications) were successful due to three reasons (Ellis et al., 1995):

1. Common functions were grouped together into a single coherent group.
2. Icons included verbal clues due to the use of the first letter of the function.
3. The font style was reproduced on the icon, providing an additional visual clue.

Users of the iconic interface did, however, have a higher error rate, although the detected difference was not significant (Ellis et al., 1995). User satisfaction responses indicated that many users felt more in control when using the icons as they did not then have to navigate through deeply nested menus in search of the correct menu option (Ellis et al., 1995). In the case of functions that did not have an easily interpretable icon, it was felt that the menus were easier to use due to the verbal representation of the function (Ellis et al., 1995). Another problem that manifested itself and should be kept in mind was that many users did not understand the terminology exclusive to word processing functions such as alignment and font (Ellis et al., 1995). The researchers in this study recommended that the study be reproduced with a larger sample and with the development of a testing instrument which allows for automatic capturing of task completion time and erroneous selections.

The current study is a replication of this previous study but on a much larger scale. More representative users and different icon configurations were tested, including different icons sets, different icon types and with menus and tooltips in different languages.

2.12 Summary

This chapter provided a summary of some of the available literature on which the current study was based and which provided motivation and a strong theoretical foundation for the study at hand.

The growth of the global market and the dissemination of computer application throughout all regions of the world have resulted in strong emphasis being placed on the need for the development of usable applications. With this globalisation effect a set of icons has emerged as the accepted de facto standard set of icons for the group of functions represented by these icons. However, just because these icons have been accepted as the standard does not necessarily imply that they are the best icons for the functions and, as mentioned in the previous chapter, a supposedly intuitive and easy to use interface is not automatically intuitive and easy to use for all users. Following this logic a set of preferred alternative icons were identified by a group of non-computer literate questionnaire respondents during an icon appropriateness ranking test. The set was completed by making use of available icon design guidelines and designing according to these during a production test. The Microsoft Office icons were used as the accepted standard and the second set of icons provided an alternative set whose usability could be compared with those of the standard icons.

Images are processed, memorised and recalled more easily than text is, therefore pictorial icons should be easier to use than text icons. However, text icons should foster recognition over recall in the same way that pictorial icons or menus do, while simultaneously lightening the load of interpretation since no representational meaning has to be extracted from text buttons. A number of studies have proven that this is not the case since there was no noticeable difference between user performance on pictorial and text icons, apart from a perceived increase in user accuracy and recall. However, different users responded to different verbal and nonverbal cues based on their ability to assimilate these cues. Therefore, it is important to test the effect of different icons and different icon types on all potential product users and the third set of icons added to this study were those of text buttons which contained no images but instead simply displayed the function name. The role of menus and tooltips in usability was also discussed and they were in turn included in the study as interface components whose usability had to be tested.

The discussion on bilingualism indicated that English second language speakers may be at a disadvantage from the outset when using an English interface. The unequal ability to assimilate L1 and L2 immediately places the second language speaker at a disadvantage when using an English interface. However, most South African studies indicated no need for translation into the first language of a bilingual user. Extended use of the English interface may also negate the need for translation, which could eventually impede the performance of the user. The final aim of the study was therefore to determine whether translation of the interface would increase the usability of the product for English second language speakers. Menus, tooltips and text buttons were all available in English, Afrikaans and Sesotho.

The following chapter will discuss the research methodology of the study in detail and how the established theory of this chapter could be tested using sound theoretical research methodology.

Chapter 3

Research methodology

3.1 Introduction

The preceding chapter offered an overview of some of the available literature and argumentation of the theory that provided a foundation and motivation for this study. Once the aim and motivation for a research study have been established, the methodology to be used must be carefully planned and grounded within established research methodology theory. This chapter explores the various options available for conducting a research study and motivates the methodology used in this study.

3.2 Research design

The goal of research in human-computer interaction is the promotion of scientific understanding of the interaction between humans and computers, which includes the tools that are used during the interaction (Giacoppo, 2001). Although HCI research can be grounded and based on the scientific method of research, many non-scientific disciplines have been adapted to HCI research (Giacoppo, 2001). Much of the driving force behind development of HCI research methods stems from the field of scientific research in psychology (Faulkner, 2000; Giacoppo, 2001). As such HCI research leans heavily towards the scientific experimental method (Faulkner, 2000). HCI, much the same as psychology, is an interdisciplinary field with many fields lending themselves to that of HCI (Giacoppo, 2001). Consequently, there are a number of research methodology tools available for research in the HCI field and the applicable methods must be decided upon based on the type of research and the phenomenon that is being investigated (Olivier, 2004).

The research design is an exact blueprint or plan of how the research will be conducted (Mouton, 2001), including how the units will be divided into groups (Trochim, 2002). Some texts indicate that the research design is decided upon based on whether the sample members were randomly assigned to groups or not (for example, Trochim, 2002; Urdan, 2005). However, Mouton (2001) offers a well-structured and comprehensive framework for determining which research methodology is applicable to the study in question and therefore that source was heavily relied upon during the formulation of the research design for this study.

In order to plan out the research study, the first objective is to decide on a research problem, after which the focus of the research is narrowed down by means of stating a research question (Mouton, 2001; Trochim, 2002), and then the research design must be decided upon (Mouton, 2001). A sample must be selected from which data can be collected, and then an analysis of the data must be made (Trochim, 2002). Conclusions can be drawn from the data analysis and generalisations can be inferred back to the original broad research problem (Trochim, 2002).

3.2.1 Research problem

The unit of analysis is the entity which will be investigated and analysed during the course of the research study (Mouton, 2001; Trochim, 2002). Based on what the unit of analysis is, the research problem falls either into a world of real-world objects or into a conceptual world (Mouton, 2001). Since this research problem is focused on the usability of interface components within a software application, it is focused on human beings and social interventions and is therefore concerned with real-world objects (Mouton, 2001).

3.2.2 Research question

Research problems are often formulated as research questions (Mouton, 2001) and are usually formulated in terms of a hypothesis (Trochim, 2002). Trochim (2002) lists only three types of research questions – descriptive, relational and causal. However, Mouton (2001) first distinguishes between empirical and non-empirical questions and then further subdivides empirical and non-empirical into a number of different types of questions.

Since it has already been established that the research problem is focused on real-world objects, the research question for this study was an empirical question.

Empirical questions are subdivided into exploratory, descriptive, causal, evaluative, predictive and historical (Mouton, 2001). Since the aim of the study was to determine whether or not language, icon type and icon sets had an effect on the usability of a word processor, the research question was causal in nature.

3.2.3 Sampling

Samples can be either probability samples or nonprobability samples (Berenson and Levine, 1979). A probability sample is obtained from a sampling method that utilises randomised selection (Trochim, 2002). A sample is random if the probability that it was selected from the population was the same as the probability that any other sample of equal size had been selected from the same population (Dowdy and Weardon, 1983). Therefore each equal sample size has the same likelihood of being picked from the population (Dowdy and Weardon, 1983). A random sample is ideal but it is however not always possible to obtain a random sample (Dowdy and Weardon, 1983). There are four common types of probability samples; the simple random sample, the systematic sample, the stratified sample and the cluster sample (Berenson and Levine, 1979).

Nonprobability samples are easier to obtain but they have the disadvantage that there is no way to determine how representative the sample is of the whole population (Berenson and Levine, 1979). They do not involve random selection and can broadly be divided into two types: accidental or purposive (Trochim, 2002). Accidental sampling is also known as convenience sampling and is the method used when the participants are included in the study based on proximity, ease-of-access and willingness to participate, that is, they are the most convenient participants to use (Urdan, 2005). Purposive sampling is sampling done with a purpose in mind and where sampling is restricted to a predefined group of people (Trochim, 2002).

This study made use of three different convenience samples. The first sample comprised high school students who were readily accessible in the area in which the study was conducted. The second and third samples were both used due to their availability via the university where the study was conducted. The second sample consisted of a group of cleaners employed by the university and the third sample was made up of students who had registered for a specific course.

3.2.4 Data collection

The next step in the research design is to determine how the data will be collected. To resolve an empirical question, new data, known as primary data, has to be collected or existing data about the world has to analysed (Mouton, 2001). Since the study involved collection of new data and no existing data was used to analyse the effect of the language or icons on the usability, an empirical study with primary data was undertaken. Primary data for an empirical study can be collected via experiments, surveys, case studies, programme evaluation or ethnographic studies (Mouton, 2001). However, since it has already been established that the research question is causal, the data collection methods can be narrowed down to surveys; comparative studies and experiments, where experiments can be conducted either in the field or in a laboratory setting (Mouton, 2001). Comparative studies investigate differences and similarities between the unit of analysis (Mouton, 2001) and were thus not relevant to this study. Therefore, this study could use experiments and surveys.

3.2.5 Research in information technology

When the research problem and design were analysed independently of the field in which it is grounded, the conclusion was reached that it was an empirical research problem investigating a causal relationship which required experiments and surveys to investigate further. However, it still had to be established whether or not the chosen methodology was relevant to a study in information technology. Olivier (2004) lists a number of research methods that are applicable to information technology and which include literature survey, arguments, mathematical proofs, surveys, experiments and case studies (Olivier, 2004).

According to Olivier (2004), this study could be considered to have social goals. Research studies with social goals most often make use of literature surveys, surveys, experiments and case studies, while empirical studies in information technology are those that are concerned with observation and include surveys, experiments and case studies (Olivier, 2004). Therefore, in terms of Mouton's (2001) classification of research studies, the correct methodology for this study would be surveys and experiments and when viewed with respect to the information technology field (Olivier, 2004), it was found that these studies were equally applicable in the information technology setting, a finding which is supported by Atwater and Babaria (2001) and Kuter and Yilmaz (2001). In fact, comparative experimentation is particularly useful when a number of alternative designs must be compared to one another (Faulkner, 2000).

3.3 Surveys

A survey is an examination of a system in which the investigator has no power to change the conditions to the objects of study (Dowdy and Weardon, 1983) and in information technology it is used for obtaining information from a large number of users (Usability glossary, nd).

There are two types of data which can be collected from a survey, namely qualitative and quantitative (Berenson and Levine, 1979). Qualitative data stems from categorical responses, while quantitative responses are numerical and can be either discrete or continuous (Berenson and Levine, 1979). Discrete quantitative data are derived from a counting process (Berenson and Levine, 1979), while continuous quantitative data stems from a measuring process (Berenson and Levine, 1979).

When viewed in broad terms, there are two types of surveys namely, questionnaires and interviews (Trochim, 2002).

3.3.1 Questionnaires

A questionnaire is defined as being a form or a set of questions which can be filled out by participants (Usability glossary, nd). The form allows for demographic information, views and interests of the respondents to be obtained (Usability glossary, nd).

In terms of usage within HCI, questionnaires are used to elicit user reactions to and opinions on an application or website (Kuter and Yilmaz, 2001) since they are a reliable source of subjective measurement (Faulkner, 2000) and provide for a means of obtaining large amounts of data (Faulkner, 1998). Questionnaires can be administered via an interviewer, or they may be self-administered, where the participant reads and answers the questionnaire without assistance (Faulkner, 2000). Furthermore, they can be divided into mail and on-line questionnaires, with on-line questionnaires being further subdivided into e-mail, computer-direct and web-based questionnaires (Kuter and Yilmaz, 2001).

The response format of a questionnaire determines whether the questionnaire is unstructured or structured (Trochim, 2002). An unstructured questionnaire consists of open questions and encourages respondents to answer in any way that they want to (Faulkner, 1998; Preece et al., 1994). Structured questionnaires are compiled from closed questions which require users to select an answer from a set of predetermined answers (Faulkner, 1998; Preece et al., 1994).

This study used questionnaires to establish user demographics (Section 4.5.1), to determine the anxiety (Section 4.6) and attitude (Section 4.7) levels of the test participants as well as the subjective satisfaction (Section 4.9.6) of the participants after usage of the different interfaces. All questionnaires were structured, self-administered, computer-direct questionnaires. Response formats used in the questionnaires included multiple-choice, Likert and LPC (Olivier, 2004) scales. Questionnaires were analysed by first rescaling negatively worded questions if there were any and then calculating the sum of the responses as the total of the variable being measured by the questionnaire (Trochim, 2002). In some cases, the answers first had to be multiplied by a factor-loading value which indicated the weight of the question with respect to the questionnaire in total.

Although cognisance should be and was taken of the numerous guidelines available for setting up a questionnaire (for example, Faulkner, 1998; Trochim, 2002), apart from the user demographic questionnaire, this study used only those questionnaires that had already been designed and used in previous studies and had had their validity and reliability verified (Sections 4.6, 4.7 and 4.9.6).

3.3.2 Interviews

An interview differs from a traditional questionnaire because it is completed by the interviewer, based on the responses of the interviewee (Trochim, 2002). An interview can either be personally or telephonically administered (Trochim, 2002) and can take the form of a structured or unstructured interview (Faulkner, 2000). Interviews were not used in the study.

3.4 Experiments

An experiment constitutes the collection of data about the population that is treated or controlled by the experimenter in the form of measurements and observations (Dowdy and Weardon, 1983). When conducting an experiment the sample is usually divided into two or more groups and then compared to one another according to one or more variables (Urdan, 2005).

A variable is anything that can be categorised and is capable of having more than one value (Urdan, 2005).Variables can either be dependent or independent (Walsh and Ollenburger, 2001). A dependent variable is one whose value is dependent on the value of another variable, while an independent variable is the variable on which the dependent variable depends on and by which it is affected (Walsh and Ollenburger, 2001). Experimental research involves isolation of an independent variable which may influence the dependent variable (Urdan, 2005). Researchers then manipulate the independent variable to ascertain the effect that it has on the dependent variable (Urdan, 2005).

Experiments are usually quantitative in nature (Mouton, 2001) and are considered to be the most robust of the research methods (Trochim, 2002) since they allow for isolation of independent variables which could be the cause of variance in dependent variables (Urdan, 2005). An experimental design is one which affords the highest internal validity, which is an assurance that what was done was the cause of what was observed (Trochim, 2002).

Experimentation in HCI can be in the form of a comparative experiment, where the system is tested against an existing system or a different design (Faulkner, 1998). It can also be in the form of an absolute experiment, where the system is tested in isolation (Faulkner, 1998). Prior to the conducting of an experiment, a hypothesis must first be established, which will then establish a cause and effect relationship (Faulkner, 1998). Experiments are particularly useful to compare a number of alternate designs to gauge the effect of certain variables on another one (Olivier, 2004).

The major part of this study constituted comparative experiments for collection of primary data. Experiments in HCI are generally performed as usability evaluations.

3.4.1 Usability evaluation techniques

"Usability testing is the cornerstone of user-interface design" (Cox and Walker, 1993).

The above quote emphasises the important role that usability evaluation plays in system design, so important a role that Shneiderman (1998) defines three pillars of design, of which expert reviews and usability testing comprise the final pillar.

Evaluation refers to the gathering of data with the aim of evaluating the usability of a system by a specified group of users (Preece et al., 1994). It allows for determination of the extent to which a product meets the needs of the users (Preece et al., 1994).

There are many evaluation techniques available to designers (Al-Qaimari and McRostie; Shneiderman, 1998) and it is important that the correct, most appropriate method be chosen to evaluate the system in question (Redmond-Pyle and Moore, 1995). The methods of evaluation depend on whether designers make use of usability experts, models or potential product users to assess the usability of a product (Scholtz, 2004). Experts are involved in usability evaluation techniques known as expert reviews and models are used in model-based evaluation (Scholtz, 2004). Whenever users are involved in the evaluation of a system it is known as user testing (Scholtz, 2004). This study relied heavily on user testing as a usability evaluation technique.

Therefore, the bulk of the discussion on usability evaluation will concentrate on user testing. The remainder of the afore-mentioned usability evaluation techniques were not appropriate for the study at hand and as such, each will only be briefly discussed in the following sections.

3.4.1.1 Expert reviews

Expert reviews involve using the services and expertise of a usability expert to test the usability of a system (Newman and Lamming, 1995), either by means of determining conformance of the system to guidelines and standards or by simulation of a targeted user. Heuristic evaluation, guideline reviews, pluralistic usability walkthrough, consistency inspection and a cognitive walkthrough are all examples of expert reviews.

Whilst having the advantage of being faster and cheaper (Newman and Lamming, 1995; Scholtz, 2004), these methods exhibit two considerable disadvantages. Firstly, at no point are the intended users of the system included in the testing of the product. Secondly, these methods require the participation of a usability expert who at the very least could experience difficulty in making time available for the testing of the application (Newman and Lamming, 1995).

The use of expert reviews was not employed during this study.

3.4.1.2 Model-based evaluation

Model-based evaluation consists of using a model to predict user behaviour (Scholtz, 2004). An example of a model-based evaluation is the GOMS model which consists of Goals, Operators, Methods and Selection rules. The GOMS model uses a model of the human information processor to predict performance on a user interface (Scholtz, 2004). Model-based evaluation is less expensive than user testing but the designed models must first be tested for validity, and this is a time-consuming task (Scholtz, 2004).

Model-based evaluation was not used in this study.

3.4.1.3 User testing

The usability engineering model as proposed by Nielsen (1992) includes a step that focuses exclusively on empirical user testing. User tests are accepted as the standard evaluation technique used to gauge the usability of an application (Al-Qaimari and McRostie, 2001) since they are the most useful method of studying usability (Nielsen, 2003b).

This form of empirical testing involves conducting a formal usability test where actual intended users of the product are required to interact with the product and then an analysis is performed on their interaction in terms of predefined usability measures (Al-Qaimari and McRostie, 2001; Dillon, 2001; Preece et al., 1994; Shneiderman, 1998). Users can also be asked for their subjective opinion of the application after completion of the test (Dillon, 2001) which allows for objective performance measures to be complemented by the capturing and analysis of subjective satisfaction measurements (Shneiderman, 1980). Measures of effectiveness, efficiency and satisfaction can be measured and analysed in this manner (Dillon, 2001). These measurements form the basis of most definitions of usability (Section 2.3). Objective measures are usually measured in terms of set usability measures which can be captured and analysed (Section 3.4.1.3.5). User opinion is obtained by means of administration of questionnaires (Preece et al., 1994). In this way it is possible to ascertain whether or not the product meets the needs and specifications of the user (Shneiderman, 1998). User testing is also known as user-centred evaluation due to the involvement of targeted users, which is the greatest advantage associated with this type of evaluation (Scholtz, 2004).

Empirical user-based testing is generally believed to deliver the most reliable and valid determination of a product's usability (Dillon, 2001) and even provide better performance measures than heuristic evaluations (Nielsen and Phillips, 1993). Dumas and Redish (1999) summarised a number of studies comparing the different available usability methods and found that user testing compares well with other methods, such as heuristic evaluations and cognitive walkthroughs. Furthermore, user testing resulted not only in more usability problems being discovered but the problems were also unique (Dumas and Redish, 1999). User tests do, however, take longer to conduct than other methods (Dumas and Redish, 1999). Another advantage of user tests is that they have been conducted successfully with users from diverse cultures (Nielsen, 1996a).

User testing has at least two obvious disadvantages, the first being that the test is generally the first time users come into contact with the product and the second being that, usually, only a subset of the applications functionality is tested (Shneiderman, 1998). User testing can also prove to be an expensive exercise and finding representative users may be a difficult task (Scholtz, 2004). Another consideration of user testing is whether the chosen tasks are in fact representative of the eventual usage of the application. However, the overwhelming advantage of user testing, which is the involvement of actual users (Scholtz, 2004), ensures that user testing is the recommended and most comprehensive form of usability testing (Nielsen, 2003b; Nielsen and Phillips, 1993).

The disadvantages present in user tests were not applicable to this study since the ideal situation was to target users who were unfamiliar with the product to test the intuitiveness and usability of the product. Secondly, a convenience sample was used which eliminated the expense associated with user tests and provided conveniently allocated times for conducting of the tests. Therefore, this method of usability evaluation was heavily relied upon during this study and the results of the usability testing will be discussed in Chapters 6 and 7.

3.4.1.3.1 Formative and summative evaluation

User testing is evaluation by means of observation or conducting of experiments and is therefore considered to be empirical in nature. This empirical testing can be in the form of formative or summative evaluation (Nielsen, 1992; Scholtz, 2004).

Formative evaluation involves getting verbal feedback from the users and requires elicitation of opinions, problems that exist with the system and general strengths and weaknesses of the system (Faulkner, 2000; Scholtz, 2004) and are most often carried out during the design process (Preece et al., 1994). Formative evaluation is synonymous with the gathering of qualitative data (Faulkner, 2000; Nielsen, 1992) and allows designers to identify usability problems by getting verbal feedback from the users instead of conducting formal full-scale empirical testing (Nielsen, 1992; Scholtz, 2004).

Conversely, summative evaluation concentrates on the collection of quantitative data (Faulkner, 2000; Nielsen, 1992), and usually takes place after completion of the system. It is designed to make judgements about the finished product (Preece et al., 1994) by measuring effectiveness, efficiency and user satisfaction (Scholtz, 2004). Both formative and summative evaluation should be in the form of empirical testing with users performing representative tasks on the application (Nielsen, 1992; Scholtz, 2004). Formative evaluation is slightly less formal than summative and requires fewer users (Scholtz, 2004). The test format remains the same regardless of which type of empirical testing is being performed, that is, users are required to complete a number of core tasks that have been identified as representative tasks (Nielsen, 1992).

Formative evaluation was conducted in the study and the results will be discussed in Chapter 5. The results of the more extensive summative evaluation will be discussed in Chapters 6 and 7.

3.4.1.3.2 Conducting user tests

When conducting user testing it is imperative that real users be tested while working on real tasks (Nielsen, nd). Users are observed interacting with the system and then evaluated on how well they were able to complete the tasks (Cox and Walker, 1993).

A user test must have the following structure:

1. The primary goal of improving product usability (Dumas and Redish, 1999).
2. Participants are real users (Dumas and Redish, 1999) which involves the procedure of identifying the target group and recruiting the users (Faulkner, 2000; Nielsen, 2003a).
3. Users are recruited (Faulkner, 2000) and real tasks are given to the users to complete (Dumas and Redish, 1999; Nielsen, 2003a).

4. Evaluation is carried out (Faulkner, 2000) by observing the participants and recording all their interaction (Dumas and Redish, 1999).

The final step differs for qualitative and quantitative evaluation. Qualitative evaluation has a final step of listening to user response (Nielsen, 2003a) while quantitative user testing has the following final step:

5. Data is analysed, problems are identified and solutions are recommended (Dumas and Redish, 1999) by reporting on the findings (Faulkner, 2000).

When employing usability studies for research purposes, the goal of the user test becomes the aim to discover whether a predetermined phenomenon exists (Dumas and Redish, 1999). The five steps listed above served as the framework for the conducting of the user testing during this study.

Before a user test is conducted it is important that a pilot test be conducted (Dumas and Redish, 1999). A pilot test is a dry run for the user test and verifies the desired procedure for the user test (Dumas and Redish, 1999). It also ensures that the facilitators are comfortable with the procedure and know what is expected of them and checks that all the equipment is functioning properly (Dumas and Redish, 1999). Data collected from the pilot test is not analysed (Dumas and Redish, 1999) it merely provides a way for evaluators to ensure the success of the main study by providing the opportunity to iron out all the problems in the design of the experiment (Preece et al., 1994). The formative evaluation which is discussed in Chapter 5 was conducted during the pilot test of the developed application.

3.4.1.3.3 Recruiting users

It is important that the participants in the user test are people who currently use or will use the product in the future (Dumas and Redish, 1999; Nielsen, nd).

Once the targeted users have been identified and recruited it is usually required that participants be classified according to their expertise (Faulkner, 2000). When testing users according to their computer experience or experience with the tested product, it is common practice to divide users according to those who know “a little” and those who know “a lot” (Dumas and Redish, 1999). A number of different definitions exist for novice and expert users, but the majority of studies simply classify users according to their level of experience, with more experienced users being classified as expert users and less experienced users being classified novices (Al-Qaimari and McRostie, 2001). This classification is based on the premise that, as users gain more experience and exposure, so their level of accuracy and performance increases and as such they can eventually be referred to as expert users (Chi, Glaser and Farr, 1988).

The division between “a little” and “a lot” is based on the impressions of the evaluators gathered during interviews and data collection (Dumas and Redish, 1999). Thus, the division of users into groups remains a subjective matter and one which is redefined for each study. One group that normally does not feature in these tests is the middle group – those between “a little” and “a lot” (Dumas and Redish, 1999). Should the same observations and trends exist for both the users that know “a little” and those that know “a lot” then there is no problem generalising the results to include the intermediate group (Dumas and Redish, 1999). Dividing users for the purposes of user testing between those that know “a little” and those that know “a lot” corresponds closely to the user classification proposed by Shneiderman (1998), who offers a more detailed analysis of user classification based on the level of knowledge of both the interface and the task domain. Therefore, the classification proposed by Shneiderman (1998) can easily be used to classify participants of a user test.

Care must be taken when selecting the participants to ensure that the testers do not fall into the trap of selecting participants that are either too experienced or those that are not experienced enough (Dumas and Redish, 1999). Should the participants be too experienced, then not all usability problems may be identified, and should the participants have too little experience, then changes may be made to the product that are not beneficial to the actual end users (Dumas and Redish, 1999).

When conducting user testing, it is imperative that data be collected about the users (Dumas and Redish, 1999). Should anomalies then surface during analysis of the data, the collected information may assist in identifying the variable responsible for the anomalies (Dumas and Redish, 1999). However, when dividing the participants into groups for analysis of the data, it is usually only possible to select a few characteristics by means of which participants can be divided into homogenous groups (Dumas and Redish, 1999).

At least twenty users per tested design is recommended when conducting quantitative user testing (Nielsen, 2001a), although some literature insists that 3 to 5 participants in each subgroup is enough to draw meaningful conclusions (Dumas and Redish, 1999).

Recruiting of the targeted users will be discussed in Chapters 5 (Section 5.2), 6 (Section 6.2) and 7 (Section 7.2). The classification of users proposed by Shneiderman (1998) offers a more comprehensive explanation of how to divide users into those who know "a little" and those who know "a lot" and it has been shown that the Shneiderman classification can be applicable for use in user testing. Therefore, participants in this study were classified using the definitions and classifications of Shneiderman (1998) as outlined in Section 2.4. To prevent the user testing being too one-sided by testing only those who know "a little" or those who know "a lot", representative users of all Shneiderman's (1998) categories were included in the user testing for this study. The methodology used to distinguish between these participants and to categorise them into a certain user class will be discussed in Section 4.5.2. As recommended, user demographics, not all of which were used when analysing the data, were captured and will be covered in Section 4.5.1.

3.4.1.3.4 Designing tasks

A goal is the desired state of the system or, more specifically, what the user wants to achieve (Preece et al., 1994). A task consists of the activities and actions that are deemed necessary to achieve the goal (Faulkner, 2000; Preece et al., 1994). An action is a "simple task" that requires no problem solving (Preece et al., 1994).

Task analysis is hierarchical, which means that all tasks can be broken down into a number of subtasks (Dumas and Redish, 1999; Faulkner, 2000). The tasks that are used in user testing are known as benchmark tasks and are considered to be standard tasks that can be performed in the system (Preece et al., 1994). Tasks that are included in the user test should follow in the natural order in which they will generally be used in the product and tasks that are important to the evaluation of the product should feature early on in the test (Dumas and Redish, 1999). Since it is likely that not all participants will be able to finish all the tasks, it is imperative that the most important tasks appear early in the test (Dumas and Redish, 1999).

The tasks that were used during the evaluation of the users will be discussed in Section 4.8, where it will clearly be seen that the tasks were decomposed into the simplest form possible. This was done to avoid a situation where the task required many actions to complete and thus could be partially completed – and so complicate the assessment of task completion. Tasks were also ordered in a manner that appeared natural and intuitive and tasks that may have caused difficulty were shifted to the end precisely for the reason that they may have prevented users from completing the test (Section 4.8).

3.4.1.3.5 Usability measurements

The ISO stresses that in order to test the usability of the product both the performance and satisfaction of the end-users must be measured in some way (ISO, 1998). In order to do this, effectiveness, efficiency and satisfaction must be defined in terms of measurable attributes (ISO, 1998). These measurements are in fact performance measures of the users and are known as usability measurements or usability metrics (Preece et al., 1994).

Performance measures are quantitative in nature and can include such measurements as (a) time to finish a task (Dumas and Redish, 1999; Faulkner, 1998), (b) the number of errors that are made during completion of the task, (c) the number of incorrect menu items and icons chosen during completion of the task, (d) the time spent using the help facility, (e) the number of incorrect choices made in a dialog box (Dumas and Redish, 1999), and (f) the ratio of success to failure (Faulkner, 1998).

Using the usability measures provided by the ISO (1998) as a basis and by combining the usability models of both Nielsen (1994) and Dix, Finlay, Abowd and Beale (1993) as well as taking all the different usability definitions into account, Abran et al. (2003a) identified a broad range of usability measures and proposed a comprehensive and consolidated usability model (Figure 3.1). The consolidated model also indicates the level of importance of each usability measurement. This usability model incorporates all proposed usability measurements and subdivides them amongst the different usability components. Each usability measurement is also ranked according to the importance of the measurement and how highly it is recommended that the measurement be captured.

For the purposes of this study, a number of usability measurements were identified to be captured and analysed. These usability measurements were identified using the consolidated model as depicted in Figure 3.1 and taking into serious consideration the importance level of each measurement. Also playing an important role in deciding which measurements to include in this study were the following invaluable guidelines and previous findings:

- Shneiderman (1998) maintains that there are measurable objectives for each task in a set of benchmark tasks and for each user of the system. These measurable objectives are:
  - **Time to learn** – How long does it take users to learn the commands or actions necessary to complete the task?
  - **Speed of performance** – Time taken to complete the task.
  - **Rate of errors by users** – Measurement of the number of errors made by the user in an attempt to carry out the task.
  - **Retention over time** – The extent to which users retain the knowledge they have gained.
  - **Subjective satisfaction** – The level of satisfaction, or how much they enjoyed working with the system or parts thereof.
- The measurements used most often are those related to time, which could include time taken to complete a task, rate of errors and attitude towards the system (Preece et al., 1994).
- The simplest, easiest and most cost effective usability metric is measuring the success rate, that is, whether users are able to complete the task successfully or not (Nielsen, 2001b).

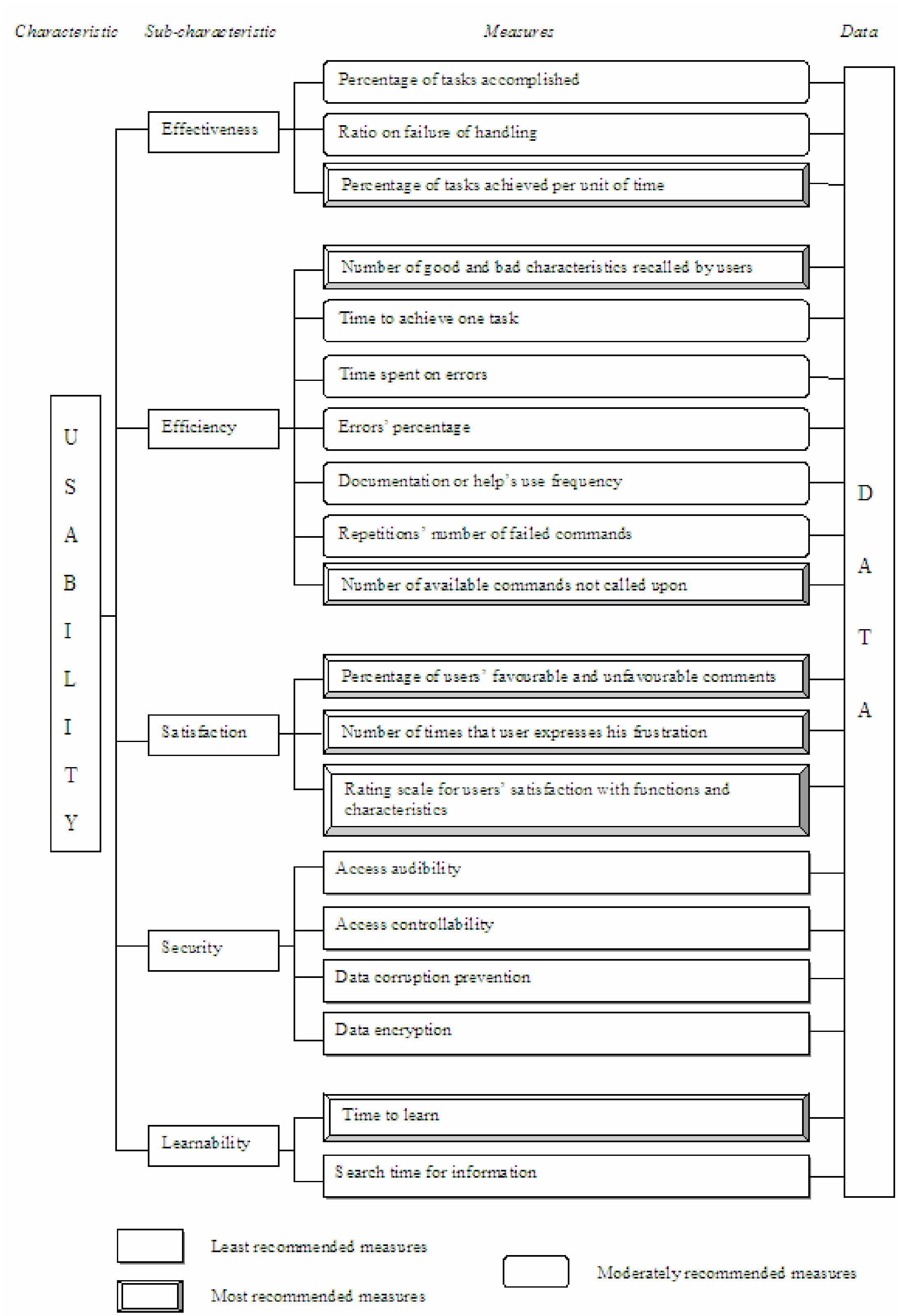


Figure 3.1: Consolidated usability model

Source: Abran et al. (2003a; 2003b)

- Bohmann (2000) indicates that most usability studies need only evaluate the time taken to complete a task or a set of tasks and the number of errors per task.
- The number of actions carried out by the user during completion of a task are seldom part of an informal usability study but may form part of a formal usability study (Bohmann, 2000). The number of actions can be calculated as the number of keystrokes, critical paths etc. (Bohmann, 2000).
- Nielsen (2001a) lists the basic usability metrics as success rate, time to complete a task, error rate and subjective satisfaction.
- Several sources stress that for a comprehensive usability study to be conducted it is highly recommended that at least one measurement be captured for each of the three usability components of efficiency, effectiveness and satisfaction (for example, Bevan and Macleod, 1994; ISO, 1998; Scholtz, 2004).
- Error rate is a classic measurement of the efficiency of a system, based on the belief that a system that allows a task to be completed with minimal errors being made allows for faster and more efficient use (Faulkner, 2000). An error is defined as something that has gone wrong, either by means of a wrong decision being made or a physical mishap (Faulkner, 2000). Errors can be measured in terms of the number of users completing the task successfully and the ratio of successes to failures (Bohmann, 2000).
- Time is another classic usability measure (Faulkner, 2000) and can be subdivided into the time taken to complete the task (Bohmann, 2000; Faulkner, 2000), the time spent on errors (Bohmann, 2000; Faulkner, 2000), the percentage of tasks completed (Bohmann, 2000) and the task success rate (Bohmann, 2000).

In summary, the most important points to consider when deciding on the usability measurements to include in the usability analysis are the following:

- It is considered obligatory to select at least one measurement for each of the usability components of effectiveness, efficiency and satisfaction when attempting to evaluate the usability of a product (Bevan and Macleod, 1994; ISO, 1998; Scholtz, 2004).

- Measurements of time and error rates are valuable measurements to include in the analysis (Bohmann, 2000; Nielsen, 2001a), as are success rates (Nielsen, 2001b) and possibly the number of actions when conducting formal user tests (Bohmann, 2000).
- The workload of the users comprises not only physical but also mental facets of the tasks and the workload should be factored into the usability measures (Bevan and Macleod, 1994; ISO, 1998). Furthermore, the cognitive resources which are required by the users during completion of the tasks can also be measured (ISO, 1998).
- The measurements can be evaluated per task or per set of tasks (Bohmann, 2000).

Data obtained from the interaction of users with the system should be used when measuring usability (ISO, 1998). This data can be collected either via objective – for example, capturing events – or subjective means, such as surveys and questionnaires completed by the users (ISO, 1998). As is evident by the types of objective and subjective measures, effectiveness and efficiency are measured by objective means and satisfaction by subjective means (ISO, 1998).

All the aforementioned guidelines were used to determine which usability measurements were required to ensure comprehensive and valid user testing results for the current study. The usability measurements included in this study were score, with a built-in evaluation of the workload required to complete each task (Section 4.9.1), time to complete each individual task (Section 4.9.2), number of actions required to complete each individual task (Section 4.9.3), the number of errors incurred by the users during completion of each individual task (Section 4.9.4), the ratio of successes to failures for each individual task (Section 4.9.5) and the subjective satisfaction of the users (Section 4.9.6). By using these measurements it was possible to measure the effectiveness and efficiency of the application and the satisfaction experienced by the user during interaction with the application.

3.4.1.3.6 User tests as experiments

User testing provides a means to compare alternate versions of a design element (Galitz, 2002) with the purpose of establishing which design is the better one (Preece et al., 1994). This methodology is known as competitive user testing, competitive analysis or comparison testing and consists of running comparable tests, such as empirical user testing, against previous versions of the application or similar products which are also available (Shneiderman, 1998; Nielsen, 1992) to determine which design best suits the needs of the users (Levi and Conrad, 1998).

Comparison testing is a typical scenario for using classic experiments in HCI (Galitz, 2002). When conducting an experiment, all the variables excluding the variables of interest, must be controlled for in the design of the experiment (Preece et al., 1994). The variable of interest, known as the dependent variable, is expected to be influenced by the independent variables which can be manipulated by the experimenter (Preece et al., 1994). When comparing designs, two or more prototypes are developed which are the same in all respects apart from the design element which must be tested (Galitz, 2002), and which provides the manipulable variable in the experiment. By control of the environment, limiting the number of independent variables, careful measurement of the dependent variables and performing statistical tests, it is possible to test the validity of a previously stated hypothesis (Shneiderman, 1980).

The bulk of the experimental design consisted of comparison testing with the different interface components providing the dependent variable.

3.4.1.3.7 Collecting data

Data collected during evaluation of a system can be either qualitative or quantitative in nature (Preece et al., 1994; Trochim, 2002). Quantitative data is concerned with user performance and attitudes which can be measured in numerical format (Preece et al., 1994). Qualitative data is non-numerical and is usually takes the form of reports and opinions (Preece et al., 1994) or text (Trochim, 2002).

Nielsen (1992) lists some of the more widespread methods of obtaining usable data for all measures of empirical testing as:

1. observing users while they complete a set of tasks on the system;
2. automatically capturing the interaction of the user with the system; and
3. having users complete questionnaires where they rate their experience on a rating scale.

Observation involves watching the users interact with the system (Faulkner, 1998). Two methods exist for logging of user actions and performance during interaction with the system (Faulkner, 1998). The actions can be logged by hand by a facilitator, which is the simplest method of data capturing (Redmond-Pyle and Moore, 1995) and consists of the facilitator physically measuring and recording all desired measurements (Faulkner, 2000). Logging can also be done automatically via the system itself (Dumas and Redish, 1999; Faulkner, 1998; Faulkner, 2000). Automatic logging is preferable since logging done by a facilitator can interfere with the way a task is completed by the user (Faulkner, 1998). Logging by the system is less intrusive than that done by a facilitator and is recommended by a number of usability guides (for example, Cox and Walker, 1993; Faulkner, 2000; Preece et al., 1994).

All three of the above-mentioned methods were used in this study. (1) When designing alternative icons, the initial developed set was first tested with a large group of users. Observation of the interaction allowed for a number of problems to be identified. Early identification of some usability problems allows developers to redesign these aspects of the system and in so doing continually improve on the design via iterative designing (Nielsen, 1992). (2) Automatic logging of user actions was used to capture a number of usability measurements in order to analyse the interaction of the user with the system using objectively captured methods without interference or manipulation by the user. This method has been employed successfully in previous studies as it provides a non-intrusive method of capturing the interaction of the user with the system and in so doing does not distract the user from the task at hand (for example, Zammit, 2000) (3) Questionnaires using rating scales were used to determine the subjective satisfaction of the users as well as the anxiety and the attitude levels of the users.

3.4.1.3.8 Data analysis

The final step in usability analysis is analysis of the collected data. Since both experimental and survey methods require statistical analysis of the collected data (Olivier, 2004) the statistical analysis methods applicable to the analysis of the data needed to be determined.

Statistics are used to analyse the data collected from an experiment or investigation. Statistics can either be descriptive or inferential (Dowdy and Weardon, 1983). Descriptive statistics are only applicable to the sample or the population from which the data is selected (Urdan, 2005) and are used to describe the results of the experiment or investigation (Dowdy and Weardon, 1983).

Inferential statistics are used when part of the population is not available for evaluation (Dowdy and Weardon, 1983) and then the results of the statistical analysis are then used to infer or generalise to the whole population (Urdan, 2005; Walsh and Ollenburger, 2001).

Inferential statistics could be used to analyse the data collected for this study, especially since some user tests are particularly appropriate for inferential statistics to be conducted on the collected data, for example, comparison of two designs (Dumas and Redish, 1999).

Measurement consists of observing and recording observations (Trochim, 2002). Four levels of measurement are available for recording, namely nominal, ordinal, interval and ratio (Trochim, 2002; Urdan, 2005). Nominal measurements simply name the attribute and there is no particular significance associated with the numbers that are used (Trochim, 2002). Ordinal scales refer to rank-ordered observations (Dowdy and Weardon, 1983; Trochim, 2002) and the values used have weight (Urdan, 2005). Ratio and interval scales use measures that have equal distance between the units (Urdan, 2005). Ratio measurements are those where there is a meaningful absolute zero, such as with weight or temperature (Trochim, 2002; Urdan, 2005).

A number of factors need to be considered to determine which statistical test is the most appropriate for analysis of the data (Olivier, 2004). The first consideration is to determine whether the data is normally distributed or not (Olivier, 2004). Should the data be normal, then a parametric test can be applied, otherwise a nonparametric test must be used (Olivier, 2004). However, many of the parametric tests can be used for suitably large data sets, regardless of whether the distribution of the data is known or not (Olivier, 2004). The second factor which determines which statistical test must be used is the number of variables. When the design has only a single independent variable it is analysed by means of univariate statistics; when more than one independent variable is present then multivariate statistics are appropriate for analysis. The experimental design also influences the type of statistical test that is used to analyse the data. Three well-known and often-used experimental designs are the independent subject design, the matched subject design and the repeated measures design (Preece et al., 1994). In the independent subject design, the group of participants identified for inclusion in the study are randomly assigned to two or more experimental conditions (Preece et al., 1994). The matched subject design requires that each subject be paired with another, after which each pair is randomly assigned to an experimental condition (Preece et al., 1994). The repeated measures design requires that all subjects feature in all the experimental conditions (Preece et al., 1994).

T-tests

A t-test is used to analyse a design with one categorical or nominal independent variable and one continuous or interval scaled dependent variable, provided that the data is normally distributed (Berenson and Levine, 1979; Dowdy and Weardon, 1983; Urdan, 2005). A t-test can either be an independent samples t-test or a dependent samples t-test (Berenson and Levine, 1979; Dowdy and Weardon, 1983; Urdan, 2005).

Since both the independent and dependent samples t-tests rely on the fact that the data must be normally distributed, a nonparametric equivalent to both the t-tests must be identified for use should the data violate the normality assumption. The Mann-Whitney U test is the most powerful alternative to the independent samples t-test (Statistica Electronic Manual, 2005). The Wilcoxon matched pairs is the nonparametric alternative to the dependent samples t-test (Berenson and Levine, 1979; Statistica Electronic Manual, 2005).

Analysis of variance

One-way analysis of variance (one-way ANOVA) is used when there is one independent variable, which is distributed amongst one or several groups, and one dependent variable (Berenson and Levins, 1979; Urdan, 2005). The factorial ANOVA is used for more than one independent variable, each distributed amongst one or several groups, and one dependent variable (Urdan, 2005). Both the one-way and the factorial ANOVA are parametric tests since they assume normality of the data (Dowdy and Weardon, 1983).

The nonparametric alternative to the one-way ANOVA is the Kruskal-Wallis ANOVA by ranks and the Kruskal-Wallis median test (Berenson and Levine, 1979; Statistica Electronic Manual, 2005).

Chi-square test

A contingency table lists the frequency of certain occurrences, for example the number of successes and failures for a set task and for each group which appears in the sample (Berenson and Levine, 1979). The Chi-square test evaluates the difference between the proportions found in the contingency table (Berenson and Levine, 1979). Since the Chi-square analysis only works with frequency of observations it makes no assumption about the distribution of the data.

3.5 Summary

By following well-established procedures, the research methodology that had to be followed for successful completion of the research study was determined. The research problem pertinent to this study was focused on investigation of real-world objects and was therefore empirical in nature. Since the aim of the study was to determine whether or not language, icon type and icon sets had an effect on the usability of a word processor, the research question was causal in nature. A causal question requires collection of data by means of experiments, surveys or comparative studies. The first two methods were used to collect new or primary data to evaluate the research question. A nonprobability convenience sample was used for all three groups tested during the study. The first sample comprised high school students who were readily accessible in the area in which the study was conducted. The second and third samples were both used due to their availability via the university where the study was conducted. The second sample consisted of a group of cleaners employed by the university and the third sample was made up of students who had registered for a specific course. The methods of surveys and experiments were confirmed as viable methods for research in information technology.

Surveys could be administered either as questionnaires or as interviews. Questionnaires were found to be most appropriate to establish user demographics and to measure anxiety and attitude levels as well as for measuring user satisfaction. Experiments in HCI were identified as being mostly usability evaluation techniques. The various evaluation techniques were discussed and user testing was found to be the most relevant to the study. Therefore, user tests were employed to compare the different designs with one another to establish which one was the better design. Observation, automatic logging and questionnaire analysis were identified as feasible methods of data capturing and available statistical analysis techniques were identified and discussed.

The research methodology has therefore been determined and the following chapter will give an in-depth discussion of how this research methodology was implemented in the study. The entire experimental design will be discussed based on the research methodology and the findings discussed in Chapter 2.

Chapter 4

Experimental design and methodology

4.1 Introduction

The previous two chapters provided the theoretical foundation for the study and the research methodology on which the study was based. The research methodology techniques that were relevant to this study were determined to be questionnaires and user testing. Observations, automatic logging of user actions and questionnaires were identified as possible means of implementing user testing. Once the theoretical background for the research design had been established, the methodology had to be applied and implemented in a practical manner which was suitable to the current study. This chapter outlines the experimental methodology followed in this study by discussing how the theory was applied to the study at hand. The in-depth discussion will focus on the development of the test instrument and the test variables, in terms of icons, menu and tooltips. A detailed description of the test method that was followed and the measurements that were captured and analysed will also be given.

4.2 Objectives

The objectives of this study were threefold (Sections 1.3 and 1.4). In broad terms, these objectives were:

1. to test whether the images depicted on icons used in an application influence the usability of that application;
2. to test whether pictorial icons or text buttons are more conducive to easy usage of an application; and
3. to determine whether translation of an interface into the first language of the end-user facilitates ease of use and ease of learning of the application.

In order to test these three objectives, the following procedure was followed:

1. A set of alternative icons which could be compared to the standard icons, as found in the current Microsoft Office package, were designed and developed.
2. To test which type of icon, pictorial or textual, is the most effective, a set of text buttons was developed to compare user performance when working with pictorial or textual icons.
3. To determine whether translation of the interface into the first language of the user had an impact on the usability of the application, the interface was implemented in three different languages via the text buttons, menu and tooltips.

4.3 Test instrument

The application environment chosen for the evaluation of the above-mentioned objectives was a word processor. This application type was chosen for a number of reasons, not least of which was the popularity and diversity of interaction afforded by such an application. From the early use of graphical user interfaces (GUIs) it was evident that the word processor had become an integral part of everyday life and has since then proven to be one of the most used applications (Roberts and Moran, 1983), with users spending the majority of their time using editors or editor-based applications (Rosson, 1984a). Objective and subjective methods of evaluation, as defined in a preceding section (Section 3.4.1.3.5), can also be successfully used to measure the "quality of the user-editor interaction" (Roberts and Moran, 1983). In other words, usability measures are applicable when establishing the usability of a word processor or text editor. The different user categories, as defined previously, are also applicable to word processor users (Roberts and Moran, 1983). Word processing studies can be classified according to their goals (Sullivan, 1989), with this study falling into the research category of improving the user interface design of word processing applications.

4.3.1 Word processor application

A small word processor application was developed which incorporated minimal functionality, whilst still being representative of a fully-fledged word processor or advanced text editor (Figure 4.1). Common word processor functions which were included were document handling (for example, **Open**, **Close** and **Save**), text manipulation (**Cut**, **Copy** and **Paste**), font formatting (**Font name**, **Font size** and **Font colour**), font styling (**Bold**, **Italic** and **Underline**) and paragraph formatting (**Left align**, **Right align** and **Centre align**).

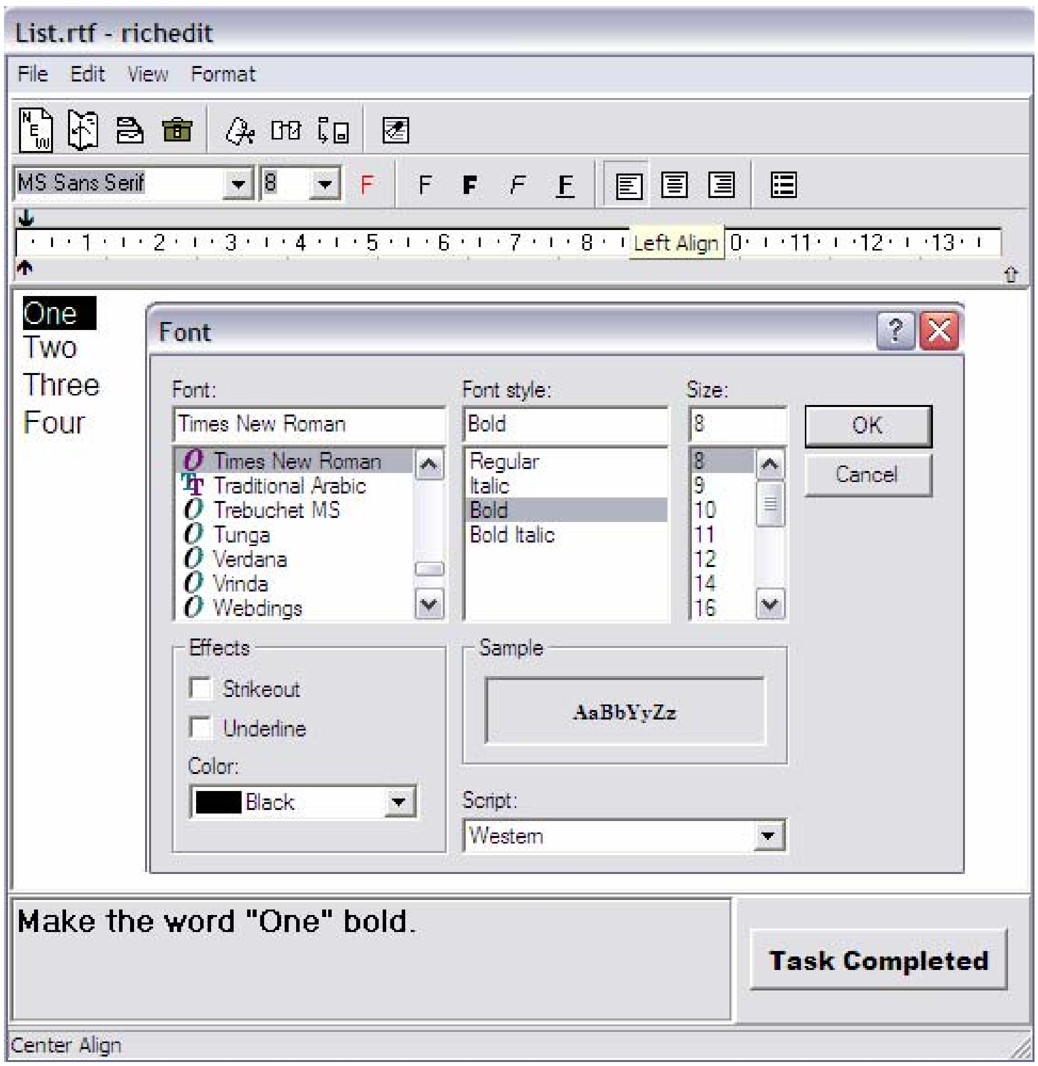


Figure 4.1: Word processor application with alternative icons and English menu

A single standard application was used after which a number of different interface options were developed which could be used interchangeably with the application.

The interface used in the word processor application could only be changed via a password-protected menu option. The password was known only to the researcher.

The application allowed for selection of a number of different tests which had been loaded into a database beforehand. Test subjects were required to complete a test which consisted of a set of predetermined word processing tasks. Tasks included in the selected test were then loaded sequentially and individually in a separate pane situated at the bottom of the application window (Figure 4.1).

The functionalities present in the application were all represented by action icons (ISO, 1999) and were selected due to their common use in word processing applications. Using the action icon classification as supplied by the ISO (1999), the functionalities which were incorporated into the application were as follows:

Table 4.1: Application functionality

Icon type	Word processing function
Filing action icons	New
	Open
	Close
	Save
Editing action icons	Cut
	Copy
	Paste
Command reversal action icons	Undo
Text action icons	Font colour
	Bold
	Italic
	Underline
	Left align
	Centre align
	Right align
	Bullets / Itemise

All functions that were available for selection via an icon were also available for use via a menu option, when a menu was present in the interface.

4.3.2 Data collection

The word processor allowed for real time evaluation of the tasks. This meant that once the user had completed the task, the application immediately evaluated the result in order to determine if the task was completed successfully or not. Apart from the task result, a number of measurements were also captured for each task. These measurements were:

- Which menu options were selected and in which order they were invoked;
- which toolbar icons were selected and in which order they were invoked;
- the number of keystrokes;
- the number of mouse clicks; and
- the time taken from the start of the task to when the user indicates the task has been completed.

From these measurements, the required usability measurements (Sections 4.9.1 - 4.9.6) could be extracted.

4.3.3 Integrating the data collection

All data collection was conducted electronically and was performed in the background. The automatic logging of user action and responses allows the participants to interact with the application without interference for data collection purposes (Cox and Walker, 1993; Faulkner, 2000; Preece et al., 1994).

All questionnaires and tasks were integrated into a single application and users were led from one part of the test to the next without any additional effort on their part. The application flowed from one part of the test into the next by simply displaying the required part of the test in the current window. Designing the system in this way allowed for the impression to be presented that the entire test was a single unit and not actually a number of smaller tests built into a larger test.

In this way the application allowed for capturing of the users' demographic details (Section 4.5.1), the measuring of user anxiety (Section 4.6) and attitude (Section 4.7), capturing of usability measures (Sections 4.9.1 – 4.9.5) and subjective satisfaction (Section 4.9.6) with the application.

4.4 Interface

The four possible general interface configurations which were identified in terms of the objectives (Section 4.2) and tested were:

1. Pictorial icons with no tooltips and no menus;
2. Pictorial icons with tooltips but no menu;
3. Pictorial icons with tooltips and a menu;
4. Text buttons with tooltips but no menu.

The subsequent development process of these interface configurations and resultant subdivision of the general configurations will be discussed at length in the following sections.

4.4.1 Icons

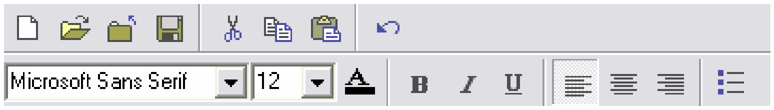
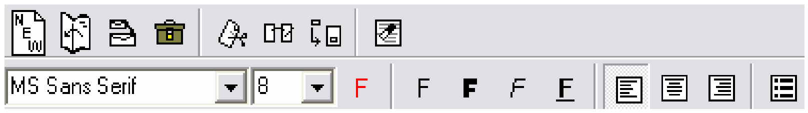
Three sets of icons (Table 4.2) were used in the different interfaces, namely (i) the standard icons found in the Microsoft Office package, (ii) an alternative set of icons obtained from previous studies (Teklebrhan and Blignaut, 2005) and via two brainstorming sessions (Section 4.4.1), and (iii) text-based icons (Section 4.4.1).

The icon toolbars used were divided into the standard and formatting taskbar as found in Microsoft Office. The two toolbars were displayed separately with the formatting toolbar situated directly below the standard toolbar (Figure 4.1).

Iterative development of the icons

Development of the icons was done using an iterative method for each individual icon set. Once the icons were placed on the standard and formatting toolbars and the text buttons were translated into the three languages, the standard, alternative and text interfaces appeared as follows for the final testing of the interfaces:

Table 4.2: Icon sets

Standard interface	
Alternative interface	
English text interface	
Afrikaans text interface	
Sesotho text interface	

Standard icons

The set of standard icons were obtained by using the existing icons found in the Microsoft Office package currently available on the market. There were, however, two changes made to the current toolbar, both concerning the **Font colour** icon.

Firstly, this icon was shifted next to the drop-downs of the **Font name** and **Font size**. This was done in order to group similar functions together as recommended by ISO guidelines (1999) and other authors (Ellis et al., 1995; McDougall and Curry, 2004). Grouping of similar icons into a set allows for inferences about the category which then aids the interpretation of the individual icon within the group (McDougall and Curry, 2004). The second change was more of an omission than a change and was implemented for the sake of simplicity. The line under the **Font colour** icon consistently remained black instead of changing to the colour selected from the font colour dialog box.

Alternative icons - first iteration

An initial brainstorming session was held where participants suggested a number of alternative icons which could possibly be used in place of the standard icons. The resultant set of icons was then distributed amongst potential word processor users. Respondents were required to indicate which icon they would choose for each of a number of listed word processor functions. Alternative icons for **Open**, **Close**, **Save**, **Cut**, **Copy** and **Paste** were determined in this way. There were two icons chosen by the same number of respondents for the **Close** function. Therefore, the icon for **Close** was selected by a process of elimination since one of the icons was also chosen as the preferred **Save** function by a large margin. The alternative icons for **Open**, **Cut** and **Copy** were confirmed as the preferred icons in a second study where respondents were given a set of icons for each of these functions and were required to select their most preferred icon from the respective sets (Teklebrhan and Blignaut, 2005).

The remainder of the icons were developed during a second brainstorming session. During this session a number of alternatives were developed for each function and then the icon that was considered to be the best was chosen to be included in the interface. The best icon was decided upon by evaluating the detail of the icon, the context it provided and whether it succeeded in depicting the icon function. Due to time constraints and lack of available test subjects, confirmation by non-computer literate users of these icons was forgone and the icons were included in the design as is.

The icons were developed to provide more context for novice users, as recommended in the literature (Lodding, 1983). Icons such as bullets and the three paragraph alignment icons used the same principle as the standard Office icons but were placed within the context of a document or page by placing a rectangular border around the icon image.

The font styling icons (**Bold**, **Italic** and **Underline**) were designed to show the transition between the regular font style and the resulting style of the font once the icon has been pressed. So for example, the **Underline** button consists of a small letter *a* in regular font followed by a small letter *a* that has been underlined. An arrow was included to indicate that there is a transition between two phases that will occur when the icon is selected. One problem that manifested itself with these design decisions was the size limitation of the icon which severely constrained the amount of detail that could be placed on the icon. Once it was placed on the toolbars, the final icon set looked like this:

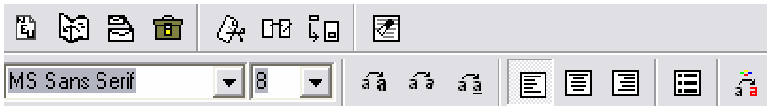


Figure 4.2: First iteration of alternative interface

Testing the first iteration of alternative icons

The set of icons shown in Figure 4.2 was tested with a group of learners ranging from Grades 8-11. The majority of the learners had no experience with a word processing application other than the few lessons they had been given in a typing class. However, the class did not cover actual usage of a word processing application, so these users may have merely been familiar with the look of the application but had no knowledge of the actual workings or functionality provided by the application.

From simple observation of the test participants as well as verbal feedback from the users it became apparent that the font styling icons were not a successful design choice as participants struggled to understand and interpret the concepts depicted on the icons. Users did not realise that the arrow included in the icon was there to indicate that a transition or change occurs to the text when the icon is selected. There also appeared to be confusion between the standard icons for Font colour and Underline. As established earlier, this type of formative evaluation serves well to identify usability problems early on in the design (Nielsen, 1992). Apart from the font styling icons, no other icons appeared to present a problem to the users.

Due to the problems experienced by these users when using the font styling icons, these icons had to be redeveloped.

Alternative icons - second iteration

The icons which caused the most difficulty – those of Bold, Italic and Underline – were changed to the second choice from the brainstorming session as it seemed that the idea of showing the transition of font styling was unsuccessful.

The second choice icons were also developed with the purpose of placing the font styling in context and displaying the transition or different effect of the font styles. However, instead of the individual icons showing the transition, the font styling set as a whole was used for this purpose. Each of the icons merely showed the result of the font style associated with the icon shortcut. This design used the same general idea as the standard Office icons. However, in contrast to the standard Office icons that apply the font styling to the first letter of the function, for example, a bold letter “B”, the alternative icons all used the same letter, a capital “F”, throughout and simply applied the different font styles to the letter.

The capital letter “F” was chosen as the letter which most clearly displayed the changes in font styling when placed side by side. Therefore, the second iteration for the icons for Bold, Italic and Underline consisted of a bold, italic or underlined capital letter “F” respectively. A regular icon was also included on the toolbar which, when invoked by the user, would remove all font styles from the selected text.

Using a single letter, with different font styles applied and placing them adjacent to one another on the toolbar allows for easier visualisation of the font styling (Figure 4.3). This ability to group logically similar icons into a single physical coherent set is one of the distinct advantages present in iconic and menu-based interfaces (Gittins, 1986). Furthermore, it reduces the amount of resources required by the user in order to locate the correct icon since the set is processed as a single unit (Hemenway, 1982), which is the reason why grouping of similar icons into a single set by means of placing the icons together in a row, column or box, is a recommended guideline when designing usable interfaces (Bickford, 1997; ISO, 1999). Special care must be taken when designing these icons to ensure that it is evident that each individual icon in the set does indeed belong to the group (Lodding, 1983). Thus the font styling icons were designed using the same letter to indicate related functions, without removing the ability of the icon to function as a stand-alone icon. That is, it is evident that the Italic icon belongs to the group (further reinforced by the separation bar), but when viewed on its own it still manages to convey the italic styling effect. The increased meaningfulness of the icon could also assist the user in determining the correct functionality of the icon.

All of these aspects were taken into consideration when designing the font styling icons in an effort to convey to the user the font changes that would occur if the function were invoked. It was hoped that by developing icons in this manner, novice users would easily relate to the concepts depicted by the icons.

F *F* F

Figure 4.3: Alternative font styling icons

The Font colour icon was also moved to appear alongside the other font functions (the name and size drop-downs), instead of appearing at the right-hand end of the toolbar. The second set of alternative icons as they appeared on the toolbars for the final tests are shown in Table 4.2.

Text buttons

The next set of icons that were developed could technically not be considered to be icons as they did not contain an image of any kind. Instead they consisted of buttons displaying the name of the function they represented. Although the strict definition of an icon precludes the sole use of alphabetic characters as a depiction, the icon taxonomy used in this study did make provision for icons which consist only of alphabetic characters or text. Therefore, these buttons could be included in the study as an icon set.

The text buttons had the first letter of each function printed in bold. The original intention was to ease the transition between text buttons and the standard icons and then possibly provide training to first-time and novice users on the text buttons. The Bold, Italic and Underline buttons would then have a bold "B", italic "I" and underlined "U", with the result that they would appear to be similar to the standard icons for Bold, Italic and Underline. This concept would, however, only work for these three functions and a trial interface with these buttons in this format resulted in a distinct disruption of the uniformity of the icons. Since consistency is of paramount importance when creating a usable product (Nielsen, 1992), all the buttons were designed with the first letter of the function as a bold capitalised letter (Table 4.2). Current interfaces underline the shortcut key in the menu structure, for example, the New menu item has the function name printed in the following format: New. Therefore, the keyboard shortcut of the **New** menu item is Ctrl+N. The distinction between a bold capitalised letter and an underlined letter was considered to be clear enough not to confuse the users into thinking that the bold letter represented a keyboard shortcut. The intention was to increase the readability of the icons as the bold first letter clearly demarcates the start of the next button and function names do not blend together.

4.4.2 Language

Because of the diverse nature of the South African population where there are eleven official languages, it is impossible to accommodate all these languages in a single study of this size. Since the largest test user base came from a university of dual English and Afrikaans medium instruction, it was decided to have all interfaces in both English and Afrikaans. Furthermore, a large majority of these students were native Sesotho speakers; therefore Sesotho was added as an additional interface language.

The language was implemented in terms of menus, tooltips and text buttons (Table 4.2). Each of the interface components was implemented in the three afore-mentioned languages. The Sesotho translations were obtained from existing software currently available on the market that had already been translated into Sesotho. Where the translation for a particular function was not available, a Sesotho language expert provided the necessary translation.

Users were tested on an interface either in their first language or in English, considered for these purposes as their L2 (Section 2.8). Since the interface was only available in English, Afrikaans and Sesotho, the participants were divided into one of these three groups according to their L1. Afrikaans users completed the test on either an Afrikaans (L1), or an English (L2) interface. Sesotho users completed the test either in Sesotho (L1), or in English (L2). For the purposes of this research Setswana speaking users completed the test either in Sesotho (L1), or in English (L2), as Setswana is widely regarded to be very similar to Sesotho and does in fact belong to the Sotho language group (Lye and Murray, 1980). The remainder of the users completed the test in English, where English would be either their L1 or L2. The graphical division of the language is depicted in Figure 4.4.

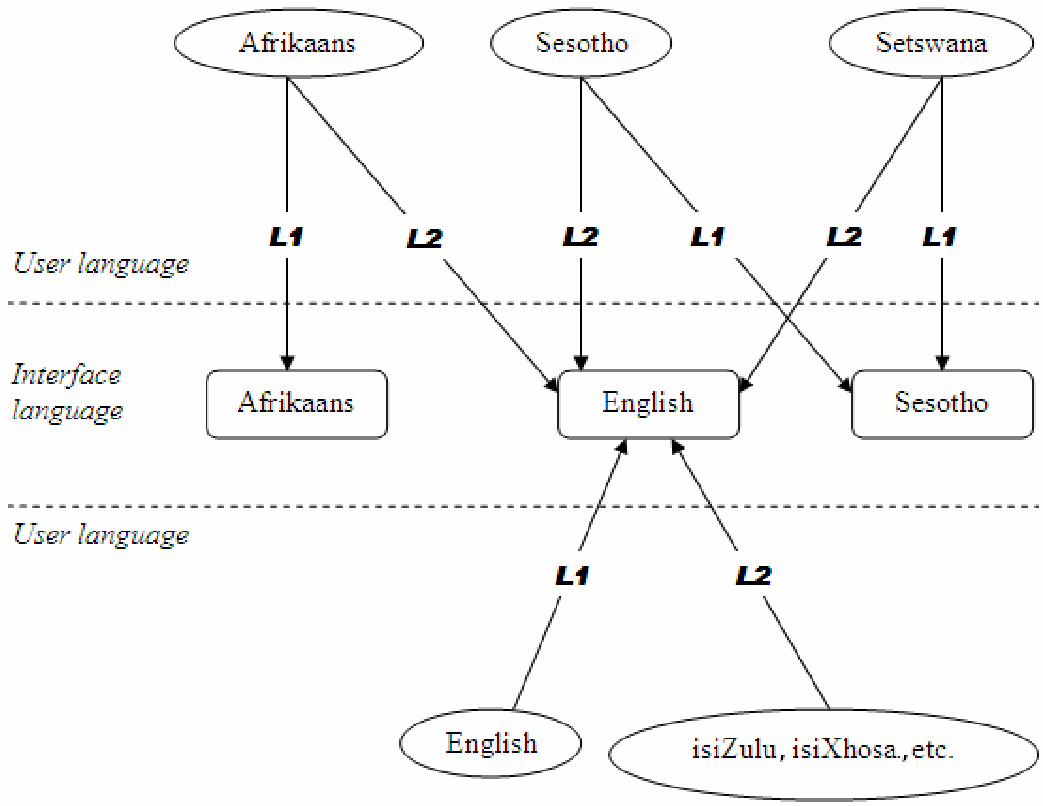


Figure 4.4: Language division of the users amongst the interfaces

4.4.3 Menus and tooltips

When menus were included in the interface, the menu structure used was the same as that found in Microsoft Office. The menus were implemented in English, Afrikaans and Sesotho.

A tooltip is a small label that appears below an icon when the mouse pointer is positioned or moved over that icon. The label contains explanatory text for that icon to assist the user in deciphering the function of the icon. The tooltips were implemented in English, Afrikaans and Sesotho. Where the interface contained a menu and tooltips, these two interface components were implemented in the same language. The tooltips appeared immediately when the mouse pointer moved over the icons, instead of after a short delay as is common practice in current software applications.

4.4.4 Interface sets

According to the icon classification of Zammit (2000), the standard and alternative icons were abstract pictorial icons, while the text-based icons fell under the word subsection of text icons as either a nominal or a verb icon. To summarise, the icons used in the study for each of the functions are shown in Table 4.3.

Table 4.3: Icon iterative development

	Function	***Standard icon***	***Alternative icon***		***Text button***
			First iteration	**Second iteration**	
Standard toolbar	**New**				New
	Open			*	Open
	Close			*	Close
	Save			*	Save
	Cut			*	Cut
	Copy			*	Copy
	Paste			*	Paste
	Undo				Undo
Formatting toolbar	**Font name**				Font Name
	Font size				Font Size
	Font colour			F	Font Colour
	Regular			F	
	Bold	**B**		**F**	Bold
	Italic	*I*		*F*	Italic
	Underline	U		F	Under line
	Left align				Left
	Centre align				Centre
	Right align				Right
	Bullets				Bullet

* Icons chosen by non-computer literate respondents

Following the example of previous studies (Benbasat and Todd, 1993), the placement of the icons on the toolbar and other interface components such as menu items, remained constant between the different interfaces used in this study, with only the actual icons and languages changing (see Table 4.2).

To enable the effect of the icons to be tested without interference from other interface components, each pictorial icon set was included in an interface with neither menus nor tooltips. This ensured that the user had to rely entirely on interpretation of the icon when using this interface. The next group of interfaces to be added to the test interfaces used the same pictorial icon sets but tooltips were added in English, Afrikaans or Sesotho. To complete the set of interfaces for testing, the afore-mentioned interfaces were used as a base to which a menu was added in the same language as the tooltips for that particular interface.

The text buttons were never used in conjunction with a menu but always used together with tooltips. This was to compensate for the fact that very often the entire function name did not fit on the button; in particular the Sesotho translations were too long. To ensure legibility of the icons, a shortened version of the function was placed on the button and the full-length version was displayed in the tooltip to provide verbal context (Figure 4.5). This phenomenon is one which is repeatedly mentioned in the literature as a guideline towards localisation and translation, where warnings are issued that the translated language may require more space than the original language and that provision must be made for this possibility (Ott, 1999).

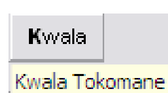


Figure 4.5: Sesotho text icon for Close **with expanded tooltip**

To summarise and with reference to the general interface configurations described earlier (Section 4.4) which can now be subdivided into more specific categories, the following interface configurations were available for testing:

1. Standard icons with a menu and tooltips;
2. Standard icons with tooltips but no menu;
3. Standard icons with neither a menu nor tooltips;
4. Alternative icons with a menu and tooltips;
5. Alternative icons with tooltips but no menu;

6. Alternative icons with neither a menu nor tooltips;
7. Text buttons with tooltips but no menu.

Of these, interfaces 3 and 6 were the only two that had no language component since there was neither a menu nor tooltips available for these interfaces. The five remaining interfaces were each implemented in English, Afrikaans and Sesotho, resulting in a total of 17 interfaces. Since users worked with an interface in either their L1 or L2, there were eventually 12 distinct interface groups to be tested – the five language interfaces all in either L1 or L2 and interfaces 3 and 6, which had no language attached to the interface.

4.5 User profile

4.5.1 User demographics

The application incorporated a screen for the capturing of the users' demographic information at the start of the test. These demographics included aspects such as age, gender, qualifications, word processor experience and frequency of use as well as the socio-economic standing of the user.

Each user was assigned a unique identifying number, known as the student number, which formed part of the users' demographic details. Hereafter, the user was identified by that number. This number facilitated easy identification of the user and afforded anonymity of the participants. Moreover, since all information was stored in a centralised database, this prevented users from entering their personal information multiple times.

4.5.2 Classification of users

Using the definitions and classifications provided by Shneiderman (1998), this study identified 4 types of users, namely:

1. Novice users
2. First-time users
3. Intermediate users
4. Expert users

For the purposes of this study, novice users were defined as those having no knowledge of either the task or interface domain of a word processor. The rest of the users were those users who had a working knowledge of the task domain and as such could be classified as first-time, intermediate or expert users. This group of users will henceforth collectively be referred to as the task-knowledgeable users. Since expertise with a software application, and therefore a word processor, is a function of both the length and frequency of use (Rosson, 1984b), their interface domain knowledge was determined by means of two scales which measured their word processor experience and frequency of use prior to the test. A combination of the two scales then determined their classification into one of the three user groups.

The experience the user had gained on a word processor prior to the administration of the test was measured on a scale of 0 to 5 and was defined as the length of time that they had been using a word processor, with possibilities ranging between months and years. The frequency of use of a word processor was measured on a scale of 0 to 4, with 0 being the lowest frequency possible. The frequency was defined as the regularity of use of a word processor. The experience and frequency scales were then cross-multiplied to obtain 30 rankings which were referred to as the expertise of the user. The expertise rankings obtained after cross-multiplication are tabulated in Table 4.4.

Table 4.4: Cross-multiplication of frequency and experience

		Frequency scale				
		0	**1**	**2**	**3**	**4**
Experience scale	**0**	0	0	0	0	0
	1	0	1	2	3	4
	2	0	2	4	6	8
	3	0	3	6	9	12
	4	0	4	8	12	16
	5	0	5	10	15	20

A new expertise scale was constructed by using the cross-multiplied values and eliminating duplicate values. Therefore, by using only the *distinct* expertise ratings, the new expertise ranking was a 14-point scale. Users were then classified according to this scale as first-time (low expertise rating), intermediate (medium expertise range) or expert (high expertise rating) users. The boundaries between first-time, intermediate and expert users was determined in such a way as to give the best distribution between the users. As is common practice (Dumas and Redish, 1999) and to provide the added insurance of eliminating possibly false results, the intermediate group, as the group which is bordered by the less and more experienced users and could contain users who were uncertain of their actual expertise, were removed from the analysis.

In order to test the interfaces comprehensively, all the above-mentioned user groups had to be tested on as many interfaces as possible.

A group of novice word processor users were identified, who incidentally also happened to be novice computer users. These users had never before made use of a computer, and therefore had never used a word processor application, and had had very little exposure to technology in general. With regard to their education level, age and socio-economic standing, these users provided a homogeneous group of novice users. These users will be discussed in greater detail in the Chapter 6.

The second group of users had been exposed to technology and had at least used a computer. These identified users were then used as the first-time, intermediate and expert users of the study. These users had had varied amounts of exposure to computers and word processors and they were classified using the above-mentioned expertise scale. They consisted of high school students and first year university students. These users will be discussed in greater detail in the Chapters 5 and 7.

4.6 Measuring user anxiety

At the start of the test, users were required to complete a questionnaire designed to establish their level of anxiety experienced when working with computers. The instrument used to measure their anxiety was a 20-point questionnaire known as the Computer Anxiety Scale (CAS). This questionnaire was developed by Marcoulides (1989) and divides computer anxiety into two sections. The first section focuses on general computer anxiety and includes questions concerning the everyday use of a computer, the societal importance placed on computers and the influence of a computer in a work-related context. The second factor was named equipment anxiety and focuses more on the physical equipment or hardware, such as printers (Marcoulides, 1989).

Respondents are required to indicate their degree of anxiety in a given situation by selecting the anxiety level from a Likert rating scale which moves from negative responses on the left to positive responses on the right. Each question is then assigned a factor loading based on its contribution to its respective anxiety category.

The original questionnaire designed by Marcoulides (1989) required respondents to rate each question using a 5-point scale resulting in a total anxiety score range of 20 to 100. However, the questionnaire used in this study adjusted the rating scale to a 6-point Likert scale in order to ensure that both the anxiety and attitude questionnaires (see following section) provided the users with the same number of options for answering the questionnaires. This same method of standardising these two specific questionnaires was successfully implemented in previous studies (for example, Blignaut, 1999) and rescaling of questionnaire response scales is common practice (for example, Blignaut, 1999; Dukes, Discenza and Couger, 1989). With this in mind, the new possible ranges of anxiety scores, as calculated using the factor loadings provided by Marcoulides (1989) are set out below:

Table 4.5: Computer anxiety score division

	Minimum anxiety score	**Maximum anxiety score**
Anxiety subsection		
General anxiety	10.06	60.36
Equipment anxiety	3.69	22.14
Overall anxiety	13.75	82.50

Higher scores correspond to higher anxiety for each category and vice versa (Marcoulides, 1989).

The questionnaire was administered to users in electronic format and appeared as an integral part of the application. The questionnaire was displayed to the user in its entirety, with questions being positioned directly below one another. Users then responded to each question by selecting the corresponding rating from a group of radio buttons ranging from 1 to 6 (Figure 4.6). The questionnaire results for each user were stored in a centralised database. The questionnaire as it was administered in the study can be found in Appendix A.

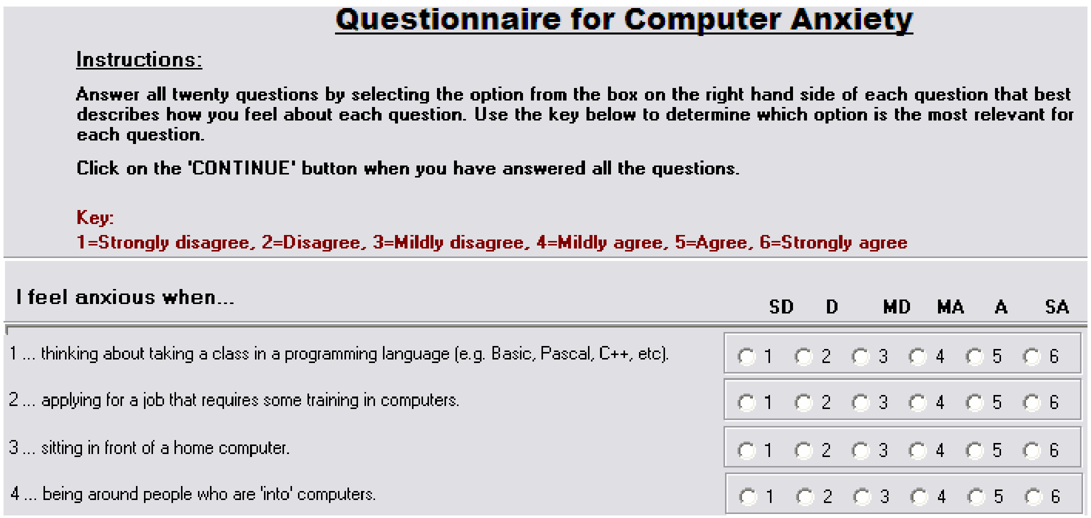


Figure 4.6: Screenshot of questionnaire for computer anxiety

4.7 Measuring user attitude

Test participants were required to complete a questionnaire aimed at ascertaining their attitude towards computer usage.

The questionnaire that was used for this purpose was originally designed as a 30-question scale, known as the Computer Attitude Scale (CAS) (Loyd and Gressard, 1984b). The computer attitude scale subdivided computer attitude into three distinct factors, namely (i) anxiety, (ii) liking and (iii) confidence (Loyd and Gressard, 1984b). Each question was assigned a factor loading according to its respective contribution to the relevant attitude subsection. Bandalos and Benson (1990) later refined the questionnaire, reducing it to just 23 items. The factors were also revised, resulting in the three factors being named (i) liking, (ii) confidence and (iii) achievement (Bandalos and Benson, 1990).

Not only is the CAS suitable for use on a heterogeneous sample but its reliability and validity has been confirmed (Gardner, Discenza and Dukes, 1993; Woodrow, 1991) and efforts to create a superior attitude scale were unsuccessful (Gardner et al., 1993). A comparison between the CAS and three other anxiety or attitude scales showed that the scales exhibit high convergent validity, thus allowing for any of the scales to be administered as a valid and reasonable measurement of computer anxiety and attitude due to the robust nature of the anxiety and attitude construct (Dukes et al., 1989). Furthermore, the scores obtained on the separate subscales are stable enough to be able to use for individual evaluation (Loyd and Gressard, 1984b) and it has been used successfully in a number of previous studies (for example Blignaut 1999; Busch, 1995; Orr, Allen and Poindexter, 2001; Pope-Davis and Twing, 1991).

The questionnaire used a Likert scale that started with a value of 1 which corresponded to an answer of "strongly disagree", whilst the maximum response of 6 corresponded to an answer of "strongly agree". The possible minimum and maximum scores for the attitude questionnaire when using the factor loadings as given by Bandalos and Benson (1990) are shown below. In order to ensure consistency with the anxiety questionnaire, scores were converted to 2 decimals:

Table 4.6: Computer attitude score division

	Minimum attitude score	**Maximum attitude score**
Attitude subsection		
Liking	6.16	36.94
Confidence	7.19	43.12
Achievement	4.46	26.78
Overall	17.81	106.85

The attitude questionnaire had to be answered after completion of the anxiety questionnaire. This questionnaire was also in electronic format and formed part of the application. The appearance (Figure 4.6) and storage of the attitude questionnaire was handled in the same manner as the anxiety questionnaire (Section 4.6). The questionnaire as it was administered in the study can be found in Appendix B.

4.8 Testing the interface

The ISO defines a task as being “the activities undertaken to achieve a goal” and further states that in order to test usability, it is commonly decided to select a set of integral tasks which represent “the significant aspects of the overall task” (ISO, 1998).

Usability tests are widely believed to be the best form of testing for an application (Al-Qaimari and McRostie, 2001; Nielsen, 2003b). Conducting usability tests allows for the determination of whether any particular interface configuration increases the usability of the application. In order to achieve this, a number of simple tasks that were considered to be indicative of common everyday word processing tasks were formulated, as the formulation and administration of representative tasks is considered to be the best method of conducting empirical tests (Nielsen, 1992). Testing a user interface by requiring users to complete a number of tasks that are representative of common tasks, also referred to as benchmark tasks, is a tried and tested method which has been implemented in many studies with great success (for example Benbasat and Todd, 1993; Duncan and Fichter, 2004; Roberts and Moran, 1983).

Each task was assigned a difficulty index based on the minimum number of actions and inferences required by a user in order to complete the task successfully. For example the task “Make the word ‘One’ bold” would consist of the following actions and inferences:

1. Understand the concept of “bold”
2. Select the word “One”
3. Find the correct icon or menu option
4. Click the correct icon or menu option

This results in the above-mentioned task being assigned a difficulty index of 4. The difficulty index of the tasks ranged from 3 to 8.

Two separate tests were compiled, each consisting of a number of individual tasks. The first test had ten tasks which had to be completed and the second had 11 tasks.

The first test was compiled for use by novice users. The novice users who were used in this study had a very specific profile (Sections 6.2 and 6.4) and the test was compiled using specifications to meet their needs. Suffice it to say at this stage that the test was set up using simple and direct language and with the task being very specific in what was required of the user. The tasks for the first test, with their respective difficulty indices, are listed below. Since all interaction with the novice users was in Afrikaans and this was the language with which they were the most comfortable (Section 6.4), the original formulation of the questions was in Afrikaans but for the sake of this dissertation the questions used are reproduced in English.

Table 4.7: Task list for novice users

Task no	**Task description**	**Difficulty index**
1	Open the document named "VRUGTE.rtf"	5
2	Highlight the word "Appel" and make it bold.	4
3	Highlight the word "Perske". Cut the word "Perske" out of the text and paste it above the word "Piesang".	8
4	Highlight the word "Lemoen" and make it italic.	4
5	Close the file.	3
6	Highlight the word "lekker" and underline it.	4
7	Highlight the word "REKENAARKURSUS" and change the font size to 28.	5
8	Highlight the word "REKENAARKURSUS", if it is not already highlighted, and change the font type to Impact.	5
9	Highlight the word "REKENAARKURSUS", if it is not already highlighted, and change the colour of the word to green.	5
10	Highlight the word "REKENAARKURSUS", if it is not already highlighted, and centre align the word.	4
TEST TOTAL		**47**

As is evident from Table 4.7, users were instructed when to highlight or select a piece of text or phrase when it was necessary in order to successfully complete the task at hand. Because it was found during initial interaction with these users that they struggled to select a word, either by means of a double-click or by holding the Shift key in and using the arrows to select, it was decided that the tasks which required selection of a word would be loaded into the application with the required word already selected. The questions did however still include the instruction that the users must ensure that the correct word is selected. As such these tasks have another action added to the difficulty index of the task for the selection, or checking the selection, of the word. Furthermore, although it was not necessary to highlight the text in order to centre align for task 10, since it was specifically relayed to the users in that form and these users tended to do exactly as instructed, task 10 was assigned a difficulty index using the same evaluation as for the preceding tasks where the selection was in fact a required action.

Test 2 had eleven tasks which had to be completed by the users and was designed for use by the task-knowledgeable users. A general description of the tasks, together with their respective difficulty index, is listed below:

Table 4.8: Task list for task-knowledgeable users

Task no	**Task description**	**Difficulty index**
1	Open a document	5
2	Bold a single word	4
3	Centre align a word	4
4	Underline a single word	4
5	Create a bulleted list	5
6	Close a document (without saving)	3
7	Create a new document	3
8	Underline a single word	4
9	Italicise a phrase	4
10	Right align a paragraph	3
11	Save and close a document	8
TEST TOTAL		**47**

Usability measurements were then captured and analysed in order to determine the effect of the different interface components on the effectiveness and efficiency of the users. These usability measures will be discussed in the following section.

4.9 Usability measures

Only those tests in which all tasks were completed, whether successfully or not, were included in the analysis. A number of usability measures were captured and used to evaluate the usability of the application when used with different interfaces. The usability measures evaluated in this study were:

1. the score obtained by the user for a completed test;
2. the time taken to complete each task;
3. the number of actions required to complete a task;
4. the number of errors incurred by the user during completion of the task;
5. evaluation of the ratio between correctly and incorrectly answered tasks; and
6. measurement of users' subjective satisfaction.

4.9.1 Score

Each user was assigned a score for their completed test. The score was calculated as the sum of the difficulty indices of all the tasks that were completed successfully. Both tests had a maximum attainable score of 47.

4.9.2 Time

The time taken to complete the task was measured as the space of time between when the task was first loaded into the application and when the user indicated that the task had been completed. The measurement of time taken to complete a task has been used successfully in previous studies (for example, Benbasat and Todd, 1993; Dillon and Song, 1997; Kacmar and Carey, 1991; Kang, 2003; Read, MacFarlane and Casey, 2001; Roberts and Moran, 1983) as has the time-capturing method, of measuring the time from the start to the completion of the task (Read et al., 2001).

The users had the power to decide when they thought that the task had been successfully completed since this proved to be the most prudent method of moving on to the next task. If users had been prevented from moving to the next task until the current task had been successfully completed then many users would have been stuck on the same task for an unacceptably long period of time, thereby possibly preventing the completion of the test. Restricting the users to a certain amount of time per task may have unduly influenced the time analysis. Time measurement regardless of whether the task was completed successfully or not is an established testing method (for example, Kacmar and Carey, 1991). Similar to other studies (Benbasat and Todd, 1993), the time used was the entire time the user spent on the task with no provision made for time spent on errors.

The time was measured in seconds and then converted to 1/time for further analysis to increase the probability that the time data would follow a normal distribution.

4.9.3 Number of actions

An action by the user was considered to be one of the following:

a. Selection of a toolbar component be it an icon, button or drop-down;
b. selection of a menu option;
c. a mouse click (excluding those for toolbar and menu selections); and
d. a key press.

Actions were calculated per user and per task. Total actions for a task constituted the sum of (a), (b), (c) and (d) listed above. Furthermore, the actions also determined whether or not the user was considered to have at least attempted the task. At least one action had to be performed for the user to have been considered as having attempted to complete the task.

4.9.4 Number of errors

Evaluation of the errors in some form or another has been employed in a number of studies (Benbasat and Todd, 1993; Ellis et al., 1995; Kang, 2003; Roberts and Moran, 1983).

An error by the user for a task constituted one of the following:

a. Selection of a toolbar component – icon, button or drop-down – that was not required in order to complete the task successfully. For example, selection of the **Bold** icon when attempting to open a document was considered to be an error;
b. selection of a menu option other than that specific menu item required in order to complete the task successfully;
c. a key press within the editable region of the word processor.

Determination of errors (a) and (b) follows the example of previous studies (Ellis et al., 1995).

Error (c) is a special case because no task required typing of any kind but key presses could be due to selection of a word or phrase. They were, however, still deemed to be an error for the following reasons:

1. Novice users struggled to select a word or phrase and as such the final test was amended so that any task which required selection of a word or phrase was loaded into the application with the required text already selected. Users were then instructed to check whether the correct text was selected. Thus, any key press that was registered during completion of the task would have caused the cancellation of the selection, after which the user would have had to reselect the needed text – thus an erroneous action was committed.
2. The second user group, the task-knowledgeable users, were required to select the needed text themselves and as such some tasks could very well have had legitimate key presses. Therefore, key presses only accounted for errors on the tasks where no selection was required and thus, no key presses should have been made.

Key presses only accounted for a single error, irrespective of the number of key presses made by the user. Toolbar component and menu selection errors were calculated as the sum of the *distinct* incorrect toolbar component and menu selections. That is, if the user selected the same incorrect toolbar component or menu item numerous times, it would only account for a single error.

4.9.5 Task results

The number of correctly and incorrectly answered tasks was determined for each task and the ratio of correct to incorrect was evaluated as a measurement of the success rate.

4.9.6 Subjective satisfaction

In order to gauge subjective user satisfaction with the application, users were required to complete the Questionnaire for User Interaction Satisfaction (QUIS) as originally developed by Shneiderman (1987, as cited in Chin, Diehl and Norman, 1988) and later refined by Chin et al. (1988). Not only has the reliability of this questionnaire been established (Chin et al., 1988; Harper and Norman, 1993), it also been used satisfactorily in several studies in order to evaluate the usability of a wide range of software products (for example, Alonso, Rose, Plaisant and Norman, 1998; Granić, Glavinić and Stankov, 2004; Sittig, Kuperman and Fiskio, 1999; Sylaiou, Almosawi, Mania and White, 2004; Wallace, Norman and Plaisant 1998). No difference in satisfaction ratings were detected between paper-based and on-line administration of the QUIS, indicating that the format of the questionnaire does not affect the reliability of questionnaire response; thus the QUIS can be administered in either format without compromising the satisfaction ratings (Slaughter, Harper and Norman, 1994). However, the QUIS has proven to be more effective when administered in on-line format, as was administered in this study, as more valuable feedback was forthcoming from users during on-line administration (Slaughter et al., 1994).

The questionnaire divides the usability of a product into a number of distinct usability subsections such as system experience, overall user reactions, terminology and system information, learning to operate the system and the system capabilities (Shneiderman, 1998). Each question is rated using a 9-point LPC scale (Olivier, 2004), where the scale is anchored on either end with opposite adjectives. Negative responses correspond to the lower ratings and each scale point accounts for an increasing positive reaction (Shneiderman, 1998).

A subsection of the questionnaire as found in Shneiderman (1998) and containing only those questions pertinent to the application used in this study was used. The sections included in the questionnaire covered the overall reaction to the system and satisfaction with the learning and system capabilities.

The questionnaire was the final step in the application and was administered to users only after they had completed their test. It too was in electronic format as an integral part of the application and used radio groups to allow users to indicate their satisfaction rankings (Figure 4.7). The full questionnaire as presented to the users is listed in Appendix C.

Figure 4.7: QUIS in electronic format

4.10 Summary

As user tests were previously identified as the best method for usability testing, this study aimed to achieve the objectives of the research plan by implementing extensive user testing. As it is imperative that the user testing be conducted in context, a test instrument was developed which allowed for interchangeable interfaces which could be assessed for their usability worth. The interfaces which were tested consisted of different interface configurations made up of different icons sets and translated into different languages. The standard set of Microsoft Office icons provided the first set of icons. The second set of icons was obtained by means of using the results of previous icon preference studies and completing the set by designing icons during a brainstorming session. These icons were tested during a pilot study conducted with a group of potential word processor users. The pilot study revealed certain problems with the first design of the alternative icons and a second set of alternative icons were developed. The third set of icons used was actually a set of text buttons which contained no images but simply used the function name as the icon depiction. Translation of the interface was then achieved through translation of the tooltips, menus and text buttons. A complete set of testable interfaces was obtained by combining the menus, tooltips and icon sets into a number of different configurations.

User testing is not complete without setting up benchmark tasks that can be completed by target users. Two tests were designed which consisted of a number of simple tasks that had to be completed by the users. The first test was for use with the novice users and as such was formulated using simple language and clear instructions on how each task was to be completed. Tasks which were considered to be more difficult were placed at the end of the test to allow for as many tasks as possible to be completed. The second test was designed for use with the task-knowledgeable users. Both tests ensured that the tasks followed a natural sequence of actions which could be carried out in a word processor. Each task was assigned a difficulty index based on the minimum number of actions and inferences required by the users to complete the task. Both tests had a maximum attainable score of 47.

Automatic logging was chosen as the method of data capturing and functionality was built into the test instrument. From the automatic logging of user interaction a number of usability measurements could be derived for evaluation of the usability of the interface. The usability measurements were obtained from studying available usability models and taking usability recommendations into account. The usability measurements were (i) the score obtained for the test, (ii) the time taken to complete each task, (iii) the number of actions required to complete each task, (iv) the number of errors made by the users during the task, (v) the success rate of the tasks and (vi) the subjective satisfaction of the user after interaction with the application. The procedure followed to extract each of these usability measures from the captured user interaction was discussed.

Four types of users were identified that were used to test the usability of the developed interfaces. These user groups were novice, first-time, intermediate and expert users. The users were divided amongst the interfaces according to their first language. In this way, users could either complete the test on an interface in their home language or on an English interface, where English was considered to be their second language.

The following chapter will discuss a pilot study which was conducted to test the developed test instrument and confirm the layout of the test questions. Results obtained from formative evaluation and observation techniques will be discussed.

Chapter 5

Pilot study

5.1 Introduction

User testing was identified as the best method of testing the usability of alternative word processor interfaces. Since it is vital to allow users to complete benchmark tests within the context in which they would normally do so, a word processor application was developed which would facilitate user testing. The development of this word processor and the interface configurations tested in this study were discussed in the previous chapter. The first step in user testing should always be a pilot study to confirm the correctness of the experimental design before involving the test users. This chapter contains the results of the pilot study that was conducted to test the word processor application developed for administration of the user testing. A number of valuable findings and observations were made which will be discussed in detail. Participants of the pilot study were also asked to complete an icon preference test, the results of which will also be discussed in this chapter.

5.2 Participants

The participants for the pilot study consisted of high school learners from two local high schools. Learners ranged from Grades 8-11, with most having little to no prior experience with a word processor. These learners were, however, accustomed to having to complete projects and having to write essays and letters as part of their high school education. Because of this, these users were assumed to have a working knowledge of the task domain of a word processing application. Therefore, according to Shneiderman (1998), these participants could be classified as first-time, intermediate or expert users. The expertise rating scale (as defined in Section 4.5.2) was used to further classify the users into one of the three user groups. Many of the users had been using a word processing application for typing classes. However, these typing classes did not cover the actual workings of a word processor application and as such the users were only familiar with the look and feel of the programme but did not really have an active working knowledge of a word processor.

5.3 Testing environment

The testing of the learners took place in their school computer laboratory. The test was administered separately to each individual class, so each test session consisted of, on average, 40 participants. This allowed for close observation of the interaction between the users and the application. These observations uncovered a number of problems with the application, the particular interface designed for the study and the interface of a word processor in general. Thus, the testing of the first iteration of the interface proved to be very valuable in terms of simple observation of the usage of the product as well as in terms of verbal feedback from the participants. Furthermore, not many of the learners managed to complete the entire test and of those tests that were completed many tasks were simply skipped over with no attempt being made by the users to complete them. Thus, the data was not statistically analysed and instead the test sessions were regarded as a learning process which led to changes being made to the alternative icons and the word processor prototype.

5.4 Application shortcomings

After discovering that users were skipping questions, the prototype was changed to force the users to at least attempt the task. Thus, the test participant could not move on to the next task before at least one action had been carried out in an attempt to complete the task. It was decided not to keep the current task in place until users had completed the task successfully as this would cause an unnecessary delay in the completion of the task, since some users would struggle on a single task for a prolonged amount of time. So users merely had to perform a single action before the prototype would allow them to move to the next task.

5.5 Interface shortcomings

The changes made to the designed icons have already been discussed in Section 4.4.1, but to recap, the first choice for the alternative font styling icons was unsuccessful and the second choice for the icons was included in the interface instead. The original idea to display a transition between the states on a single icon image proved to cause difficulties for the users during attempted interpretation of the icon.

These findings confirm the need to test designed icons to ensure that the icons can be interpreted within the context of the application. Although the results of the usability tests were not analysed and as such it was impossible to determine whether the icons impeded the performance of the users, it was still clear by observation of the users and the feedback provided by the users that the icons were not understandable. This confirms the findings that some icons are not successful (for example, Zammit, 2000).

5.6 Qualitative observations

Different users respond differently to interfaces and the changes that occur due to the interaction with the system. The most valuable information gleaned from these sessions was not the problems identified with the icons or the prototype. Instead, other issues were revealed during observation and subsequent discussions with the participants. The problems were identified while monitoring the response of the users while they completed certain tasks. These problems are certainly not exclusive to this prototype and not necessarily exclusive to a word processor either. Problems with the following concepts or interface components were identified and will be discussed in detail in the subsequent sections:

- Applying changes to selected text
- Font
- Font colour
- Font size
- Complex dialog boxes
- Subtle changes
- No visible feedback

5.6.1 Applying changes to selected text

The prototype was designed to apply font changes only to selected text – the same way as in Microsoft Office for applications to multiple words. Many users did not realise that text had to be selected before changes could be applied to the desired piece of text. For example, many users successfully found the **Italic** icon when attempting to italicise a word or phrase but failed to select the text before clicking on the icon. This would of course have no effect on the text and users would then enquire what they were doing wrong. Once it had been explained to them that the word or phrase first had to be selected in order for the effect to be applied, few users had further difficulty in selecting the text. The stumbling block was the initial realisation that a selection had to take place.

5.6.2 Font

Participants did not fully understand the entire concept of *font*. There was little or no comprehension for the fact that *font* is a complete character set that has a number of attributes, such as a font type, size and colour and that these attributes can be changed by the user.

5.6.3 Font colour

Once users had successfully changed the colour of the text, the fact that the selection distorts the actual colour of the text caused much confusion and consternation amongst the users. For example, green-coloured text appears to be purple when selected. Users would then return to the dialog box and attempt to change the colour again, only to find the same result on return to the application. Participants had to be instructed to remove the selection in order to reveal the actual colour of the text.

5.6.4 Font size

Changing the font size caused a problem for the simple reason that when users saw the ruler stretching across the top of the window; they assumed that the numbers on the ruler (used for indentation of paragraphs) corresponded to the integer value of the font size.

5.6.5 Complex dialog boxes

The complexity of the font dialog box created two distinct problems for the users, namely:

- **Drop-down boxes**

  Users often identified the drop-down box, such as the font colour, that was required in order to complete the task. However, they did not understand that the drop-down expanded to reveal more options and could not understand why there was only a single colour visible in the option box.

- **Scroll boxes**

  Once users had been assisted to expand the drop-down lists, they still did not comprehend that the list could be scrolled through to reveal more options that were still hidden after the initial expansion. To these users, only the options which were immediately visible when expanding the drop-down were viable options. Users had to be shown that the list could in fact move both upwards and downwards to expose more available options.

5.6.6 Subtle changes

Subtle changes that occur within the application were not noticed by the users. For example:

- Creating a new document caused the background to change from grey to white and the editor to become enabled.
- The document name, as displayed in the title bar, changed from "Untitled" to the name provided by the user when first saving the document.

These types of changes were not noticed by the majority of these users, who then waited for confirmation that their action was in fact successfully carried out.

5.6.7 No visual feedback

Subtle changes went unnoticed by the users and also caused concern; a situation which was more pronounced when there was no visual feedback at all, such as when copying a piece of text or subsequently saving an already named file. These actions provide no feedback to the users and users continued to repeatedly click the icon in an attempt to elicit some response. This caused frustration on the part of the users, since they were convinced that they had identified the correct icon but nothing happened and they were then unsure of whether or not their instruction had been carried out.

These problems are not unique to the current study and by counteracting these perceived usability problems the usability of the end product could be substantially increased. Possible solutions to overcome these afore-mentioned problems provide an area of further research but time limitations prevented further in-depth explorations into these difficulties. Proposed solutions for these problems will be provided in Chapter 9.

5.7 Icon preference

In order to ascertain whether there is a preference for a certain icon amongst word processor users, a questionnaire pertaining to this was distributed to a single group of Grade 11 participants. The questionnaire consisted of images of the standard and alternative (after the redesign of the font styling images) icons as well as of the text buttons for all 16 of the functions found in the word processor prototype. Users were then asked to rank the icons for each function according to their order of preference.

Similar paper-based questionnaires have been used to great effect where respondents were required to match icons with the correct command (Potosnak, 1988) and measuring user preference by having users select the icon that they felt was best from an assortment of icons not only constitutes the second step of the ISO icon evaluation method but is also a tried and tested method (for example, Caron, Jamieson and Dewar, 1980; De Carolis et al., 1995; Green, 1979; Kim and Lee, 2005; Tudor, 1994). The questionnaire, in its final form, can be found in Appendix D.

There were a total of 36 questionnaires which were completed correctly and which did not contain a patterned response. Of these questionnaires, 30 were completed by Afrikaans, 5 by English and 1 by Sesotho first language speakers. In order to eliminate unnecessary factors and since the observations were limited, all but the Afrikaans speaking user questionnaires were discarded. Of the 30 remaining questionnaires, 20 belonged to first-time users, 4 to intermediate users and 6 to expert users. As with the rest of the analysis, the results obtained from the intermediate users were not considered for the analysis.

As the larger group and since they had had less exposure to a word processor and thus had less knowledge of the standard icons, the questionnaire results of the first-time users are graphically depicted below. The accumulated choices for the standard icons are situated at the bottom of each stack, followed by the accumulated choice for the alternative icons, and lastly the accumulated choice for the text buttons is found at the top of each stack.

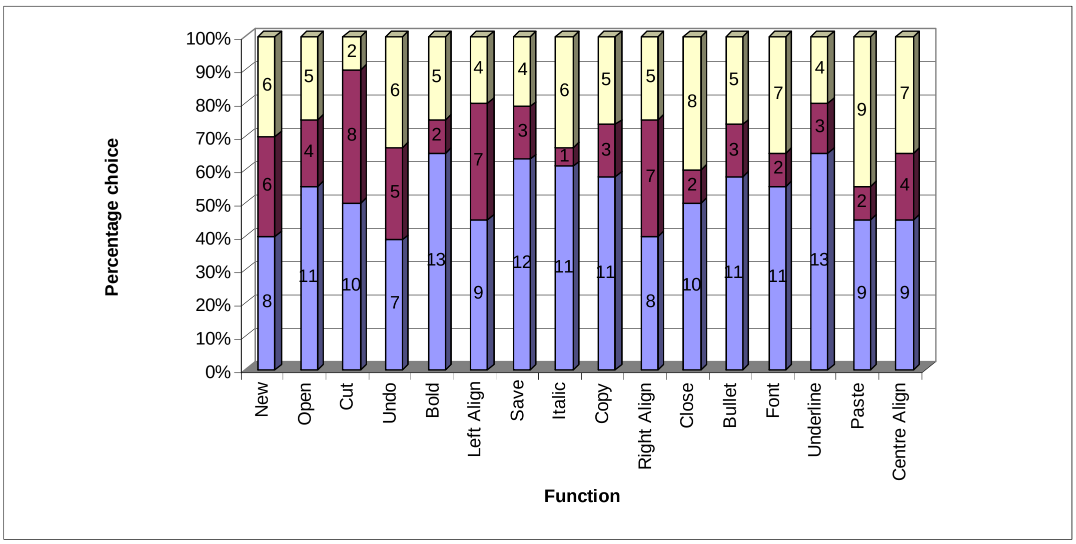


Chart 5.1: First-time user icon preference – First choice

All of the icons have a first choice majority for the standard icon, with more than half having a preference percentage of above 50%. The **Save**, **Bold**, **Italic** and **Underline** icons are the four icons that stand out as those with the highest percentage choice for the standard icon. The text buttons have the second most selections for the first choice of first-time users, followed by the lowest preference for the alternative icons. Table 5.1 tabulates the icon preference percentage for first-time users and experts.

First choice icons were ranked as follows for both first-time and expert users:

1. Standard icons
2. Text buttons
3. Alternative icons

Table 5.1: First-time and expert users icon preference percentages

	First choice	
	First-time users	**Expert users**
Standard	52	57
Alternative	20	15
Text	28	28

Therefore, it would seem as though the standard icons are the most preferred by word processor users. However, since it is well know that preference does not automatically reflect performance, it is still necessary to determine whether the standard icons promote better user performance.

5.8 Summary

A pilot study was conducted with a group of high school learners, ranging from Grade 8 to Grade 11, using the first iteration of the alternative icons. A number of interesting findings were made during observation of the users' interaction with the system. Some of these findings were unique to the study while others could be generalised to commonly used word processor applications as well as other applications.

Problems unique to this study constituted the discovery that the first iteration of the alternative icons was ineffective and should not be used in further user testing. These icons were subsequently redesigned to improve the alternative icons for the final testing of the interfaces.

Informal observations of user responses led to the discovery of certain factors which impeded the interaction of the user. These problems were not considered to be unique to the developed word processor but could also be extended to use of other word processors and even other applications. These included the confusion caused by the lack of visible feedback, subtle changes in the environment, lack of understanding of complete concepts such as font, difficulties experienced with selection of text and unexpected changes occurring in the environment, such as the distortion of text colour caused by a selection. Possible solutions to these problems will be discussed in a later chapter in the hope that a more usable environment can be created for all user groups.

A preference questionnaire designed to elicit the order of preference between the standard, alternative and text buttons was distributed amongst the users. The results indicated a high preference for the standard icons as a first choice, followed by the text buttons and finally the alternative icons. However, preference does not always reflect performance and whether the icons actually influenced user performance and learnability will be discussed in subsequent chapters.

The next chapter discusses the results of user testing conducted with novice users and how the interface language and icon sets affected their ability to learn and use the system.

Chapter 6

Testing novice users

6.1 Introduction

It was established in previous chapters that user tests were to be conducted with representative users. Furthermore, it was determined that in order to guarantee a comprehensive presence of representative users, members of all user groups had to be identified, sampled and tested on the different interface configurations. The first step in conducting successful user tests was to conduct a pilot study to ensure that the eventual usability tests would run smoothly and without incident. The previous chapter discussed the results of the pilot study and the valuable information that was obtained through the pilot study. Once the testing procedure had been finalised and confirmed by the pilot study, the four user groups had to be tested on the application. This chapter discusses the results of user testing conducted with the first of these groups, the novice users. Due to the limited number of available novice users only a subset of the developed interfaces could be tested. A comparison of user performance on pictorial icons and text buttons was also drawn to determine whether novices benefited from a certain icon type. Finally, a number of observations in terms of difficulties experienced during general use of certain interface components will be discussed.

6.2 Participants

In order to test the effects of the interface on the usability and learnability of a word processor, a group of potential users who had never used a computer before were approached and asked to participate in the study. These participants were sourced from the maintenance department of a South African university and consisted of employees who were responsible for the cleaning of buildings.

The task knowledge of these participants was considered to be minimal, if they possessed any at all, since they had a limited educational background and the nature of their work does not demand a working knowledge of any tasks normally associated with word processing. Since they had never before used a computer they were assumed to have no knowledge of the interface domain. Using the classification of users, as defined by Shneiderman (1998), these users were considered to be novice users.

Of the total of 45 users in this study, 40 were female. The vast majority of the users (31 users or 69%) spoke either Sesotho or Setswana as their first language. 20% of the users were native Afrikaans speakers, 1 participant was English speaking and the remainder (4 users) spoke an African language other than Sesotho and Setswana. Regarding education and socio-economic standing, these users all belonged to the same ranges.

6.3 Interfaces

The maximum number of allowable interfaces, as compelled by the limited number of novice users, was three different interfaces. This allowed for a reasonable number of users to be assigned to each interface group. The interfaces chosen for inclusion in this study were selected in order to give the best possible diversity of interface components while still ensuring that only a single interface component changed between two consecutive interfaces. The interfaces used in this study were:

1. Alternative icons with English tooltips and no menu;
2. Standard icons with English tooltips and no menu;
3. Standard icons with Sesotho tooltips and no menu.

6.4 Training

Since these participants had no previous experience with a computer, they were provided with limited training with the aim that they would eventually be able to complete simple tasks within the word processor application. Each user was assigned to an interface group at the start of the study and each group then received training only on their designated interface group. The different interface groups received training in separate sessions. As the only language to be fully understood by both the participants and the facilitators, Afrikaans was chosen as the instruction medium.

Users were however, taught the English and Sesotho terminology so that they could interact with their designated interfaces to the greatest potential that they offered.

The training course was divided into a series of one-hour training sessions which were spread out over a two-week period. Each session consisted of a short lecture, with plenty of time for the users to interact with the application. This was followed by a short test at the end of each session, which was designed to gauge their understanding and retention of the material covered in the previous day's session.

It was quickly discovered that the day-to-day knowledge retention of these users was extremely poor. Another important observation was the low self-confidence of these users concerning the use of the application. Many of the users refused to continue on their own, instead waiting for one of the facilitators to assist them or to confirm that their thinking was correct. These problems, together with the inherent problems experienced by a user with this profile, such as mouse handling, prohibited the collection of usable data from these training sessions. There were however many lessons learnt from these sessions, some of which provided valuable insight into the difficulties experienced by this user demographic and some of which could be implemented and tested. It was not possible to implement and test all of the observations and suggestions obtained from these groups and these may thus provide topics for further research.

6.5 Test administration

In order to benefit from these users, they were brought back for three one-hour follow-up sessions on consecutive days two months after the original course. The first day consisted of a recap of the work covered in the previous course. On the second day the users were required to complete test 1, as set out in Section 4.8, Table 4.7, on the same interface on which they first received their training. The following day the users then completed the same test but on the interface that made use of the English text buttons. Although testing users on consecutive days is not the ideal situation, should a difference in user performance be detected between user performance on the second and third days it would be expected that the third day's performance would be better than that of the second day. This increase in user performance could then be attributed either to the fact that the text buttons facilitated easier use or to the fact that the users had already been exposed to the application and had learnt to work on the application.

One lesson learnt from the first course and implemented in test 1, was the simplification of the language used in the formulation of the tasks. Tasks were compiled using only the material that was covered during the training sessions, with no extra or unknown actions being required of the users.

Once analysis of the results was started, only those tests that had been completed were included. The division of the resultant users according to their first language was as follows:

Table 6.1: Novice test interfaces

Interface	Afrikaans	Sesotho or Setswana	Other African language
Alternative icons, English tooltips		7	1
Standard icons, English tooltips	1	4	1
Standard icons, Sesotho tooltips		7	

As established in a previous section (Section 4.4.2), Sesotho and Setswana were treated as a single language group. Since the novice users for languages other than Sesotho and Setswana were minimal, and in an effort to simplify the data analysis, the results of the Afrikaans and other African language users were disregarded for the purposes of this study. Removal of these languages ensures a high degree of homogeneity for the remaining members of the group. This left only 18 tests of the original 45 users. The 18 users were all female and all belonged to the same socio-economic group. Four of the users had finished at least Grade 11, nine had a highest educational level of Grade 10, and the remainder (five users) had a highest education level of Grade 7.

6.6 Analysis

The usability measurements (Section 4.9) of (i) score, (ii) time taken to complete each task, (iii) number of actions needed to complete each task and (iv) number of errors made in each task were evaluated for both the icon and the language influence. The final usability measurement that was assessed was the ratio of correct and incorrect answers for each task. Due to the limited time available with these users subjective user satisfaction was not measured.

Since the interfaces had been selected in such a way that there was a difference of a single factor between the interfaces and only Sesotho first language speakers were included in the analysis, the interface language determined whether it was the L1 or L2 of the user. So, interface 1 had all L2 users, interface 2 had all L1 users and interface 3 had all L2 users. This resulted in an incomplete ANOVA design, so it was necessary to divide the analysis into two separate analyses each comparing two interfaces. Comparison of interface 1 with interface 2 would allow for isolation of the influence of the icon set, while comparison of interface 2 with interface 3 would isolate the influence of the interface language (L1 vs. L2).

Literature indicates that an interface in the L1 of the user should increase the efficiency of the user (Section 2.8), so if any difference was detected between interfaces 2 and 3 it was expected to be in favour of interface 3 as this was in the L1 of the users. Whether the alternative icons would affect the usability either positively or negatively could not be predicted. With no expectations either way and since the size of the groups were the same, the influence of the icon set was evaluated first.

6.6.1 Icon influence

It was decided to assess whether or not the icons did in fact impact on the performance of novice users and whether the icon design may eased the learning process for novice computer users if they were successfully designed to convey the intended meaning of the function linked to that particular icon. In order to do this, and only test the influence of the icons on user performance, a comparison analysis had to be done on the interfaces that differed only by means of the icon set. Since these interfaces were the same in all respects excepting for the icons used, any differences in performance between these two user groups could be attributed to the icons used in the interface. So the interface with a differing language, i.e. Sesotho, was removed from the analysis in order to study the impact of the icons on user performance.

The interfaces included in this analysis were as follows:

Table 6.2: Novice icon influence test user distribution

Interface	Number of completed tests
Alternative icons, English tooltips	7
Standard icons, English tooltips	4
Total	**11**

The dependent variable for each usability analysis was the relevant usability measure that was being investigated. The independent variable for all of the analyses was the interface used which in fact implies standard or alternative icons. This allowed for a standard unpaired t-test to be conducted. However, since a t-test is a parametric test and requires that observations follow the normal distribution, five normality tests were first conducted on the data. These tests were:

1. the Kolmogorov-Smirnov one sample test,
2. the Shapiro-Wilks test,
3. the Anderson-Darling test,
4. the Lilliefors test, and
5. the Jarque-Bera test.

If any one of these allowed for the null hypothesis of the sample distribution following a normal distribution to be rejected, the non-parametric equivalent of the t-test was used instead to analyse the observation. The most powerful non-parametric equivalent to the independent samples (unpaired) t-test is the Mann-Whitney U Test (Statistica Electronic Manual, 2005).

6.6.1.1 Score

The test score for each user was calculated as the sum of the difficulty indices of the tasks that were completed successfully. Users could score a minimum of 0 (zero), and a maximum score of 47 was achievable if all tasks were completed correctly.

The null hypothesis for analysis of the score was:

1. H_0: The icons of the interface do not affect the score of the user.

In order to assess whether the scores were normally distributed the following null hypothesis was tested:

1. H_0: The sample follows a normal distribution.

The p-values of the normality tests on the scores obtained are shown below:

Table 6.3: Novice users – Icon influence; score normality tests

Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera
> .20	0.269	0.402	0.586	0.517

* p < 0.05

A p-value of less than 0.05 meant that the null hypothesis could be rejected. As is evident from the above table, none of the five normality tests required rejection of the null hypothesis, thus it can be concluded that the scores for the sample follow a normal distribution and a t-test can be used to compare the performance of the two user groups.

The null hypothesis H_0 of the interface icons having no effect on the test score, could not be rejected, $t(9) = 0.849$, $p = 0.418 > 0.05$, which indicates that the interface, or in this case the icons, did not affect the overall score of the users.

6.6.1.2 Time

The analysis procedure for the time was as follows:

1. Each task was analysed separately (Bohmann, 2000; Ellis et al., 1995).

2. Since the observations were limited and very few users completed the tasks successfully, the task results of all the users who attempted the task were included in the analysis. An attempt to complete the task was determined based on the number of actions performed by the user during the task. The user must have carried out at least one action in order to be considered as having attempted the task.
3. The time taken to complete each task was measured in seconds for each user.
4. The measured time was converted to 1/time for further analysis (Section 4.9.2).
5. The normality of the observed time was evaluated using the same five normality tests as mentioned above in Section 6.6.1.1.
6. The statistical tests used for evaluation of each task were selected on the following basis:
    a. Should the observed time have followed a normal distribution as determined by the normality tests used in (5), the time of the two interface groups was compared by means of an unpaired t-test.
    b. The Mann-Whitney U test is the non-parametric alternative to the t-test and was used for analysis should any one of the five normality tests have disputed the normality of the data.

The following hypothesis could be formulated for this analysis:

1. H_0: The interface icons have no effect on the time taken to complete a task.

The results of the individual task analysis are shown in Table 6.4.

Tasks with less than 5 observations were not evaluated (Berenson and Levine, 1979). This resulted in the final three tasks being excluded from the analysis. These tasks were purposefully placed at the end of the test since it was observed during the course of the user training that users struggled with the complex dialog box and with understanding the concept of font changes. These proved to be the greatest stumbling blocks for the users and to facilitate the easier completion of the test it was decided to move these questions to the end of the test.

Table 6.4: Novice users – Icon influence; time analysis

Task no	n	*Normality tests*					*Hypothesis tests*	
		Kolmogorov-Smirnov	**Shapiro-Wilks**	**Anderson-Darling**	**Lilliefors**	**Jarque-Bera**	**t-test p-value**	**Mann-Whitney p-value**
1	10	> .20	0.001*	0.002*	0.007*	0.088		0.909
2	9	> .20	0.000*	0.000*	0.017*	0.046		0.796
3	7	> .20	0.000*	0.001*	0.010*	0.260		0.480
4	7	> .20	0.004*	0.008*	0.038*	0.365		0.289
5	7	> .20	0.384	0.458	0.589	0.579	0.877	
6	8	> .20	0.001*	0.001*	0.010*	0.160		0.386
7	5	> .20	0.078	0.085	0.161	0.640	0.453	
8	3	--	--	--	--	--	--	
9	4	--	--	--	--	--	--	
10	3	--	--	--	--	--	--	

* $p < 0.05$

-- Insufficient observations for meaningful analysis

The null hypothesis H_0 could not be rejected for any of the tasks, showing that the interface, or in this case the icons, did not affect the time needed by novice users to complete a task.

6.6.1.3 Number of actions

The same criteria as for the time analysis were used to select the user tasks for inclusion in the evaluation of whether the number of actions needed by the users to complete the task differed significantly between the icon sets. Similarly, the same criteria as for the time analysis were used for the determination of which statistical test was to be used in order to test the following null hypothesis:

1. H_0: The interface icons have no effect on the number of actions required by the users to complete the task.

The results of the analysis for each individual task are tabulated in Table 6.5.

H_0 could not be rejected for any of the tasks at an α level of 0.05, indicating that the interface did not impact on the number of actions needed by a novice user to complete a task.

Table 6.5: Novice users – Icon influence; number of actions analysis

Task no	n	*Normality tests* Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera	*Hypothesis tests* t-test p-value	Mann-Whitney p-value
1	10	> .20	0.000*	0.000*	0.008*	0.025*		0.305
2	9	> .20	0.000*	0.000*	0.017*	0.046*		0.197
3	7	> .20	0.391	0.411	0.514	0.617	0.307	
4	7	> .20	0.125	0.167	0.242	0.536	0.947	
5	7	<.15	0*	0*	0.001*	0.217		0.724
6	8	> .20	0.325	0.427	0.573	0.603	0.751	
7	5	> .20	0.795	0.732	0.882	0.708	0.105	
8	3	--	--	--	--	--	--	
9	4	--	--	--	--	--	--	
10	3	--	--	--	--	--	--	

* $p < 0.05$
-- Insufficient observations for meaningful analysis

6.6.1.4 Number of errors

Using the same selection criteria as for the time and number of actions, the following hypothesis was tested:

1. H_0: The interface icons have no effect on the number of errors made by the users during completion of a task.

The same statistical tests were used to test H_0 as for the hypotheses for the time taken to complete the task and the number of actions required during completion of the task. The results of the analysis are shown below.

Table 6.6: Novice users – Icon influence; number of errors analysis

Task no	n	*Normality tests* Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera	*Hypothesis tests* t-test p-value	Mann-Whitney p-value
1	10	<.05*	0*	0*	0*	0.376		0.21
2	9	<.10	0.001*	0*	0*	0.944		0.302
3	7	<.15	0*	0*	0*	0.494		0.86
4	7	<.05*	0*	0*	0*	0.197*		0.596
5	7	//	//	//	//	//	//	
6	8	<.10	0*	0*	0*	0.188*		0.885
7	5	> .20	0.146	0.135	0.146	0.683		0.083
8	3	--	--	--	--	--	--	
9	4	--	--	--	--	--	--	
10	3	--	--	--	--	--	--	

* $p < 0.05$
-- Insufficient observations for meaningful analysis
// Sample has no variance

None of the tasks had a normal distribution, thus the Mann-Whitney U test was used to analyse all the tasks. H_0 could not be rejected for any of the tasks, thus the icons used in the interface did not affect the number of errors made by the users during completion of the tasks. The errors for task 5 had no variance; therefore no statistical analysis could be done on the number of errors made during completion of that task.

6.6.1.5 Task results

The number of correctly answered tasks was too small ($n < 5$) to perform a meaningful analysis on this usability measure for any of the tasks.

6.6.1.6 Discussion

Since the icons used in the interface did not have an effect on the test score, the time taken to complete a task, the number of actions needed to complete a task or the number of errors made during completion of the task, it can be concluded that novice users do not need an alternative set of icons designed for their use. Alternative icon users achieve the same level of performance as standard icon users, indicating that the icon sets function equally well for training of novice users.

6.6.2 Consolidation of the interfaces

Since it was shown that the icons had no impact on user performance whatsoever, the interfaces no longer had to be grouped according to the icons used. All three of the interfaces were included in this analysis and division of the users was based on the language of the interface tooltips. Since only the Sesotho and Setswana speaking users were included in the analysis, the users could either have an interface in their L1 or in English, considered for these purposes to have been their L2. The final distribution for this analysis is shown below.

Table 6.7: Novice users – Consolidation of the interfaces

New interface group	Original interface configuration	Number of completed tests	Group Total
English (L2)	Standard icons, English tooltips	4	
	Alternative icons, English tooltips	7	11
Sesotho (L1)	Standard icons, Sesotho tooltips	7	7
Total		**18**	**18**

6.6.3 Language influence

The only remaining factor to be tested was the language used for the tooltips. The premise was that by making the interface available in the home language of the user, the time required by the novice user to master the use of the system would be lessened. Adaptation to a simple change in language should not be a problem, provided that the user has already learnt the functionality of the icons and the terminology associated with the application. The same usability measures of (i) score obtained, (ii) time taken to complete a task, (iii) number of actions required to complete a task and (iv) number of errors made during completion of the task were analysed in terms of the language of the interface.

As in the previous analyses the dependent variable for each usability analysis was the relevant usability measure that was being investigated. The independent variable for all of the analyses was the language of the interface which indicated whether it was the L1 or L2 of the user.

6.6.3.1 Score

The score for each user was again calculated as the sum of the difficulty indices for all the correctly completed tasks. The same analysis procedure was followed as for the icon analysis (Section 6.6.1.1), that is, first testing the normality of the observations and then using either an unpaired t-test or Mann-Whitney U test to evaluate the results.

The following null hypothesis was tested:

1. H_0: The interface language has no effect on the score of the user.

The results of the normality tests are contained in the table below:

Table 6.8: Novice users – Language influence; score normality tests

n	Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera
18	> .20	0.202	0.203	0.139	0.451

* $p < 0.05$

According to all five normality tests used, the distribution of the score was normal at an α level of 0.05, since $p > 0.05$. The t-test was not significant at an α level of 0.05, $t(16) = -1.011$, $p = 0.327 > 0.05$, so H_0 could not be rejected, thus the language of the interface had no effect on the score achieved by the user.

6.6.3.2 Time

The same analysis procedure as for the testing of the icon influence (Section 6.6.1.2, steps 1-6) was used to test whether the language had an effect on the time taken to complete a task. The following hypothesis was formulated for this analysis:

1. H_0: The interface language has no effect on the time taken to complete a task.

The results obtained from the individual analyses are tabulated in Table 6.9.

Tasks 8, 9 and 10 had sufficient observations (n >= 5) for analysis for this user group. However, as it could not be established whether or not the icons influenced the performance of the users for these tasks, there was not enough motivation to group the interfaces based only on the language for these tasks.

Table 6.9: Novice users – Language influence; time analysis

Task no	n	Normality tests					Hypothesis tests	
		Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera	t-test p-value	Mann-Whitney p-value
1	14	> .20	0.001*	0.002*	0.008*	0.014*		0.724
2	15	<.10	0.000*	0.000*	0.000*	0.001*		0.906
3	12	<.10	0.000*	0.000*	0.000*	0.057		0.123
4	14	> .20	0.006*	0.004*	0.008*	0.312		0.655
5	11	> .20	0.062	0.104	0.244	0.462	0.810	
6	14	> .20	0.050	0.131	0.572	0.218	0.217	
7	11	> .20	0.051	0.066	0.043*	0.459		0.273
8	9	--	--	--	--	--	--	
9	8	--	--	--	--	--	--	
10	7	--	--	--	--	--	--	

*p < 0.05

-- Insufficient observations for meaningful analysis

Using a significance level of 0.05, the null hypothesis, H_0, could not be rejected for any of the tasks, thereby proving that the time required by users to complete a task was not affected by whether the interface was in the L1 of the user or not.

6.6.3.3 Number of actions

The number of actions required by the users to complete the tasks was also evaluated using the same procedure as for the group when testing icon influence (Section 6.6.1.3). The results for the analysis of the following hypothesis are contained in Table 6.10.

1. H_0: The interface language has no effect on the number of actions required to complete a task.

The results contained in Table 6.10 confirm that the interface language did not impact the number of actions required to complete a task since the null hypothesis could not be rejected for any of the tasks.

Table 6.10: Novice users – Language influence; number of actions analysis

Task no	n	*Normality tests* Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera	*Hypothesis tests* t-test p-value	Mann-Whitney p-value
1	14	<.10	0.000*	0.000*	0.000*	0.000*		1
2	15	> .20	0.018*	0.018*	0.058	0.388		0.263
3	12	> .20	0.012*	0.030*	0.143	0.043*		0.372
4	14	> .20	0.126	0.152	0.306	0.428	0.772	
5	11	<.10	0.000*	0.000*	0.000*	0.000*		0.925
6	14	> .20	0.026*	0.026*	0.019*	0.388		0.302
7	11	> .20	0.013*	0.017*	0.098	0.337		0.465
8	9	--	--	--	--	--	--	
9	8	--	--	--	--	--	--	
10	7	--	--	--	--	--	--	

*p < 0.05
-- Insufficient observations for meaningful analysis

6.6.3.4 Number of errors

Using the same techniques as for the evaluation of icon influence (Section 6.6.1.4), the errors were also evaluated for these two groups of users. Table 6.11 contains the results of the analysis of the following null hypothesis:

1. H_0: The interface language has no effect on the number of errors incurred during completion of a task.

Table 6.11: Novice users – Language influence; number of errors analysis

Task no	n	*Normality tests* Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera	*Hypothesis tests* t-test p-value	Mann-Whitney p-value
1	14	<.05*	0.000*	0.000*	0.000*	0.282		0.437
2	15	<.05*	0.001*	0.000*	0.000*	0.920		0.157
3	12	> .20	0.001*	0.001*	0.013*	0.125		0.104
4	14	<.01*	0.000*	0.000*	0.000*	0.000*		0.655
5	11	--	--	--	--	--	--	
6	14	<.01*	0.000*	0.000*	0.000*	0.000*		0.439
7	11	> .20	0.001*	0.001*	0.007*	0.143		0.361
8	9	--	--	--	--	--	--	
9	8	--	--	--	--	--	--	
10	7	--	--	--	--	--	--	

* p < 0.05
-- Insufficient observations for meaningful analysis

H_0 could not be rejected for any of the tasks, so the number of errors made during completion of the task was not determined or influenced by the language of the interface. The error rate of task 5 could not be evaluated as it had no variance thus not allowing for any statistical analysis to be done.

6.6.3.5 Task results

The final analysis that was conducted was to test whether there was a significant difference between the number of tasks answered correctly and those that were completed incorrectly. More often than not, the number of correctly completed tasks was very low and as can be seen from Chart 6.1 only four tasks had a success rate above 50% when including only the attempted tasks. Therefore the results did not allow for a meaningful comparison of the number of correctly and incorrectly completed tasks and the following discussion will be based on interpretation of the success rate as depicted in Chart 6.1.

Task 2, which required users to make a single word bold, had the lowest success rate. Two other tasks, tasks 4 and 6, required users to italicise and underline a word. There was a marked increase in the success rates between tasks 2 and 4, after which the success rate remained somewhat constant for task 6. This would seem to indicate that the users struggled with the first font styling task, after which they used that knowledge to assist them in completing the subsequent font styling tasks.

Task 3 had the second lowest success rate. This task had by far the highest difficulty level and required the users to cut a word and then paste it a couple of lines higher. It is understandable that these users struggled to complete this question successfully.

Task 5 appears to be by far the task where the users performed the best. The 11 users of the 18 included in the analysis who attempted the task all managed to complete it successfully. The task required that the users close the current file. The users were also fairly evenly spread amongst the different interface groups with 3 users making use of interface 1, and 4 users each using interfaces 2 and 3.

The success rates of tasks 8, 9 and 10 are not a cause for concern, since the tasks were moved to the end of the test precisely because it was expected that users would not be able to complete them correctly, thus the success rate of these tasks was actually higher than expected.

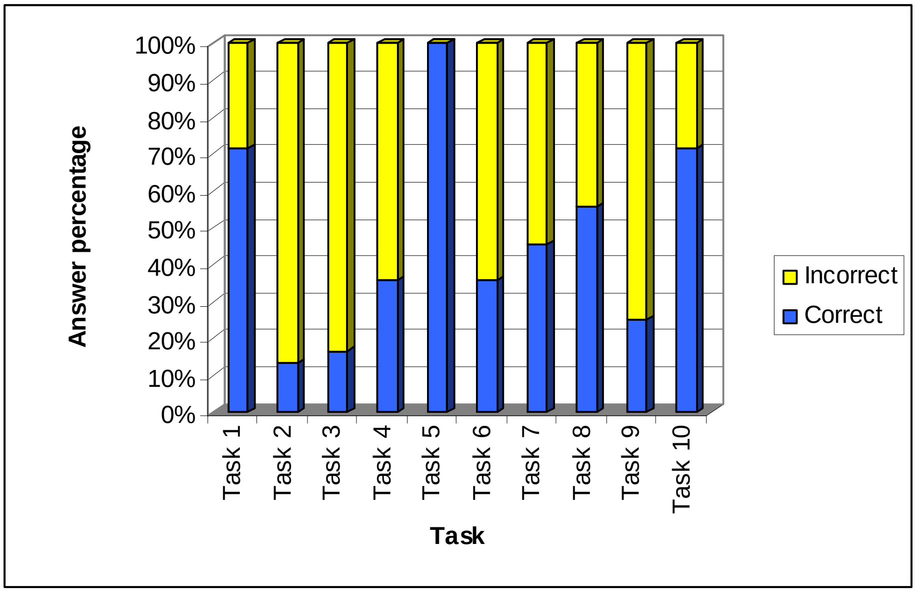


Chart 6.1: Novice users – Task result percentages

6.6.3.6 Discussion

Since none of the analysed usability measures indicated a difference in performance between L1 and L2 users, inclusion of tooltips in the first language of the user did not impact user performance either positively or negatively. The users learnt the correct terminology and interpretation of the icons during the training sessions, and it would seem as though inclusion of tooltips did not significantly enhance interaction for either language group. Therefore, novice users can be provided with tooltips in either their L1 or in English, provided that they are comfortable with the application terminology.

A possible further area of study would be to compare the performance of novice users using the same icon set, but with tooltips being absent from one interface and present in the other. This would allow for the investigation of whether tooltips, regardless of language, increase the usability of a product for novice users.

6.6.4 Comparison of pictorial and textual icons

Literature indicates that pictorial processing is quicker than textual processing (Section 2.10), but there is no empirical evidence to support this claim when it comes to icons in a graphical user interface environment. On the contrary, it is possible that the text buttons may be easier to use since they contain only a function word. In other words, if users wanted to underline, then they only have to locate the button with the word “underline” on it. There are no complex commands to learn and the text buttons provide a means of selecting the function name, which is known to the user, from a collection of buttons in place of having to select a graphical depiction of the function. Comparison of the usability measurements will allow a conclusion to be made on whether the pictorial icons are easier to use than the text buttons or vice versa.

On the third day of the follow-up course, users were required to complete the same test but this time using the English text buttons. The results of the two tests were compared in order to determine whether there was a difference between the performances of the users when using a pictorial or a textual interface. The only difference that was expected was an increased performance on the third day due either to the fact that the text buttons were easier to use or to the fact that the users had learnt from the previous days’ exposure to the application.

6.6.4.1 Score

Fourteen users managed to complete the whole test using both their predetermined pictorial interface and the text buttons. The score was calculated using the difficulty index assigned to each task. The scores obtained for these users are listed in Table 6.12.

With a higher mean score and 9 out the 14 users obtaining a higher score for the test when using the text buttons as opposed to the pictorial icons, it would be tempting to say that the text buttons were easier to use. However, it would be remiss to do so without testing whether the observed difference is in fact a significant one.

Table 6.12: Novice users – Score comparison

		Score	
	User	**Pictorial icons**	**Text buttons**
	3001	8	5
	3002	11	16
	3003	25	29
	3006	9	16
	3009	12	11
	3011	0	7
	3014	8	0
	3016	13	30
	3017	26	7
	3027	30	15
	3033	17	22
	3035	5	29
	3040	34	43
	3044	0	26
Mean score		14.143	18.286
Std. dev		10.791	12.060

The scores of all the users who completed both tests were compared by means of a one-tailed paired t-test since the normality tests ($\alpha = 0.05$) showed that the score ranges for both tests followed a normal distribution (Table 6.13).

If there is a significant difference between the scores obtained using the different interfaces, then the interface that allows for a higher score to be obtained would logically be the easier interface to use for this user group. Thus the following hypothesis was put to the test:

1. H_0: The text buttons are not easier to use than the pictorial icons.

Table 6.13: Novice users – Comparative score normality tests

	Pictorial icon test	Text button test
Kolmogorov-Smirnov	> .20	> .20
Shapiro-Wilks	0.280	0.733
Anderson-Darling	0.257	0.632
Lilliefors	0.217	0.571
Jarque-Bera	0.527	0.673

*p < 0.05

Since $t(13) = 1.190$, $p = 0.128 > 0.05$, H_0 could not be rejected, which proves that there was no significant difference between the scores of the users who completed both tests. This result indicates that neither interface was significantly easier to use, therefore neither interface facilitates easier completion of the tasks at hand, or leads to significantly higher attained scores.

6.6.4.2 Time

In order to determine whether or not tasks were completed quicker when using the interface with the text buttons, the times of all tasks that were completed successfully when using the text buttons were compared to the times of the tasks when using the pictorial icon interface, regardless of whether the task was completed successfully or not but provided that the user had at least attempted to complete the task. The following procedure was followed for comparison of the task times between the two interfaces for each individual task:

1. Following an established evaluation method (Ellis et al., 1995) each task was analysed separately.
2. The times of the tasks that were completed successfully on the text button interface were calculated in seconds.
3. The calculated time was converted into 1/time for further analysis.

4. For each of the users identified in (2), the time taken for the same task using that user's respective pictorial interface was calculated in seconds.
5. This time was also converted into 1/time.
6. Users who did not attempt to complete the task using the pictorial icons were removed from the analysis.
7. The normality of the observed time for both interfaces was evaluated using the five normality tests used in the previous analyses, namely the Kolmogorov-Smirnov one sample test, the Shapiro-Wilks, Anderson-Darling, Lilliefors and Jarque-Bera normality tests.
8. Statistical tests were carried out to compare the times using the two interfaces according to the following method:
    a. Should the observed times have followed a normal distribution as determined by the normality tests used in (7), the two measured paired times of the users were compared by means of a paired t-test.
    b. The Wilcoxon one-tailed paired test was used as the non-parametric alternative to the one-tailed paired t-test, should one or both of the time distributions not be normal.

The results of the normality tests for each test are shown in Table 6.14.

Table 6.14: Novice users – Comparative time normality tests

Task no	n	Test	Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera
1	6	1	> .20	0.254	0.177	0.094	0.633
		2	> .20	0.299	0.400	0.695	0.596
2	5	1	> .20	0.010*	0.013*	0.011*	0.590
		2	> .20	0.644	0.598	0.645	0.648
3	2		--	--	--	--	--
4	8	1	> .20	0.007*	0.008*	0.005*	0.485
		2	> .20	0.119	0.104	0.071	0.559
5	7	1	> .20	0.185	0.267	0.419	0.570
		2	> .20	0.176	0.137	0.184	0.595
6	5	1	> .20	0.141	0.127	0.132	0.653
		2	> .20	0.073	0.078	0.090	0.589
7	4		--	--	--	--	--
8	1		--	--	--	--	--
9	2		--	--	--	--	--
10	2		--	--	--	--	--

*$p < 0.05$
-- Insufficient observations for meaningful analysis

Statistical analysis results are shown in Table 6.15 for the following null hypothesis:

1. H_0: Text buttons do not decrease the amount of time needed to complete a task.

Table 6.15: Novice users – Comparative time analysis

Task no	n	t-test	Wilcoxon paired test
1	6	0.037*	
2	5		0.094
3	2	--	
4	8		0.371
5	7	0.185	
6	5	0.203	
7	4	--	
8	1	--	
9	2	--	
10	2	--	

*$p < 0.05$
-- Insufficient observations for meaningful analysis

H_0 could only be rejected for the first task ($p = 0.037$) which required users to open a document. The test 2 task was completed in a significantly shorter time than the test 1 task. As this was the first task, and no other task showed a significant difference between the completion rates, a possible explanation for this observance could be that users were hesitant at the start of the first test and took a while to familiarise themselves with the workings of the application again, thus contributing to a longer completion time for that task.

6.6.4.3 Number of actions

The tasks to be included in the analysis were selected using the same process as for the time selection of the comparison analysis (Section 6.6.4.2).

Results of the normality tests for the actions of the two tests are shown in Table 6.16.

Table 6.16: Novice users – Comparative number of actions normality tests

Task no	n	Test	Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera
1	6	1	> .20	0.085	0.098	0.089	0.617
		2	> .20	0.000*	0.001*	0.014*	0.437
2	5	1	> .20	0.106	0.093	0.057	0.603
		2	> .20	0.374	0.336	0.269	0.626
3	2		--	--	--	--	--
4	8	1	> .20	0.303	0.483	0.829	0.588
		2	> .20	0.005*	0.007*	0.018*	0.243
5	7	1	<.20	0.000*	0.000*	0.002*	0.274
		2	<.10	0.000*	0.000*	0.000*	0.223
6	5	1	> .20	0.037*	0.045*	0.057	0.645
		2	> .20	0.257	0.270	0.375	0.614
7	4		--	--	--	--	--
8	1		--	--	--	--	--
9	2		--	--	--	--	--
10	2		--	--	--	--	--

*p < 0.05
-- Insufficient observations for meaningful analysis

The following hypothesis was tested and the results of the analysis are shown in Table 6.17.

1. H_0: Text buttons do not decrease the number of actions needed to complete a task.

Table 6.17: Novice users – Comparative number of actions analysis

Task no	n	t-test	Wilcoxon paired test
1	6		0.855
2	5	0.418	
3	2	--	
4	8		0.611
5	7		0.364
6	5		0.392
7	4	--	
8	1	--	
9	2	--	
10	2	--	

*p < 0.05
-- Insufficient observations for meaningful analysis

The null hypothesis could not be rejected for any of the tasks, thereby showing that the text buttons did not facilitate fewer actions for the users to complete the tasks.

6.6.4.4 Number of errors

The same selection process as for the time (Section 6.6.4.2.) and actions (Section 6.6.4.3) paired comparison was followed.

Results of the normality tests are shown in Table 6.18.

Table 6.18: Novice users – Comparative number of errors normality tests

Task no	n	Test	Kolmogorov-Smirnov	Shapiro-Wilks	Anderson-Darling	Lilliefors	Jarque-Bera
1	6	1	> .20	0.001*	0.002*	0.002*	0.536
		2	<.10	0.000*	0.000*	0.000*	0.396
2	5	1	> .20	0.325	0.129	0.146	0.815
		2	<.20	0.000*	0.000*	0.001*	0.001*
3	2		--	--	--	--	--
4	8	1	//	//	//	//	//
		2	<.20	0.004*	0.004*	0.004*	0.002*
5	7	1	//	//	//	//	//
		2	<.20	0.000*	0.000*	0.000*	0.000*
6	5	1	//	//	//	//	//
		2	//	//	//	//	//
7	4		--	--	--	--	--
8	1		--	--	--	--	--
9	2		--	--	--	--	--
10	2		--	--	--	--	--

*p < 0.05

-- Insufficient observations for meaningful analysis

// Sample has no variance

The following hypothesis was explored:

1. H_0: Text buttons do not decrease the number of errors made during completion of a task.

The results of the analysis of H_0 are tabulated in Table 6.19.

The null hypothesis could not be rejected for any of the tasks, leading to the conclusion that the text buttons did not assist the users any more than the pictorial icons, and that the use of text buttons did not lower the error rate of the users.

Table 6.19: Novice users – Comparative number of errors analysis

Task no	n	t-test	Wilcoxon paired test
1	6		0.678
2	5		0.968
3	2	--	
4	8	0.052	
5	7	0.086	
6	5		
7	4	--	
8	1	--	
9	2	--	
10	2	--	

*$p < 0.05$

-- Insufficient observations for meaningful analysis

6.6.4.5 Discussion

No difference was detected between the performances of the users in terms of score, time to complete each task, the number of actions required or the number of errors made during completion of the task. The only exception was the significant time difference detected for the first task, where the task was completed faster using the text buttons. Since this can be explained by the fact that users had to familiarise themselves with the way the tests worked during the first test, it can be concluded that the text buttons did not facilitate increased user performance.

However, the fact that the second interface was text-based and users simply had to identify the icon with the correct function word could have played a role in the easy adaptability of these users between the two interfaces. An area which should be explored to test comprehensively the ease with which these users adapt to changes in the interfaces would be to test the users, or another group of users with the same profile, on a different pictorial interface. Should the users then display an unaffected performance, it would indicate that once users have become accustomed to the terminology and usage of the application, they can easily adapt to changes in the interface.

6.7 Qualitative observations

Problems experienced by these users during interaction with the application highlighted some issues which could provide topics for further research in this area.

It was observed that users struggled to interact with the complex font dialog boxes. Some of the same observations that were found for the school-going users tested previously were evident with these users. These were:

- the confusion experienced by the concept of a drop-down list; and
- the concept of scrolling upwards and downwards to reveal more options.

Furthermore, it was found that the numerous options present in the conventional dialog boxes, such as the various colours available for selection, hampered the users and added to the confusion of the users which made them unsure of what to do. The drop-downs presented great difficulty for a number of reasons, namely:

1. Similar to the school-going users, these users could also find the correct area in the dialog box but had difficulty in understanding that clicking on the down arrow would cause the drop-down to expand and reveal more options.
2. Clicking on the small arrow created difficulty due to poor mouse dexterity.
3. Once the drop-down had been successfully expanded, users did not understand how to scroll upwards and downwards in order to reveal more available options.
4. Some users were also overwhelmed by the sheer complexity of the dialog boxes and the number of options and information available in the small area.

Solutions to these problems will be proposed in Chapter 9.

6.8 Summary

A group of novice word processor users who also happened to be novice computer users were identified and received limited training on the word processor application.

Due to the limited number of available novices only the following interfaces were used for this part of the study:

1. Alternative icons with English tooltips and no menu
2. Standard icons with English tooltips and no menu
3. Standard icons with Sesotho tooltips and no menu

Each user received training on only a single interface included in the study. Users were then required to complete Test 1 as outlined in Section 4.8, from which usability measures were obtained that could be compared to determine which interface was the most usable.

Interfaces 1 and 2 as listed above were compared first since they only differed by a single factor so any difference that was detected could then be attributed to the influence of the icons and there were no expectations as to which interface would facilitate better performance by the users. No significant difference was detected between any of the usability measurements for any of the tasks. Therefore, neither icon set significantly increased the learnability, effectiveness or efficiency of the word processor application. This would seem to indicate that novice users adapt to the interface used and learn the meanings of icons, since they are equally comfortable using the standard icons and the alternative icons.

Since the icons had no influence on the usability of the interface there was no longer any need to differentiate between the interfaces based on the icons used. Therefore the interaction of all three interfaces were evaluated together without differentiating between the icon sets and only dividing the interfaces based on the language of the interface which was either in the L1 or L2 of the user. Storage and processing of L1 and L2 indicate that an interface in the L1 of the user should increase the usability of the interface. However, the language of the interface did not affect the score of the test, the time taken to complete the tasks, the number of actions carried out or the number of errors made during completion of any of the tasks. Therefore, it could be concluded that the interface language did not influence the usability of a word processor.

Finally, the users were required to complete the same test but this time using the interface with the English text buttons. The performance of the users on their original pictorial interface was then compared to their performance on the text buttons. Should any difference be detected it was expected to be that the performance on the text buttons would be better either due to the fact that users had learnt from their previous experience or because the text buttons were easier to use. However, results indicated that these users performed on a comparable level between the two icon types, therefore leading to the conclusion that the text buttons were not easier to use. The only difference that was detected was on the first task where the pictorial interfaces required more time to complete the tasks. This difference could, however, be attributed to the fact that users first had to become accustomed to using the application again. These results show that these users were able to adapt easily to a slightly changed interface once they had become comfortable using the original interface. All that is required by novice users is training and exposure to the product in order to familiarise themselves with the usage and workings of the application.

The following chapter deals with the results of user tests conducted with the more task-knowledgeable users where all available interface configurations could be included in the testing.

Chapter 7

Testing task-knowledgeable users

7.1 Introduction

User testing must be conducted using as many potential users as possible and ensuring that the users represent users at all levels of expertise. In this way it can be ensured that all usability problems are revealed by testing users with less expertise while simultaneously guaranteeing that the problems revealed do exert influence on the usability of the application by testing users with more expertise in the task and interface domain. The previous chapter reported the results of user tests conducted with novice users. Once it had been determined that neither the icon set used nor the interface language had influenced the usability of the product for this user group, it was imperative that it be established whether the same conditions held for users with more knowledge of the task and interface domain. This would ensure that the interface configurations could be tested on a truly representative group of users. By testing such task-knowledgeable users, it could be determined whether the alternative icons, text buttons and interface language enhanced the user interaction with the application. It could also be established whether these users could adapt to changes in the interface as easily as the novice users had. This chapter will discuss the results found during these user tests with the more task-knowledgeable users.

7.2 Participants

The subjects used as experienced users consisted of university students who were all taking a basic computer literacy course. The following is a description of only those users whose tests were acceptable for inclusion in the analysis. The participants, 403 of whom were female and 283 male, provided for different levels of word processor expertise. The average age of the participants was 18, with the youngest participant being 16 years old (n = 1) and the oldest 47 years of age (n = 1). All the participants had passed at least Grade 12, with 6% (n = 40) having a tertiary qualification in the form of a diploma or a degree. Test subjects spoke a variety of languages, including

English, Afrikaans, Sesotho, Setswana, isiXhosa and isiZulu. With a percentage of 42%, the greatest majority of the students were Afrikaans first language speakers (n = 290), followed by Sesotho and Setswana first language speakers (35% and n = 239). Regardless of what their first language was, all subjects were conversant in either English or Afrikaans as these are the tuition languages of the university.

Since the participants had all successfully completed Grade 12, it was assumed that they had experience in compiling letters, writing essays and creating a well-ordered document for completion of school projects. Therefore, it could be concluded that they at least had knowledge of the task domain of a word processor. According to Shneiderman (1998), this task domain knowledge allowed for these users to be classified as first-time, intermediate or expert users. To complete the classification of each user into one of the user groups, the interface knowledge of the users was gauged using the expertise scale as constructed and defined in Section 4.5.2.

7.3 Testing environment

The test was conducted during the first practical session for the course in the computer laboratory on the university campus grounds. Each practical session had between 90 and 120 attending students. The practical took place before the users had received any instruction in word processor packages. Up until that point they had only been taught basic Windows concepts.

All the interface configurations were tested using the experienced users. The interfaces were:

1. Standard icons with a menu and tooltips;
2. Standard icons with tooltips but no menu;
3. Standard icons with neither a menu nor tooltips;
4. Alternative icons with a menu and tooltips;
5. Alternative icons with tooltips but no menu;
6. Alternative icons with neither a menu nor tooltips;
7. Text icons with tooltips but no menu.

The application was loaded onto all the computers in the laboratory prior to the practical sessions. Each application defaulted to one of the available interfaces in such a manner that no two adjacent computers had identical interfaces. Users were assigned to the interface groups at random while ensuring that the number of L1 and L2 users for each interface was balanced. After conclusion of all the practical sessions there were 686 tests that had been completed and that met the requirements for the language division. After elimination of the intermediate users the total number of usable tests was 517.

Users were required to complete test 2 as set out in Section 4.8, which consisted of eleven simple word processing tasks that could be completed solely by making use of either a menu option (when a menu was present) or a shortcut icon on the toolbar.

7.4 Analysis

A distinction was drawn between interfaces containing the extra factor of interface language and those with no language component. Thus, the two interfaces without a language component (3 and 6) were removed from the initial analysis to be evaluated separately from the remainder of the interfaces. The subjects who participated in the study and completed the test on the interfaces that contained a language component were designated group A, and the rest of the subjects were categorised as group B. The final distribution of users is shown in Table 7.1.

The same usability measurements were analysed as for the novice users, namely (i) score, (ii) time taken to complete each task, (iii) number of actions required to complete each task, (iv) number of errors made during completion of each task, (v) ratio of correct and incorrect answers for each task and (vi) subjective satisfaction.

Table 7.1: First-time and expert user distribution

User Group		Interface	Language	First-time	Expert	Group Total
Group A	1	Standard icons, Menu, Tooltips	L1	24	20	
			L2	26	23	
	2	Standard icons, No menu, Tooltips	L1	22	26	
			L2	24	13	
	4	Alternative icons, Menu, Tooltips	L1	13	17	
			L2	26	23	
	5	Alternative icons, No menu, Tooltips	L1	21	23	
			L2	25	26	
	7	Text buttons, No menu, Tooltips	L1	15	20	
			L2	33	25	445
Group B	3	Standard icons, No menu, No tooltips		19	15	
	6	Alternative icons, No menu, No tooltips		17	21	72
Total				**265**	**252**	

For each of the two sets of data, group A and group B, the following considerations were taken into account:

- Only completed tests were used during the analysis.
- The time taken, number of actions needed and the number of errors made were measured and analysed for each individual task. The same method of evaluating each task separately has been used in previous studies (Ellis et al., 1995).
- In contrast to the novice group, this group provided for more than enough observations even when the incorrect tasks were disregarded from the analysis. To prevent the observation data from being disproportionately skewed by the incorrect tasks and because this method had been used successfully in previous

studies (Nielsen, 2006), only those tasks that were completed successfully were included in each analysis.

- When evaluating the results of the tasks, that is, comparing the number of successes with the number of failures, only those measurements where the user actually attempted to complete the task were taken into account. These tasks were identified as those where the user performed at least one action during the course of the task.
- Where data was obtained from questionnaires, such as with the measurement of subjective satisfaction, anxiety and attitude, only fully completed questionnaires were included in the analysis and questionnaires exhibiting a patterned response were summarily excluded from the analysis. A questionnaire with a patterned response was identified as one where all responses had the same scale value.

The results obtained from group A were analysed first. The decision to evaluate the language group first was based purely on the fact that it was the larger group and thus was considered to have higher reliability. A secondary motivation for analysing group A first was the possibility of consolidating the groups as with the novice users, if a single factor exerted no influence on the usability measures.

7.4.1 Analysis of group A

Group A consisted of those users who used any one of the interfaces 1, 2, 4, 5 or 7 as listed above. Furthermore, these interfaces could either be in the L1 or the L2 of the user. All of the prior mentioned usability measures were analysed for participants of group A.

7.4.1.1 Independent variables

The independent variables used in the analysis were the interface employed during the test, the word processor expertise of the user and whether or not the user completed the test in their L1 or L2 (English for all users).

7.4.1.2 Dependent variables

For the analysis of each usability measurement, that particular measurement was the dependent variable.

7.4.1.3 Normality of the data

Since there were multiple independent factors, each with two or more levels, the correct statistical test to use for data of this kind was the factorial Analysis of Variance (ANOVA) (Statistica Electronic Manual, 2005). However, the ANOVA assumes that the observations are normally distributed (Berenson and Levine, 1979; Statistica Electronic Manual, 2005). In order to test the normality of the observations, the following five normality tests were conducted on the sample data:

- the Kolmogorov-Smirnov one sample test,
- the Shapiro-Wilks test,
- the Anderson-Darling test,
- the Lilliefors test, and
- the Jarque-Bera test.

Each of these normality tests investigated the null hypothesis of normality, that is:

1. H_0: The sample follows a normal distribution.

Very often the null hypothesis had to be rejected, meaning that the sample was not normally distributed, thus violating the assumptions of the ANOVA. When this happens, there are two options available for the analysis of the non-normal data (Statistica Electronic Manual, 2005).

The first option is to make use of a non-parametric test for analysis (Statistica Electronic Manual, 2005). However, these non-parametric tests often provide less powerful analysis of the data (Statistica Electronic Manual, 2005), a point which is perfectly illustrated by the fact that the only viable non-parametric alternative for analysis of the data was the Kruskal-Wallis test. This test is in actual fact the non-parametric equivalent to the one-way ANOVA and as such it does not allow for multivariate analysis. This would mean that a series of Kruskal-Wallis analyses would have to be performed, each time including a different independent variable. Much the same as multiple t-tests instead of a single ANOVA weakens the results and increases the possibility of wrongly rejecting a true null hypothesis (Walsh and Ollenburger, 2001) the same would occur if multiple Kruskal-Wallis tests were run in place of a single analysis on all the data at once. Another disadvantage of performing multiple univariate Kruskal-Wallis tests is that this would not allow for inspection of the interaction effects between the various factors.

The second option is to ignore the fact the data is not normally distributed and use the ANOVA test for analysis (Statistica Electronic Manual, 2005). This option is supported by the application of the Central Limit Theorem which dictates that even though the variable under investigation does not follow a normal distribution within the population, the sample mean of sufficiently large samples will approximate a normal distribution (Dowdy and Weardon, 1983; Electronic Statistics Textbook, 2005; Statistica Electronic Manual, 2005). The Central Limit Theorem comes into play provided that the sample size is “large”, where “large” is considered to be a sample size larger than 30 (Statistica Electronic Manual, 2005; Walsh and Ollenburger, 2001) or according to some literature, a sample size larger than 100 (Electronic Statistics Textbook, 2005). Since the number of observations analysed was always well above 300, the Central Limit Theorem was applicable to the sample of experienced users.

Serving as further motivation to use the ANOVA, is the fact that the ANOVA has been shown to be robust under the violation of the assumptions of normality of data (Berenson and Levine, 1979; Dowdy and Weardon, 1983; Electronic Statistics Textbook, 2005; Statistica Electronic Manual, 2005), as well as the homogeneity of variances when the sample group is large enough (Berenson and Levine, 1979; Dowdy and Weardon, 1983; Electronic Statistics Textbook, 2005). Moreover, the ANOVA has been shown to be even more robust than its non-parametric equivalent under violation of the assumption of normality (Finch, 2005).

All of these factors contributed to the final decision to use the ANOVA to analyse the data whether it was normal or not. With this in mind, the results of the normality tests for this group will not be reported, even though the tests were conducted for each analysis.

7.4.1.4 Score

Each user was assigned a weighted score which was calculated as the sum of the difficulty indices of all the tasks completed successfully by that user. The maximum attainable score for the test was 47.

The weighted score for members of group A was analysed using a 2 (Expertise) × 5 (Interface) × 2 (Language) factorial ANOVA. The following hypotheses were formulated:

1. $H_{0,1}$: The word processor expertise of the user has no effect on the test score.
2. $H_{0,2}$: The interface used has no effect on the test score.
3. $H_{0,3}$: Whether the test was completed in the user's L1 or in L2 has no effect on the test score.

$H_{0,1}$ was rejected ($F_{Expertise}(1, 425) = 27.73$, $p = 0 < 0.001$), indicating that the word processor expertise of the users did indeed have an effect on the score achieved by them. Neither $H_{0,2}$ ($F_{Interface}(4, 425) = 1.10$, $p = 0.356 > 0.05$) nor $H_{0,3}$ ($F_{Language}(1, 425) = 0$, $p = 0.947 > 0.05$) could be rejected, leading to the conclusion that neither the interface nor the language had an effect on the score achieved by the user.

7.4.1.5 Time

A 2 (Expertise) × 5 (Interface) × 2 (Language) factorial ANOVA was performed for each task to evaluate the time taken to complete that task. The following hypotheses could be formulated for each task:

1. $H_{0,1}$: The word processor expertise level of the user has no effect on the time taken to complete the task.
2. $H_{0,2}$: The interface used has no effect on the time taken to complete the task.
3. $H_{0,3}$: Whether the test is conducted in the user's L1 or L2 has no effect on the time taken to complete the task.

The results of the ANOVA are summarised in Table 7.2.

Table 7.2: Group A – 1/Time ANOVA results

Task no	n	Expertise	Interface	Language	Expertise * Interface	Interface * Language	Expertise * Language	Expertise * Interface * Language
1	364	0.004*	0.897	0.577	0.570	0.214	0.321	0.930
2	395	0.000*	0.002*	0.940	0.379	0.181	0.759	0.743
3	406	0.000*	0.989	0.743	0.134	0.344	0.392	0.080
4	419	0.000*	0.779	0.886	0.377	0.873	0.441	0.835
5	270	0.000*	0.621	0.867	0.541	0.926	0.166	0.156
6	360	0.000*	0.402	0.468	0.280	0.413	0.171	0.973
7	396	0.001*	0.147	0.697	0.894	0.017*	0.962	0.876
8	425	0.000*	0.225	0.802	0.880	0.201	0.418	0.831
9	424	0.000*	0.041*	0.624	0.510	0.268	0.527	0.673
10	405	0.002*	0.755	0.567	0.895	0.794	0.386	0.625
11	409	0.002*	0.848	0.579	0.801	0.558	0.088	0.475

* $p < 0.05$

An α level of 0.05 was used throughout to distinguish between significant and non-significant differences. As would be expected, expert users performed significantly better than first-time users with $H_{0,1}$ being rejected for all of the tasks. $H_{0,2}$ could be rejected for the tasks that require a single word to be made bold ($p = 0.002$, task 2) and a phrase to be italicised ($p = 0.041$, task 9). $H_{0,3}$ could not be rejected for any of the tasks at an α level of 0.05, therefore it could be concluded that the interface language had no effect on the time needed by the user to complete the task successfully.

7.4.1.6 Number of actions

A 2 (Expertise) × 5 (Interface) × 2 (Language) factorial ANOVA was performed for each task to evaluate the number of actions required by users to complete that task. The following hypotheses could be formulated for each task:

1. $H_{0,1}$: The word processor expertise level of the user has no effect on the number of actions required to complete the task.
2. $H_{0,2}$: The interface used has no effect on the number of actions required to complete the task.
3. $H_{0,3}$: Whether the test is conducted in the user's L1 or L2 has no effect on the number of actions required to complete the task.

The p-values of the conducted ANOVAs are tabulated below.

Table 7.3: Group A – Number of actions ANOVA results

Task no	n	Expertise	Interface	Language	Expertise * Interface	Interface * Language	Expertise * Language	Expertise * Interface * Language
1	364	0.084	0.524	0.975	0.006*	0.495	0.881	0.390
2	395	0.001*	0.869	0.773	0.895	0.794	0.594	0.615
3	406	0.022*	0.104	0.639	0.128	0.749	0.487	0.644
4	419	0.010*	0.551	0.634	0.462	0.525	0.538	0.502
5	270	0.000*	0.234	0.540	0.166	0.445	0.991	0.044*
6	360	0.279	0.548	0.121	0.216	0.264	0.053	0.336
7	396	0.705	0.810	0.576	0.293	0.223	0.376	0.423
8	425	0.000*	0.615	0.743	0.439	0.167	0.756	0.518
9	424	0.003*	0.404	0.767	0.851	0.246	0.434	0.244
10	405	0.366	0.023*	0.777	0.267	0.599	0.823	0.502
11	409	0.096	0.238	0.993	0.486	0.032*	0.082	0.304

* $p < 0.05$

$H_{0,1}$ could be rejected for all but three of the tasks. $H_{0,2}$ could be rejected for a single task which required users to right align a paragraph. $H_{0,3}$ could not be rejected for any of the tasks, indicating that the interface language had no effect on the number of actions required to complete the task successfully.

7.4.1.7 Number of errors

For the analysis of this usability measure a 2 (Expertise) × 5 (Interface) × 2 (Language) factorial ANOVA was performed for each task. The following hypotheses could be formulated for each task:

1. $H_{0,1}$: The word processor expertise level of the user has no effect on the number of errors made during completion of the task.
2. $H_{0,2}$: The interface used has no effect on the number of errors made during completion of the task.
3. $H_{0,3}$: Whether the test is conducted in the user's L1 or L2 has no effect on the number of errors made during completion of the task.

The p-values of the conducted ANOVA's are shown in Table 7.4.

Table 7.4: Group A – Number of errors ANOVA results

Task no	n	Expertise	Interface	Language	Expertise * Interface	Expertise * Language	Interface * Language	Expertise * Interface * Language
1	364	0.112	0.023*	0.876	0.016*	0.494	0.239	0.727
2	395	0.097	0.010*	0.086	0.258	0.420	0.011*	0.979
3	406	0.001*	0.045*	0.103	0.109	0.106	0.518	0.641
4	419	0.078	0.046	0.693	0.152	0.765	0.083	0.050
5	270	0.431	0.668	0.438	0.329	0.612	0.876	0.244
6	360	0.840	0.330	0.228	0.189	0.766	0.170	0.364
7	396	0.794	0.715	0.192	0.052	0.508	0.052	0.132
8	425	0.034*	0.990	0.400	0.531	0.327	0.786	0.624
9	424	0.171	0.114	0.198	0.025*	0.360	0.019*	0.089
10	405	0.284	0.018*	0.523	0.733	0.533	0.661	0.173
11	409	0.522	0.607	0.197	0.652	0.017*	0.400	0.774

* $p < 0.05$

$H_{0,1}$ and $H_{0,2}$ could each only be rejected for a handful of tasks. $H_{0,3}$, on the other hand, could not be rejected for any of the tasks. This proved that the interface language did not affect the error rate of the users while completing the tasks.

7.4.1.8 Task results

For the analysis of the final usability measure, namely that of the number of correctly answered tasks versus the number of incorrectly answered tasks, a Chi-square analysis was performed for each task. For simplicity the independent variables were separated and tested individually by means of a one-dimensional Chi-square analysis. For the sake of brevity and since no other usability measure was influenced by the language of the interface, only the results of the language analysis for the task results will be reported here.

Since there were five different interfaces and two levels of word processor expertise, this resulted in ten individual Chi-square analyses per task divided as per the list below:

- A comparison between the results for first and second language users for all the first-time users of each interface.
- A comparison between the results for first and second language users for all the expert users of each interface.

The following hypotheses could be formulated for both the first-time and expert users of each interface and for each task:

1. $H_{0,1}$: Whether the test is conducted in the user's L1 or L2 has no effect on the success rate of the task.

The results of the Chi-square analysis are shown in the table below:

Table 7.5: Group A – Task results Chi-square analysis

		First-time users					*Expert users*				
Interface		**1**	**2**	**4**	**5**	**7**	**1**	**2**	**4**	**5**	**7**
Task no	**n**										
1	416	--	--	--	--	--	--	--	--	--	--
2	445	--	--	--	--	--	--	--	--	--	--
3	445	--	--	--	--	--	--	--	--	--	--
4	445	--	--	--	--	--	0.332	0.648	--	0.215	0.396
5	445	0.963	0.776	0.821	0.635	--	--	--	--	0.807	--
6	443	--	--	--	--	--	--	--	--	--	--
7	444	--	--	--	--	--	--	--	--	--	--
8	445	--	--	--	--	--	--	--	--	--	--
9	444	--	--	--	--	--	--	--	--	--	--
10	443	--	--	--	--	--	--	--	--	--	--
11	445	--	--	--	--	--	--	--	--	--	--

* $p < 0.05$

-- Insufficient observations for meaningful analysis

Five observations are considered to be the minimum allowable number of observations required in each cell in order to perform a meaningful Chi-square analysis (Berenson and Levine, 1979; Walsh and Ollenburger, 2001). Due to the high success rate of the tasks and the fact that the incorrect answers had to be further subdivided between the different interfaces and the user expertise, many of the tasks had an insufficient number of observations for a reliable Chi-square analysis to be performed. The success rate of neither the first-time users nor the expert users for the tasks which had an adequate number of observations (tasks 4 and 5) was affected by the interface language.

The chart below (Chart 7.1) shows that the success rate of all but one task, task 5, had a success rate of over 80%. Task 5 required users to create a bulleted list. Even though this task had a substantially lower success rate than the other tasks, the Chi-square analysis still failed to reject the null hypotheses for this task.

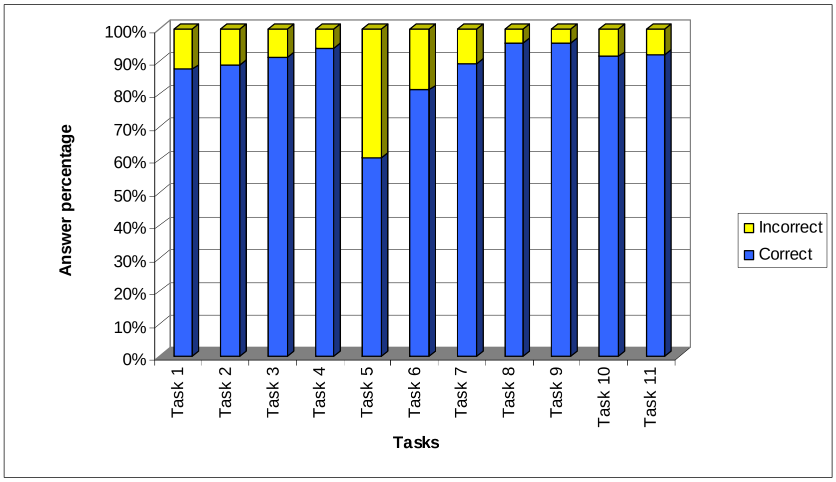


Chart 7.1: Group A – Task success rate

7.4.1.9 Subjective user satisfaction

Unfortunately, time constraints prevented many users from completing the satisfaction questionnaire, which resulted in substantially fewer questionnaires being completed than the total number of users. In total, there were 297 acceptably completed QUISs for group A.

Since each subsection consisted of a different number of questions, the mean score for each section was calculated per user. This ensured that each subsection could have a minimum score of 1 and a maximum score of 9. The total usability value was calculated as the average of all the questions together.

Descriptive statistics for the completed satisfaction questionnaires are shown below:

Table 7.6: Group A – Subjective satisfaction descriptive statistics

Usability Section	n	Mean	Minimum	Maximum	Std. dev.
Overall reaction	297	6.464	2.330	9.000	1.410
Learning	297	6.673	2.000	9.000	1.556
System capabilities	297	6.614	2.000	9.000	1.480
Total	297	6.574	2.440	8.880	1.246

For each of these sections, including the questionnaire total, a 2 (Expertise) × 5 (Interface) × 2 (Language) factorial ANOVA was performed in order to determine whether there was a difference between the satisfaction levels of the different user groups. The following hypotheses were evaluated for each subsection of the satisfaction questionnaire as well as for the overall satisfaction scores of the questionnaire:

1. $H_{0,1}$: The word processor expertise level of the user has no effect on the subsection of subjective satisfaction.
2. $H_{0,2}$: The interface used has no effect on the subsection of subjective satisfaction.
3. $H_{0,3}$: Whether the interface is in the user's L1 or L2 has no effect on the subsection of subjective satisfaction.

The p-values (α = 0.05) of each ANOVA are shown in the table below.

Table 7.7: Group A – Subjective satisfaction ANOVA results

	Expertise	Interface	Language	Expertise * Interface	Expertise * Language	Interface * Language	Expertise * Interface * Language
Overall reaction	0.000*	0.184	0.032*	0.477	0.343	0.241	0.894
Learning	0.000*	0.963	0.424	0.047*	0.154	0.117	0.230
System capabilities	0.000*	0.777	0.319	0.054	0.082	0.525	0.449
Total	0.000*	0.850	0.102	0.096	0.101	0.228	0.541

* $p < 0.05$

A significant difference (α = 0.05) was found between first-time and expert users for all of the usability subsections. In all of these cases, the experts gave a higher satisfaction rating than the first-time users. Due to the interaction effects present between the expertise and interface factors for the learning subsection, the difference between the satisfaction ratings could be attributed mainly to the significantly higher satisfaction rating given by expert users of interface 2 when compared with first-time users of both interfaces 2 ($p = 0.014$) and 7 ($p = 0.005$) and between the expert users of interface 7 and first-time users of both interfaces 2 ($p = 0.021$) and 7 ($p = 0.018$).

$H_{0,2}$ could not be rejected for any of the tasks, thus the interface used did not affect the satisfaction experienced by the user during interaction with the prototype.

Surprisingly, L1 users indicated a considerably lower overall satisfaction with the system than L2 users. This could be due to the fact that many of the users had some experience with a word processor, even though in some cases it was very limited experience. Given this fact, the terminology that now had to be used in a foreign language, where foreign means that it was not the English terminology that they had become accustomed to, lowered the satisfaction of the users.

7.4.1.10 Discussion

None of the analysed objective usability measures were affected by the language of the interface and only a single subjective usability measure was influenced by the interface language. Therefore, it can be concluded that having the interface in the first language of the user does not increase the efficiency or effectiveness of the application. Since only those interfaces which could be distinguished on the basis of the interface language were included in the analysis, and the language did not affect the performance of the users, the need to characterise the interfaces according to the language was no longer required in order to analyse the usability of the interfaces.

7.4.2 Consolidation of the groups

Group B users were removed from the initial analysis since the interfaces had neither a menu nor tooltips, thus containing no language component. Users of this group had to rely entirely on the interpretation of the icon and as such could not be included in the analysis. Since it was shown that language had no effect on any of the usability measurements, apart from the single satisfaction rating, the need to separate the groups no longer existed and the two groups could be consolidated into a single group for the remaining analysis. Therefore, the two groups were amalgamated and the results were re-analysed without taking the language, or lack thereof, of the interface into consideration. The distribution of the users amongst the interfaces, once they had been grouped together and divided only according to the interface used and their expertise rating, is shown in Table 7.8.

Table 7.8: User distribution with language factor removed

	Interface	First-time	Expert	Group Total
1	Standard icons, Menu, Tooltips	50	43	
2	Standard icons, No menu, Tooltips	46	39	
3	Standard icons, No menu, No tooltips	19	15	
4	Alternative icons, Menu, Tooltips	39	40	
5	Alternative icons, No menu, Tooltips	46	49	
6	Alternative icons, No menu, No tooltips	17	21	
7	Text buttons, No menu, Tooltips	48	45	517
Total		**265**	**252**	

The analysis of this group includes users of all seven interfaces listed previously (Sections 4.4.4 and 7.4). The same usability measurements as for group A were analysed and measurements were calculated in the same manner as for group A. Apart from these established usability measures, the results of the computer anxiety and computer attitude levels for the consolidated group will also be reported upon.

7.4.2.1 User anxiety

In order to calculate each participant's anxiety level, questionnaire responses were multiplied by their respective factor loadings, which indicates the weight of the contribution the question has to the overall anxiety score (Marcoulides, 1989). The multiplied answers for each question were summed for each category, the sum of which was then taken as the score for that category.

7.4.2.1.1 Independent variables

The independent variables for the anxiety analysis were word processor expertise and interface used.

7.4.2.1.2 Dependent variables

The dependent variables for this analysis were the individual user scores for general anxiety, equipment anxiety and overall anxiety.

7.4.2.1.3 Analysis

In order to analyse the anxiety of the test participants a 2 (Expertise) × 7 (Interface) between subjects factorial ANOVA was conducted. Results are tabulated below:

Table 7.9: Consolidated group – Anxiety ANOVA results

	n	Mean	Min.	Max.	Std. dev.	Expertise	Interface	Expertise * Interface
General	390	35.015	10.690	57.140	10.866	0.010*	0.031*	0.001*
Equipment	390	11.065	3.690	22.140	4.690	0.066	0.191	0.076
Overall	390	46.080	14.380	77.860	14.842	0.014*	0.041*	0.003*

* $p < 0.05$

7.4.2.1.4 Discussion

Significant differences between the general anxiety levels were detected for both independent variables and the interaction between the two variables for the general anxiety and overall anxiety factors. Understandably, where a significant difference between user expertise levels was present, first-time users displayed higher anxiety than their expert counterparts. These differences can, however, be attributed to certain interface groups due to the presence of significant interaction effects. Tukey's highest significant difference (HSD) post-hoc test indicated that the same groups were responsible for the detected difference in both the general and overall anxiety measures. The detected differences, at an α level of 0.05 and with the first p-value corresponding to the general anxiety and the second to the overall anxiety, were mainly due to the difference in anxiety levels between:

- The expert users of interfaces 2 and 5 ($p = 0.005$; $p = 0.012$);
- Expert users of interface 5 and first-time users of interface 7 ($p = 0.006$; $p = 0.07$); and
- The expert users of interface 5 and first-time users of interface 6 ($p = 0.017$; $p = 0.018$).

In all three of these observations the expert users of interface group 5 had significantly lower anxiety levels than users of the other interface groups for both anxiety categories.

Taking a closer look at the overall anxiety levels of the different interface groups, Chart 7.2 clearly shows two distinct anxiety level groupings. Users in interface groups 1, 4 and 5 had a lower computer anxiety level, whilst the remainder of the interface groups cluster at a higher anxiety level. Users in interface group 1 had the lowest combined anxiety level, whereas interface group 3 displayed the highest anxiety levels of the seven interface groups.

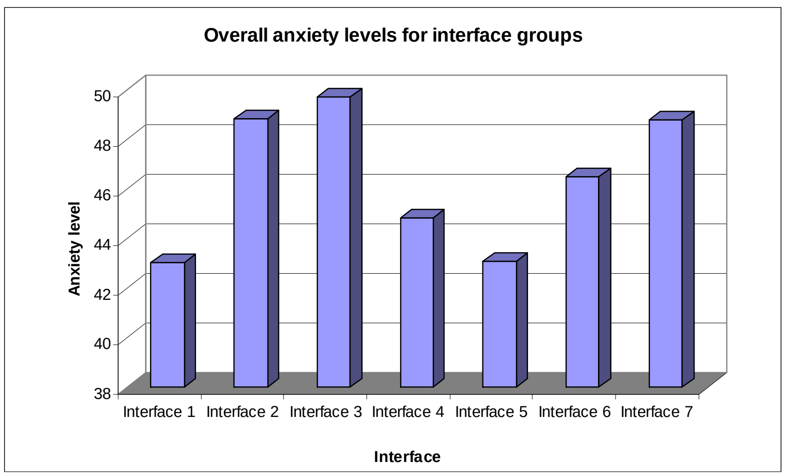


Chart 7.2: Anxiety levels for consolidated interface groups

As mentioned previously, it was determined that expert users had a lower anxiety rating when compared to the first-time users in the consolidated group. However, when comparing the first-time and expert users for the individual interfaces, some interesting trends appeared, as can be seen on Chart 7.3. For interfaces 1, 2 and 3 – all the interfaces with the standard icons – expert users had higher anxiety levels than the first-time users of those groups. Thereafter, for the remainder of the interface groups the expert users displayed lower anxiety levels than the first-time users.

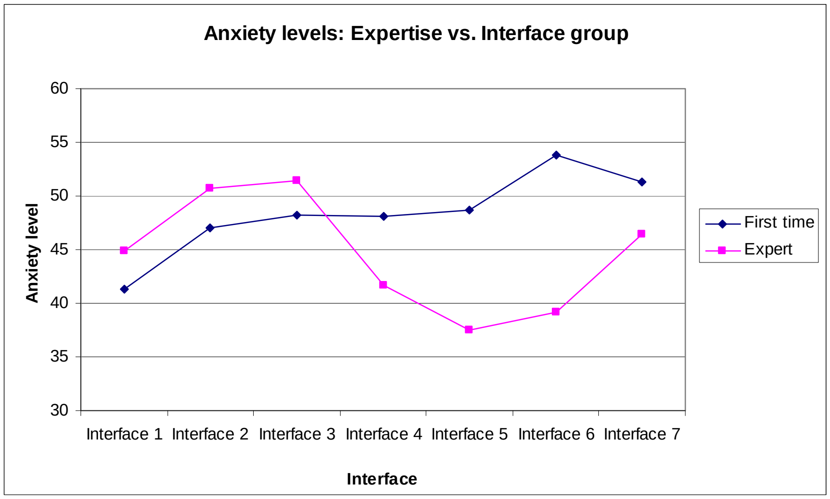


Chart 7.3: Consolidated group – Expertise vs. interface group anxiety levels

7.4.2.2 User attitude

Answers to negatively worded questions were first recalibrated by simply inverting the scale. That is, all answers of 1 were recalibrated to an answer of 6, 2 became 5 etc. Once this had been done, answers were multiplied by their respective factor loadings (Bandalos and Benson, 1990) and the sum of each category was taken as the score for that category. As with the possible maximum and minimum scores, final scores for each category were rounded off to 2 decimal points to ensure consistency with the anxiety questionnaire. The overall attitude score is the sum of the individual attitude subsections. High attitude scores correlate with a positive attitude towards computers.

7.4.2.2.1 Independent variables

The independent variables for this analysis were the interface and the word processor expertise.

7.4.2.2.2 Dependent variables

The dependent variables were the four calculated attitude scores for each user, namely liking, confidence, achievement and an overall attitude score.

7.4.2.2.3 Analysis

In order to analyse the attitude of the users towards computers a 2 (Expertise) × 7 (Interface) factorial ANOVA was conducted. A summary of the results are tabulated below.

Table 7.10: Consolidated group – Attitude ANOVA results

	n	Mean	Min.	Max.	Std. dev.	Expertise	Interface	Expertise * Interface
Liking	393	26.017	6.160	36.940	6.474	0.034*	0.110	0.093
Confidence	393	31.963	10.530	43.120	6.606	0.000*	0.925	0.291
Achievement	393	21.680	8.260	26.780	3.685	0.003*	0.890	0.503
Overall	393	79.659	30.330	106.850	14.167	0.000*	0.621	0.121

* $p < 0.05$

7.4.2.2.4 Discussion

Word processor expertise affects a users' attitude towards computers for all three subsections as well as for the overall attitude of a user towards computers. In all of these cases, the expert users showed a significantly higher, and therefore, more positive attitude towards computers. Thus, it would seem as though experience dictates a users' attitude towards a computer and once the user gains more experience their attitude becomes more positive towards a computer.

No interface group showed a particularly more positive or negative attitude towards computers and no interaction effects were present. As is evident in Chart 7.4, the attitude of the different interface groups remained relatively constant with only the attitude of users in interface group 6 being considerably higher than the rest.

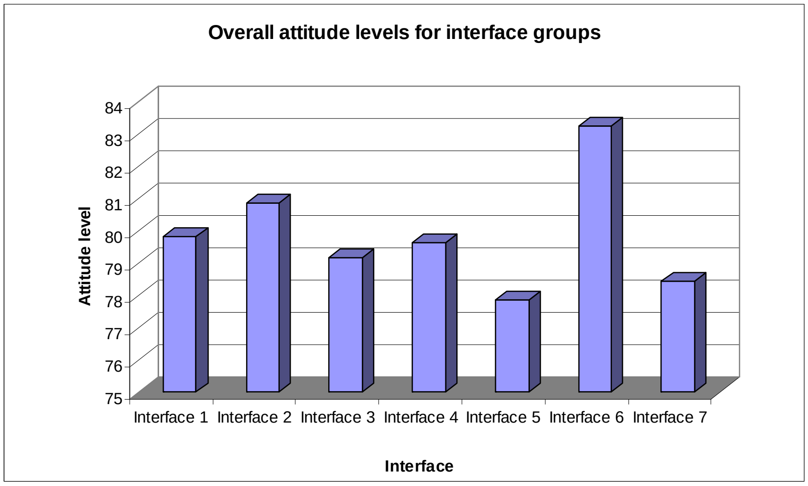


Chart 7.4: Attitude levels for consolidated interface groups

Chart 7.5 shows that first-time users of all interface groups have a more negative attitude than the expert users, an observation that was confirmed by the statistical tests. Only interface 6 and 7 indicate minimal differences in the attitudes between the different user groups.

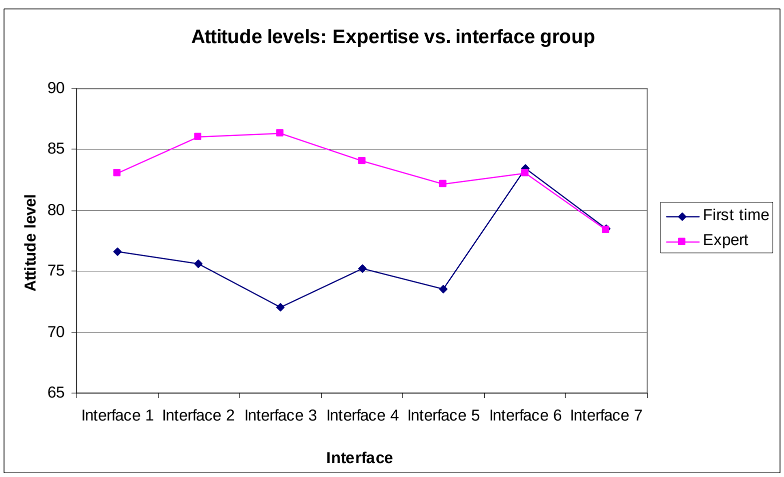


Chart 7.5: Consolidated group – Expertise vs. interface group attitude levels

7.4.2.3 Usability analysis

The usability analysis for the consolidated group consisted of evaluation of the same 6 usability measures which were analysed for both group A and the novice users.

7.4.2.3.1 Independent variables

The independent variables used in the analysis were the interface employed during the test and the word processor expertise of the user.

7.4.2.3.2 Dependent variables

For each analysis, the calculated usability measure is the dependent variable.

7.4.2.3.3 Normality of the data

The same normality rules and arguments as for group A are also applicable and were used for the consolidated group.

7.4.2.3.4 Score

The weighted score for this user group was analysed using a 2 (Expertise) × 7 (Interface) factorial ANOVA. The following hypotheses were formulated:

1. $H_{0,1}$: The word processor expertise of the user has no effect on the test score.
2. $H_{0,2}$: The interface used has no effect on the test score.

$H_{0,1}$ was rejected ($F_{Expertise}(1, 503) = 26.47$, $p = 0 < 0.001$) indicating that the word processor expertise of the user did indeed have an effect on the score achieved by the user. The interface used had no effect on the score achieved by the users since $H_{0,2}$ ($F_{Interface}(6, 503) = 2.01$, $p = 0.063 > 0.05$) could not be rejected.

7.4.2.3.5 Time

For each task, a 2 (Expertise) × 7 (Interface) factorial ANOVA was performed to evaluate the time taken to complete that task. Once again, only those tasks that were completed successfully were included in the analysis. The following hypotheses could be formulated for each task:

1. $H_{0,1}$: The word processor expertise level of the user has no effect on the time taken to complete the task.
2. $H_{0,2}$: The interface used has no effect on the time taken to complete the task.

The p-values of the resultant ANOVA are summarised in Table 7.11.

Table 7.11: Consolidated group – 1/Time ANOVA results

Task no	n	Expertise	Interface	Expertise * Interface
1	418	0.022*	0.774	0.305
2	457	0.000*	0.001*	0.375
3	467	0.000*	0.845	0.087
4	486	0.000*	0.339	0.121
5	312	0.029*	0.781	0.133
6	396	0.002*	0.001*	0.279
7	457	0.000*	0.256	0.891
8	493	0.000*	0.281	0.830
9	492	0.000*	0.006*	0.318
10	464	0.002*	0.419	0.530
11	467	0.004*	0.120	0.555

* $p < 0.05$

$H_{0,1}$ could be rejected for all of the completed tasks since expert users were able to complete the task significantly faster than first-time users for all of the tasks. $H_{0,2}$ could be rejected for three of the eleven tasks. Possible reasons for these observations are discussed below.

- Task 2: Change font style of a single word to bold (p = 0.001)

A significant difference was found between the users of interface 2 and interface 7 (p = 0.001) as well as between users of interface 2 and interface 6 (p = 0.038). This indicates that the icons contributed significantly to the performance of the user. In both cases, users of the standard icon had a shorter completion time than users of the other two interfaces. This would indicate that the standard icon for **Bold** is extremely intuitive and succeeded in conveying the concept of bold to the user, even more so than the word "Bold" on the button. Since only those users of the alternative icons where no tooltips were used showed a significantly longer completion time than those of the standard icons, it would seem as though the tooltips assisted the users in deciphering the function linked to the icons for the remainder of the alternative icon users.

- Task 6: Close a document (p = 0.001)

With an average completion time of 53 seconds, users of the alternative icons with no menu and no tooltips (interface 6) took longer to close a document than all other users, who completed the task in times ranging from 18 seconds to just marginally longer than 20 seconds. The number of correct answers to this task was also the second lowest of all the tasks. The alternative icon for the **Close** function was chosen by questionnaire respondents, but the results of this task show that it did not successfully communicate the concept of **Close** when used in an interface without any tooltips or menus to assist the user. This result indicates that user preference does not automatically ensure increased user performance, confirming the sentiment which has suggested by some authors (McDougall and Curry, 2004) In fact, the icon chosen by the respondents was actually designed as an alternative for an electronic mail interface. The icon appeared to be acceptable when used in conjunction with a tooltip which could lead the user to the correct icon required for the task at hand. Taking into consideration that choices of illiterate computer users were split evenly between the icon eventually used for **Save** and the one used for **Close**, it may be pointed out that perhaps the entire concept of closing a document needs to be explained more clearly to novice or first-time users.

The novice users (Chapter 6) of this interface did not struggle any more that the users of the standard interface, showing that the meaning of any icon can be learned by users. After these users had been taught that the icon was used to close a document, they did not struggle to complete the task any more than the users of the standard icons did. However, the users of the experienced group had to interpret the icons for themselves and this clearly shows that the icon is unsuccessful since users were unable to extract any meaning from the icon without having assistance in the form of tooltips or training.

- Task 9: Italicise a phrase ($p = 0.006$)

Tukey's HSD post-hoc test, at a 0.05 significance level, indicated that the most significant difference occurred between users of interface 2 and interface 6 ($p = 0.027$) as well as between users of interface 4 and interface 6 ($p = 0.042$). Alternative icon users with no tooltips and no menu (interface 6) had a significantly longer average completion time than the users of the other two mentioned interfaces. These results indicate that the alternative italic icon did not succeed in conveying the function to the user. However, once again, inclusion of a tooltip to indicate the icon purpose assisted the users in determining the functionality linked to the icon.

7.4.2.3.6 Number of actions

The number of actions required by the users to successfully complete each task was analysed next. The following hypotheses were formulated for each task:

1. $H_{0,1}$: The word processor expertise level of the user has no effect on the number of actions required to complete the task.
2. $H_{0,2}$: The interface used has no effect on the number of actions required to complete the task.

An individual 2 (Expertise) × 7 (Interface) between-subjects factorial ANOVA was conducted for each task, the results of which are tabulated below:

Table 7.12: Consolidated group – Number of actions ANOVA results

Task no	n	Expertise	Interface	Expertise * Interface
1	418	0.221	0.108	0.000*
2	457	0.002*	0.969	0.888
3	467	0.169	0.273	0.185
4	486	0.017*	0.766	0.634
5	312	0.159	0.529	0.036*
6	396	0.627	0.005*	0.403
7	457	0.609	0.896	0.367
8	493	0.000*	0.704	0.349
9	492	0.002*	0.515	0.909
10	464	0.282	0.066	0.494
11	467	0.184	0.596	0.977

* $p < 0.05$

$H_{0,1}$ could be rejected for four tasks, where first-time users required significantly more actions to complete the task successfully than expert users. These tasks were all font styling tasks, namely the task that required users to bold a single word, two tasks that required users to underline a single word and the task that required users to italicise a phrase.

- Task 2: Change font style of a single word to bold ($p = 0.002$)
- Task 4 and task 8: Underline a single word ($p = 0.017$ and $p = 0$)
- Task 9: Italicise a phrase ($p = 0.002$)

These four tasks were the only tasks that required users to first make a selection – either of a word or a phrase (two or more words). The results indicate that first-time users required more actions to complete the task successfully, possibly because they were unaware that the desired word had to be selected before font styling could be applied. These users may first have selected the correct option but without having selected the word or phrase. Users may then have continued to attempt to complete the task by selecting other icons or menu options before eventually realising that in order to achieve the desired effect the word or phrase must first be selected. Even once the user realised that a selection had to be made, they may have experienced difficulty in making the selection, in particular the multiple word selection.

$H_{0,2}$ could be rejected for the task that required users to close a document. In this case, the interface did have an effect on the number of actions needed by the user to complete the task successfully. Users of the alternative interface with neither a menu nor tooltips carried out considerably more actions before completing the task successfully than users of all the other interfaces. These users also took considerably longer to complete the task than users of the other interface groups (see previous section). The higher number of actions could account for the longer completion time for these users, since more actions would require more time to carry them out in – time and actions increase proportionally. Closer inspection of the number of errors made will determine whether the higher number of actions were in fact erroneous selections on the part of the users of this interface.

There were interaction effects present for the task that required users to open a document (task 1) and for that which required the creation of a bulleted list (task 5).

- Task 1: Open a document ($p = 0.000$)

  Post-hoc tests indicate that the significant interaction effect detected can be attributed to the difference between the combinations of user expertise and interface group as indicated by the listed significant p-values ($\alpha = 0.05$) in Table 7.13.

  First-time users of interface 6 required significantly more actions to complete the task successfully than users of the other interfaces in the pairings listed above. Concerning the remaining difference, expert users of interface 3 executed more actions than expert users of interface 4. First-time users may have struggled more to find the correct option since there were no leading tooltips in interface 6 to guide the users to the correct icon option. Even expert users of the standard icons without tooltips or menus required more actions than the users of the alternative icons with tooltips and a menu, showing once again, that the icons seem to work best when used in conjunction with a tooltip, which allows users to make an informed decision on the invocation of the correct icon option.

Table 7.13: Consolidated group – Task 1 post-hoc p-values

		First-time users	***Expert users***
		Interface 6	**Interface 3**
First-time users	**Interface 7**	0.022	
Expert users	**Interface 2**	0.020	
	Interface 4	0.004	0.045
	Interface 5	0.014	

- <u>Task 5</u>: Create a bulleted list (p = 0.036)

No individual user groups showed any distinct differences in the number of actions required to complete the task (the most distinct difference was between experts of 1 and first-time users of 4 with p = 0.095). The two graphs below, which plot the mean number of actions and the mean time to complete the task for each interface group, show that the time and actions experience the same fluctuations for both expert and first-time users.

Chart 7.7 further illustrates the fact that task 5 resulted in the highest number of actions required to complete the task. This is understandable since the task required that users place bullets in front of multiple lines of text. Since it was observed that many users placed the bullets individually instead of selecting the list as a whole, an average of 14 actions for this task is quite possible since the open document consisted of four to five lines of text. If the user had to navigate to each line and then insert a bullet, that would account for at least eight to ten actions.

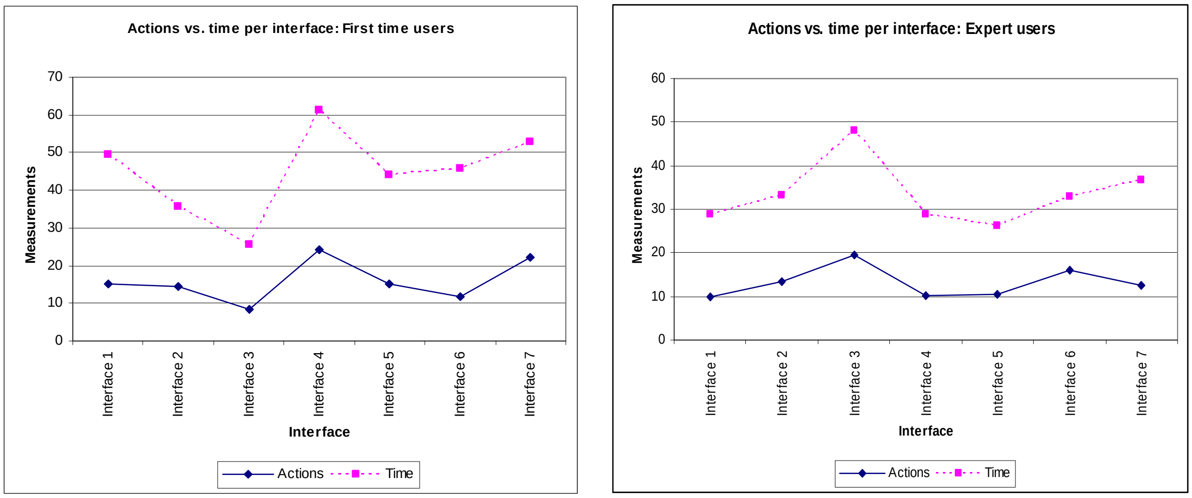


Chart 7.6: First-time and expert users – Time vs. number of actions per interface

The increased number of actions for task 5 could also be attributed to the nature of the action that is carried out when the user clicks on the bullets icon. Since a bullet is either inserted at the start of the line or removed from the line if there is already a bullet present, users may have required an increased number of actions to complete the task if the cursor was not positioned correctly when the bullets option was selected.

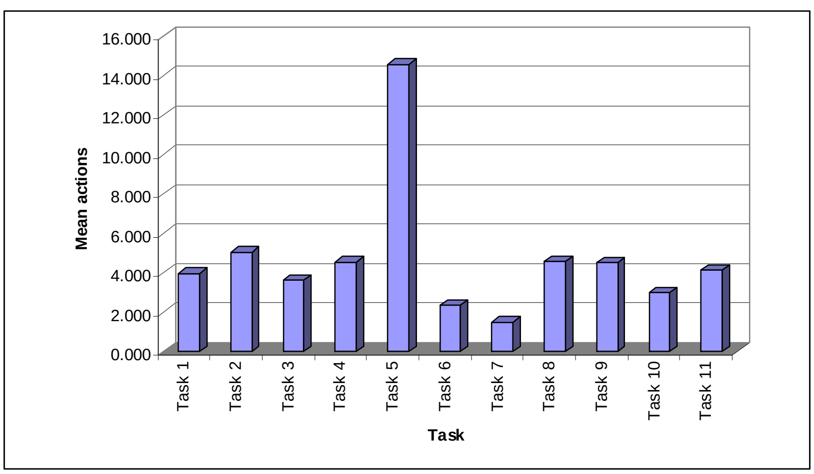


Chart 7.7: Consolidated group – Task mean actions

7.4.2.3.7 Number of errors

A 2 (Expertise) × 7 (Interface) between-subjects factorial ANOVA was conducted for each task in order to explore the validity of the following formulated hypotheses:

1. $H_{0,1}$: The word processor expertise level of the user has no effect on the number of errors made during completion of the task.
2. $H_{0,2}$: The interface used has no effect on the number of errors made during completion of the task.

The p-values of the ANOVA are shown in Table 7.14:

Table 7.14: Consolidated group – Number of error ANOVA results

Task no	n	Expertise	Interface	Expertise * Interface
1	418	0.457	0.000*	0.000*
2	457	0.105	0.008*	0.443
3	467	0.062	0.205	0.107
4	486	0.058	0.009*	0.280
5	312	0.526	0.925	0.027*
6	396	0.670	0.000*	0.611
7	457	0.977	0.796	0.114
8	493	0.015*	0.916	0.502
9	492	0.112	0.160	0.017*
10	464	0.195	0.036*	0.561
11	467	0.666	0.002*	0.890

* $p < 0.05$

$H_{0,1}$ could be rejected for the task that required users to underline a word since first-time users made significantly more errors during these tasks than their expert counterparts did. Although only the second underlining task indicates a significant difference between the first-time and expert users, the first underlining task (task 4) has a p-value of 0.058 and as such is not significant at the 0.05 level, but it is significant at the 0.1 level.

- <u>Task 8</u>: Underline a single word ($p = 0.015$)

It was found, by observation of the participants during the test, that many users did not realise that the word had to be selected before a font style or effect could be applied to the word or phrase. This may have been more prevalent in the first-time user group than the expert user group. Should the users have selected the correct icon but failed to have selected the required word beforehand, they may have become discouraged by the fact that there was no reaction by the system. This may have convinced first-time users that they had made an incorrect selection, with the result that these users would then attempt to find the correct icon in order to complete the task. The error rate for these users would have increased during this process of elimination of the icons. Since none of the other icons would yield the correct result either, users could eventually conclude that an additional action was required of them before completion of the task. With this in mind, users could then have selected the required word and then selected their original icon choice, which would then yield the desired effect of underlining.

Both of the underlining tasks showed a significant difference between the number of actions required by the first-time and the expert users during completion of the task. Since only distinct errors were analysed, and a significant difference was detected between these users for both errors and actions, the process of elimination theory is cemented as a plausible explanation for the detected differences.

$H_{0,2}$ could be rejected for a number of tasks ($\alpha = 0.05$) namely, opening and closing a document, bolding a word, underlining a word and aligning a paragraph.

- Task 1: Open a document ($p = 0.000$)

  A significant difference was detected between the number of errors made by the users of interface 1 and interface 6 ($p = 0.042$), the users of interface 3 and interface 4 ($p = 0.011$), the users of interface 4 and interface 6 ($p = 0.003$) and the users of interface 6 and interface 7 ($p = 0.026$). However, the presence of significant interaction effects means that the significant difference found between the interfaces can mainly be attributed to the difference between the interface and expertise combinations listed in Table 7.15. The table lists the significant p-values for the interaction effects at an α level of 0.05.

Table 7.15: Consolidated group – Task 1 post-hoc p-values

		First-time users	***First-time users***	***Expert users***
		Interface 5	**Interface 6**	**Interface 3**
First-time users	**Interface 1**		0.016	0.014
First-time users	**Interface 4**		0.016	0.014
First-time users	**Interface 7**		0.014	0.012
Expert users	**Interface 2**		0.020	0.018
Expert users	**Interface 4**	0.035	0.002	0.001
Expert users	**Interface 5**		0.020	0.018
Expert users	**Interface 7**		0.034	0.032

The experts of interface 3 made more errors than the users of the other interfaces involved in the interactions. Similarly, the first-time users of interface 6 made more errors than the rest of the listed interface users. Concerning the last interaction, the first-time users of interface 5 made more errors during completion of the task than the expert users of interface 4. Since neither interface 3 nor interface 6 had tooltips, the interaction effects lend even more credence to the prior assertion that the tooltips served to assist the user in selection of the correct icon.

In order to fully evaluate the nature of the errors made by the users, a closer inspection of the actual incorrect selections was done for the task. Once the errors had been listed and grouped, a number of interesting discoveries were made, all of which are discussed in greater detail below.

Although the number of erroneous menu selections only made up a small percentage of the total errors (9%), the fault percentages still indicate an interesting, very obvious inclination. As can clearly be seen on Chart 7.8, a high percentage of users incorrectly selected the menu option for a new document before correctly opting to select the open icon.

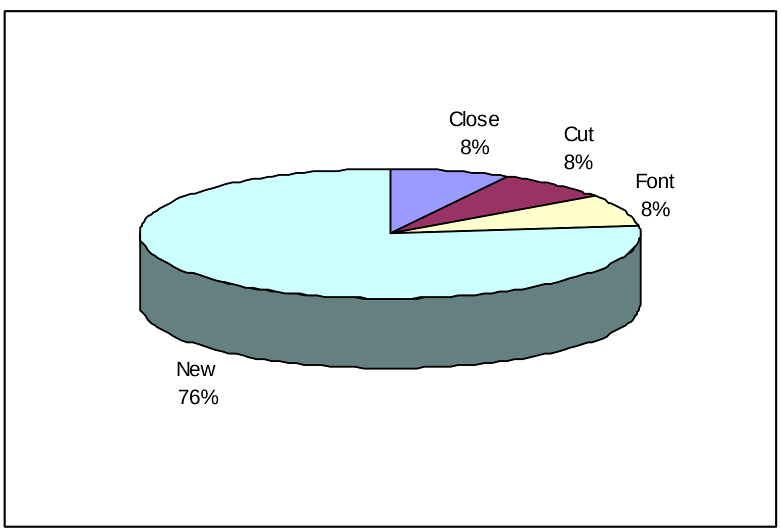


Chart 7.8: Consolidated group – Erroneous menu selection for task 1

Chart 7.9 indicates the fault percentage for the icon selection. Once again, New features prominently in the error percentage, with only two interfaces (interface 1 and

6) having **New** as the minority error chosen. All other interfaces had **New** chosen as the incorrect icon choice for well over 50% of the errors, with two of the interfaces (4 and 7) having a 100% error rate for **New**. When disregarding **New**, the remainder of the errors, listed in the graph as "Other icon", were distributed amongst a number of icon choices. Bearing in mind that the errors constituting the remainder were spread over at least another 2 choices, with one task having an error distribution over 10 other icon choices, it is clear that a problem exists between **New** and **Open**. This claim is reinforced by the fact that when calculated in total, regardless of interface, the **New** icon accounts for 58% of the iconic errors made.

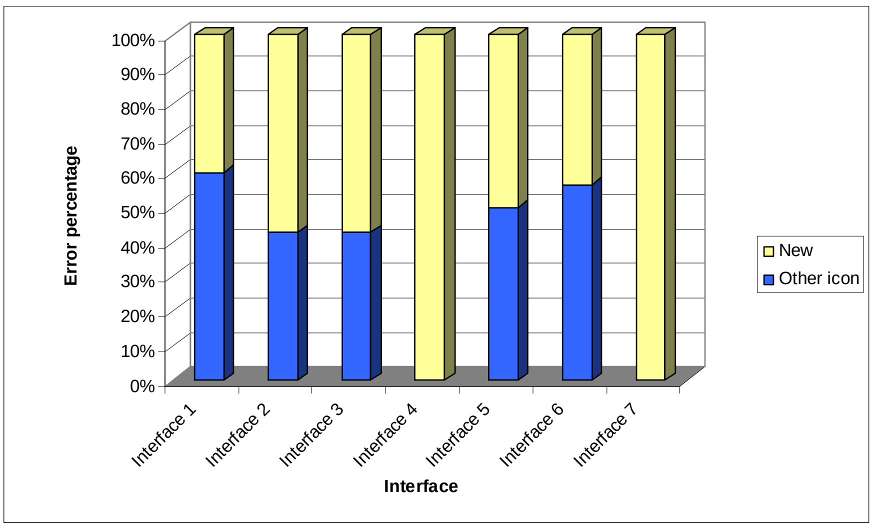


Chart 7.9: Consolidated group – Erroneous icon selection for task 1

- Task 2: Make a single word bold (p = 0.08)

Tukey's HSD post-hoc test indicates that the significant difference between the performance of the interface groups can be attributed to the difference between the users of interface 1 and interface 6 (p = 0.038). Users of interface 6, alternative icons with no tooltips, made significantly more errors than the users of the standard icons.

No incorrect menu items were selected during completion of this task. The incorrect icon selections were spread evenly among the different interfaces, with users of standard icons making 32% of the total errors, alternative icon users 39% and text button users 29%. The text button errors were distributed amongst a number of different incorrect selections, with no button having a clear selection majority. The standard icon errors had a majority error for the **Cut** icon, followed closely by the **Paste** icon. Alternative users chose the **Font colour** icon incorrectly in the majority of cases.

- Task 4: Underlining a single word (p = 0.009)

The most significant difference detected by Tukey's HSD post-hoc test was between the users of interface 6 and interface 7 (p = 0.122), with interface 7 users making fewer errors. The text-based icon interface (interface 7) also had the lowest mean error rate for this task.

Although no incorrect menu items were chosen by the users during execution of this task, comparison of the incorrect icons did however once again show that one particular icon was chosen incorrectly in the majority of cases, namely that of **Font colour**. Although the total of other incorrect icons – all incorrect choices less the selections for **Font colour** – had a higher selection percentage for some interfaces, the choices are once again spread over at least two other icon choices and even in these cases the margin between the **Font colour** icon and other icons was very small. When examining the total selections for each icon individually and not grouping the icons, the **Font colour** icon was always in the majority for the interfaces where it was erroneously chosen.

When combining the icons into standard, alternative and text icons, the error percentage for the standard **Font colour** icon is 54%, for alternative 36% and for text 0%.

Informal observation of the users during the test had already highlighted this phenomenon of incorrectly choosing the **Font colour** icon in place of the underline icon for the standard icons. When observing the users during the test, it became evident that many users chose this icon as their first choice for underline. Even when users requested assistance with the task and it was enquired of them which icon looked as though it would underline a piece of text, most users responded that the **Font colour** icon appeared to be the icon for that purpose. Since scrutiny of the errors only serves to confirm this observation, the discussion of this problem will simply focus on the standard icons. The standard icons for **Font colour** and **Underline** are shown below:


Font colour icon


Underline icon

As can clearly be seen it would be easy for a first-time user to confuse the two icons since they both consist of a single alphabetic character with a line underneath. Even though the line used for **Font colour** changes according to the font colour chosen by the user in standard word processing applications, it still defaults to black when the application is opened, with the result that the two icons are very similar to one another in appearance until a different font colour is chosen. Quite possibly, the prominence of the line under the "A" of the **Font colour** icon accounts for the fact that it seems to draw the eye of the user to the icon more easily the much thinner line under the letter in the **Underline** icon. Designing icons so that they are all visually distinct from one another so as to avoid confusion on the part of the user is a strongly recommended usability guideline (Ellis, Tran, Roo and Shneiderman, 1995). Clearly, these icons do not comply with that guideline.

- Task 6: Closing a document without saving ($p = 0.000$)

Users of interface 6, alternative icons with no menu and no hints, made more errors than users of all the other interfaces. Users of this interface also required more actions as well as more time in which to complete the task successfully. All of these facts together prove that the choice made by the non-computer literate users for the **Close** icon was in fact an unsuccessful icon as it neglected to convey the correct icon function to the user when it needed to be interpreted without the aiding presence of tooltips and menu. Thus, the icon does work in conjunction with tooltips or a menu but is in itself not intuitive enough to allow interpretation without additional interface components to guide the user.

The percentage of errors that can be attributed to mistakenly chosen menu selections is negligible (only 6%), and further analysis on the menu items will not be done. With the exception of interface 3, all the other interfaces have a majority error of the **Open** icon. The icon with the second highest error rate for all interfaces using the alternative icons is the **Save** icon. This would seem to correspond with the choice made by the non-computer literate users who completed the icon selection questionnaire compiled during the first brainstorming session, and whose choices for **Close** were split evenly between the **Close** and **Save** icons eventually used for the alternative interface. The fact that the bulk of the erroneous selections was the **Open** icon for all but one of the interfaces, once again signifies the confusion between the different document handling actions.

- Task 10: Right aligning a paragraph (p = 0.036)

Users of interface 2 made significantly fewer errors than the users of the text button interface (p = 0.013). Users of the text button interface made the greatest number of errors on average of all the interface groups. This may be due to the fact that the button for the text interface only had the word “Right” on it, as opposed to the question which required users to *align* the paragraph to the right. The tooltip did however provide more context with the full function description of “Right align” appearing when the mouse moved over the button.

However, closer inspection of the actual errors made revealed that although users of interface 7 made the most errors, these errors were widespread across a number of the icons. Overall, not many incorrect selections were made during completion of this task. Another interesting phenomenon showed that of the five interfaces where errors were made, four of those had the most errors for **Left Align**, whereas the task required users to right align a paragraph. For these interfaces, the **Left Align** either had a clear majority or was tied with another icon as having the highest error rate. In total, the **Left Align** icon accounted for 29% of the errors made, which made the **Left Align** icon the most incorrectly selected icon in total.

The problem of selecting the **Left Align** was observed amongst these test subjects, and it was also seen to be more pronounced with the school-going users. Centring did not seem to cause a problem as both the standard and alternative icon (, ) seemed to be very indicative of aligning text in the centre of a page. Once users understood what **Centre Align** meant they had almost no trouble identifying the correct icon to use. **Right**

align, however, did present a problem for these users. The jagged *right* edge of the ***Left align*** icon for both the standard and alternative icons (, ) appears to attract the attention of the user, causing them to infer that the icon is in fact the **Right align** icon. Although more experienced users will recognise that the straight, aligned edge indicates which alignment is applied by selection of the icon, first-time users may at first struggle to differentiate between the icons until they have been shown the difference between the alignment styles and what alignment specifically means in terms of text editing.

- Task 11: Saving and closing a document ($p = 0.002$)

Tukey's HSD post-hoc test showed that the most significant difference occurred between the users of interfaces 5 and 6 ($p = 0.035$), that is the alternative interface with no menu but with tooltips and the alternative interface with neither a menu nor tooltips. The most errors were made, on average, by the users of interface 6 during completion of this task. Since this task included having to close a document and users of this interface made the most errors on average for the previous task that only required users to close a document, it would seem that closing a document still caused problems and users were unable to retain the previously gained knowledge. Users of the standard interface with no menu and no tooltips made the second most errors on average during completion of this task. Thus, once again, the result is indicative of the fact that tooltips assisted users in determining the correct icon to be chosen for the task and the absence of tooltips increased the number of errors made by the users.

The icons for **Open** (30%), **New** (17%) and **Undo** (16%) accounted for the most errors made during the task. On closer inspection of the breakdown of the error selections, it was clear that these icons were incorrectly chosen by users of all the interfaces.

Post-hoc tests failed to discover which user group or which interfaces were responsible for the significant interaction effect present in task 5. The significant interaction effects present in task 9, where users had to italicise a phrase, could be traced to the significant difference between expert users of interface 5 and first-time users of interface 6, as well as between first-time users of interface 5 and first-time users of interface 6. In both of these cases, the first-time users of interface 6 made significantly more errors than the users of interface 5. Interfaces 5 and 6 both used the alternative icons, the only difference between the two interfaces being the presence of tooltips in interface 5. Once again, it would seem as though the tooltips assisted users in identifying the correct icon needed to complete the task correctly.

7.4.2.3.8 Task results

Only task 5, creating a bulleted list, had enough observations for group A to conduct a meaningful Chi-square analysis in order to determine whether or not interface language had an effect on the success rate of the task. Since language did not affect the outcome of this task, an analysis could be done for this task for the consolidated group.

Two independent Chi-square analyses were conducted for this purpose. First the observations were split into the seven different interfaces and seven individual Chi-square analyses were conducted in order to test the following hypothesis for each interface:

1. $H_{0,1}$: Word processor expertise has no effect on whether the task is completed successfully or not.

The results of the Chi-square analysis are tabulated in Table 7.16:

Table 7.16: Consolidated group – Task results Chi-square analysis results

	Chi-square p-value
Interface 1	0.348
Interface 2	0.696
Interface 3	--
Interface 4	0.030*
Interface 5	0.712
Interface 6	0.360
Interface 7	1

* $p < 0.05$
-- Insufficient observations for meaningful analysis

Word processor expertise only had an effect on the success rate for users of interface 4. The rejection of $H_{0,1}$ in this case further verifies the fact that first-time users of interface 4 had difficulty in completing this task successfully, since significantly fewer first-time users were able to complete the task than expert users.

Secondly the observations were split between the first-time and expert users and two Chi-squares were performed to determine (a) whether there was a difference between the interfaces for the first-time users alone and (b) whether there was a difference between the interfaces for the expert users alone. For these analyses the following hypotheses were formulated:

1. $H_{0,1}$: The interface used has no effect on whether first-time users complete the task successfully or not.
2. $H_{0,2}$: The interface used has no effect on whether expert users complete the task successfully or not.

$H_{0,1}$ could not be rejected ($p = 0.547$), thus the interface did not affect the success rate of first-time users. There were insufficient observations for the expert users since interface 3 only had three incorrect answers for expert users for this task. Therefore, $H_{0,2}$ could not be investigated further.

Chart 7.10 shows the percentage of correct and incorrect answers for each task. The majority of the tasks had a success rate of at least 85%. The tasks with success rates lower than 85% were:

- Task 5: Creating a bulleted list;
- Task 6: Closing a document without saving.

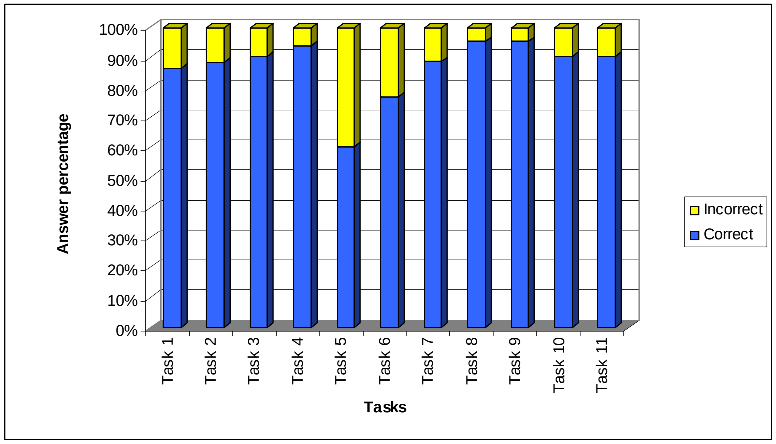


Chart 7.10: Consolidated group – Task success rate

The tasks had the following success rates, in descending order, with the success percentage rounded off to the nearest decimal:

Table 7.17: Consolidated group – Task success rate percentages

Task no	Task description	Success rate percentage	Difficulty index
Task 8	Underline a word	95%	4
Task 9	Italicise a phrase	95%	4
Task 4	Underline a word	94%	4
Task 10	Right align a paragraph	90%	3
Task 11	Save and close a document	90%	8
Task 3	Centre align	90%	4
Task 7	Create a new document	88%	3
Task 2	Bold a single word	88%	4
Task 1	Open a document	86%	5
Task 6	Close a document without saving	77%	3
Task 5	Create a bulleted list	60%	5

Task 5 had the lowest success rate. It was previously shown that this task required the most actions to complete. The difficult nature of the task, as discussed previously, may also have contributed to its low success rate of the task. Interestingly, the first font styling task, namely that of bolding a single word, had the fourth lowest success rate, but the remainder of the font styling tasks make up the tasks with the three highest success rates. Task 4, which had the third highest success rate, was the second font styling task. Tasks 8 and 9 both had a success rate of 95% and were the final two font styling tasks. These results demonstrate a progressive increase in success rates each time the users were required to apply a font style change. Consequently it would seem that the users retained the knowledge that they had previously gained and were able to apply it to similar tasks later on.

Also, tasks with low success rates were not necessarily the tasks with the highest difficulty indices. A Pearson's correlation coefficient substantiates this with a weak negative correlation between the success rate percentage and the difficulty index ($r = -0.062$). The correlation between the task difficulty index and the success rate is not a significant one at $\alpha = 0.05$ ($t = 0.185$, $df = 9$, $p = 0.858$). With this in mind it may be possible to assign difficulty levels using a method other than the allocation of the number of actions and inferences required to successfully complete the tasks.

7.4.2.3.9 Subjective user satisfaction

The mean score for each user was calculated per subsection. Each subsection could have a minimum usability score of 1 and a maximum usability score of 9. Descriptive statistics for the consolidated group were as follows:

Table 7.18: Consolidated group – Descriptive statistics for user satisfaction

	n	Mean	Minimum	Maximum	Std. dev.
Overall reaction	330	6.449	1.170	9.000	1.431
Learning	330	6.661	1.000	9.000	1.586
System capability	330	6.587	2.000	9.000	1.539
Total usability	330	6.555	2.440	8.880	1.278

To evaluate the subjective user satisfaction experience after use of the application the following hypotheses were formulated for each of the satisfaction subsections:

1. $H_{0,1}$: The word processor expertise level of the user has no effect on the subsection of subjective satisfaction for the application.
2. $H_{0,2}$: The interface used has no effect on the subsection of subjective satisfaction for the application.

For each subsection of the satisfaction scale a 2 (Expertise) × 7 (Interface) between-subjects factorial ANOVA was conducted, the results of which can be found in Table 7.19.

Table 7.19: Consolidated group – Subjective satisfaction ANOVA results

	Expertise	Interface	Expertise * Interface
Overall reaction	0.003*	0.349	0.715
Learning	0.000*	0.292	0.136
System capability	0.000*	0.174	0.138
Total usability	0.000*	0.295	0.313

*$p < 0.05$

$H_{0,1}$ could be rejected for all four of the subsections of the questionnaire (p =0.003; 0; 0; 0) with experts indicating higher satisfaction with the system in all four of these categories. The lower satisfaction of the first-time users could be due to the fact that first-time users became frustrated at the length of the test and the difficulties experienced in completing all the tasks successfully.

$H_{0,2}$ could not be rejected for any of the sections, proving that the interface used does not influence the subjective satisfaction experienced when using a system.

7.4.2.3.10 Discussion

In summary, the usability measures obtained from the tests with this group revealed the following tendencies with regard to the two variables:

Expertise

Anxiety and attitude levels were loosely linked to the experience of the user, with more experienced users having a more positive attitude towards computers and experiencing lower anxiety levels when interacting with a computer.

There was a marked difference between the performance of first-time and expert users, with the expert users being able to complete all tasks quicker and being capable of achieving significantly higher scores for their tests. Expert users required fewer actions to complete the four font styling tasks of bolding, underlining and italicising text. First-time users made more errors than the expert users during completion of the two underlining tasks. These results indicate that the first-time users struggled with the completion of the font styling tasks due to the increased errors and actions. The correct icons or menu options were not immediately recognised and users employed more of a physical elimination process in order to determine the correct options. The remainder of the tasks were completed quicker by the expert users but not with significantly fewer actions or errors. Therefore, the first-time users simply took longer to process and identify the correct icon or menu option, but without an increased error or action rate. Thus, they did not resort to selecting icons in an attempt to identify the correct icon any more than the expert users did and quite possibly followed a mental approach to eliminate the icons.

The expertise of the user influenced the subjective satisfaction, with more experienced users having a more enjoyable interaction with the system.

Interface

The alternative icon for **Close** was an unsuccessful icon, since the alternative interface with no menu and no tooltips required more time and actions and resulted in more errors being made by the users of this interface during completion of tasks that required closing of a document. This icon did, however, function equally as well as the standard icon when used in combination with a tooltip or menu. There were a few other tasks which were influenced by the interface used, but the greatest impact of the interface was felt during the tasks which required users to close a document, and the font styling tasks. Inclusion of tooltips provided positive results in terms of increased usability, since even the icons which functioned badly on their own showed comparable usability when included in an interface with tooltips.

Subjective satisfaction was not influenced by the interface used during completion of tasks.

7.5 Summary

To analyse the group of task-knowledgeable participants, the users were divided into groups according to whether the interface they used had a language component or not. The group with menus and tooltips incorporated into the interface were designated group A and were analysed first due to the fact that this was the larger group. As in the previous evaluations, users were divided into the interface groups according to their L1. Therefore, a user could either use an interface that was in his/her L1, or an English interface, where English was not the first language of the user.

The analysis of group A made provision for the independent variables of expertise level, interface configuration used and whether the interface was in the L1 or L2 of the user. As expected, expert users scored significantly better on the test than novice users did. Neither the interface nor the language of the interface affected the score of the user. The results of the analysis of the time, actions, errors and success rate of the tasks for group A is tabulated in Table 7.20. The tick marks indicate that a significant difference was detected for that task and that variable.

Table 7.20: Group A – Significant difference summary

	Expertise			***Interface***			***Language***			
Task no	**Time**	**Actions**	**Errors**	**Time**	**Actions**	**Errors**	**Time**	**Actions**	**Errors**	**Success rate**
1	✓					✓				
2	✓	✓		✓		✓				
3	✓	✓	✓			✓				
4	✓	✓								
5	✓	✓								
6	✓									
7	✓									
8	✓	✓	✓							
9	✓	✓		✓						
10	✓				✓	✓				
11	✓									

As can clearly be seen from the above table, language had no effect on any of the usability measurements. Users of the L1 interface indicated a lower overall satisfaction rating, possibly due to their limited knowledge of the terminology in that language.

Since interface language exerted no influence over any of the usability measures, the two groups were consolidated and re-evaluated as a single group. The evaluation of the consolidated group included two independent variables of interface used and expertise level of the user. Similar to group A, the evaluation of the consolidated group showed that expert users scored significantly better on the test than their first-time user counterparts, although the interface did not affect the score achieved by the users. The time, action, error and success rate analysis of the consolidated group is shown in Table 7.21. Once again, the tick marks indicate a significant difference detected for that particular task and variable.

Table 7.21: Consolidated group – Significant difference summary

	Expertise				***Interface***			
Task no	**Time**	**Actions**	**Errors**	**Success rate**	**Time**	**Actions**	**Errors**	**Success rate**
1	✓						✓	
2	✓	✓			✓		✓	
3	✓							
4	✓	✓		✓			✓	
5	✓							
6	✓				✓	✓	✓	
7	✓							
8	✓	✓	✓					
9	✓	✓			✓			
10	✓						✓	
11	✓						✓	

As would be expected, expert users required significantly less time to complete the tasks correctly than did the first-time users. Some of the tasks they also managed to compete correctly using fewer actions and making fewer errors. However, all of these results were to be expected since expert users should generally be able to manage a higher performance than first-time users. The results do, however, serve to validate the expertise scale that was developed to differentiate between those users who know "a little" and those who know "a lot". More pertinent to the study is the effect that the interface had on the performance of the user.

The following significant differences were detected for the consolidated group:

- Users of interface 6 on average took the longest and required more actions than any other users to close a document (task 6).
- Users of interface 2, standard icons with tooltips but no menu, were able to bold a single word in significantly less time than users of interfaces 6 and 7 (task 2).

- Users of interface 6 required significantly more time to italicise a phrase than users of interface 2 and 4 (task 9).
- Users of interface 6 made significantly more errors than users of interface 1 while attempting to make a word bold (task 2).
- Users of interface 6 made significantly more errors than users of interface 7 when having to underline a word (task 4).
- Users of interface 6 made more errors than all other users when closing a document (task 6).
- Users of interface 7 made significantly more errors than users of interface 2 when right aligning a paragraph (task 10).
- Users of interface 6 made significantly more errors than users of interface 5 when saving and closing a document (task 11).

From these significant relationships revealed during the analysis of user interaction, it could be determined that the alternative icon for **Close** was unsuccessful even though it had been chosen as the preferred icon by non-computer literate users. Of the users of alternative icons, it was only the users of interface 6, which had neither tooltips nor a menu, who struggled to complete the tasks with comparable performance to other interface users. From this it could be concluded that the tooltips and the menus assisted the users to select the correct icon when the depiction alone was not sufficient for correct interpretation.

The standard icons for **Bold**, **Italic** and **Underline** appeared to be the best suited to convey the meaning of the associated functions. Of the three icon sets, the standard icons facilitated faster recognition and fewer errors during font styling tasks. However, the **Underline** and **Font colour** icons are very similar to each other and resulted in misinterpretation of the **Font colour** icon.

Analysis of the errors indicated a distinct confusion between the document handling functions of **New**, **Open**, **Close** and **Save**. This could be attributed to the abstract nature of file manipulation. There is no actual physical file or document that is opened, closed, saved or created. Instead it is an abstract concept that is manipulated within the metaphor of a word processing environment. Confusion could also be caused by the fact that the terms are fairly interchangeable within an office setting which requires filing and to which the metaphor refers to (Section 9.3.2.1).

Expert users indicated a higher satisfaction rating for all four satisfaction subsections, but no interface groups had a significantly different satisfaction rating.

Overall, the findings of this chapter would suggest that there is no need for a translated interface for a word processor application for fluent bilingual users although translation does not impede or hamper the performance of the user in any way. Icons influenced the performance of the users for certain tasks and their impact on the interaction with the application will allow for suggestions to be made as to the development of an adapted interface which is more suited to the needs of word processor users with the same profile as those in this study. The adapted interface will be discussed in greater detail in Chapter 9.

In the following chapter a more in-depth analysis of the menu and icon selection during completion of the tasks will be presented. The errors made by the experienced users who were unable to successfully complete the tasks will be investigated, together with the selection rate of the menus and icons.

Chapter 8

Menu and icon error and selection analysis

8.1 Introduction

The previous chapter discussed the findings of the user tests with the more task-knowledgeable users. It was found that language had no impact, and the interface had a limited influence on the performance of the users. When a significant difference was detected between the error rates of the various interfaces an inspection analysis was conducted on the errors which revealed a number of interesting findings. However, the previous chapter only incorporated the tasks that were completed correctly in each analysis. It is important to examine the tasks of the users who were unable to complete the tasks correctly in order to establish whether there was any particular reason preventing them from successfully carrying out a task. If so, taking these problems into account when designing an interface will increase the usability of the interface as users will no longer be prevented from completing tasks successfully. This chapter will focus on the evaluation and discussion of the errors made by the task-knowledgeable users who were unable to complete the tasks correctly.

Furthermore, in the development of interfaces it is important to determine which interaction method is the most popular and widely used interaction method. This chapter will include a discussion on the selection rate of menus and icons during completion of the tasks to determine which interaction method is the most popular amongst word processor users.

8.2 Menu and icon error rate

Using the results of the same task-knowledgeable participants that were analysed in the previous chapter, the menu and icon errors made during completion of the tasks were analysed for those users who were unable to complete the tasks successfully. Therefore, the distribution of the users remains the same as for the analysis of the consolidated group analysed in the previous section and is shown in Table 8.1.

Table 8.1: Menu and icon error rate user distribution

	Interface	First-time users	Expert users	Group Total
1	Standard icons, Menu, Tooltips	50	43	
2	Standard icons, No menu, Tooltips	46	39	
3	Standard icons, No menu, No tooltips	19	15	
4	Alternative icons, Menu, Tooltips	39	40	
5	Alternative icons, No menu, Tooltips	46	49	
6	Alternative icons, No menu, No tooltips	17	21	
7	Text buttons, No menu, Tooltips	48	45	517
Total		**265**	**252**	

Given that some users were unable to correctly complete the tasks, the errors made by these users were evaluated to gain more insight into why they were unable to successfully complete the tasks, and into whether there was a particular misinterpretation that was prohibiting them from successful completion of the task. Evaluation of the task-knowledgeable users who completed the tasks correctly revealed certain trends and erroneous icons and menu options which were consistently chosen by the users. By evaluating the actions of users who could not complete the tasks correctly, it would be possible to ascertain whether there was any similarity between the errors made by the users who could eventually complete the tasks correctly and those who were unable to. Therefore, the errors of the users who managed to complete the task correctly will occasionally be compared with those of the users who were unable to complete the task correctly. Focusing on the problems encountered by these users during their unsuccessful attempts to complete the tasks could lead to the development of an environment that is more usable and easier to learn, thus lowering the time needed to learn to use the application. The number of correct and incorrect answers per task and the respective percentage of tasks completed correctly are tabulated in Table 8.2.

Table 8.2: Number of incorrect and correct answers per task

Task no	Total answers	Correct answers	Incorrect answers	Percentage correct answers
1	483	418	65	86.54
2	517	457	60	88.39
3	517	467	50	90.33
4	517	486	31	94
5	517	312	205	60.35
6	514	396	118	77.04
7	516	457	59	88.57
8	517	493	24	95.36
9	516	492	24	95.35
10	513	464	49	90.45
11	517	467	50	90.33

Similar to the analysis of the errors in the previous chapter, the comparisons and observations contained in this chapter were done by means of inspection of the error percentage of each incorrectly chosen icon or menu option. The analysis in Chapter 7 also showed that there was only a significant difference between the number of errors made by first-time and expert users for task 8. Therefore, the analysis of errors in this chapter will not distinguish between errors made by first-time users and those made by expert users. Some of the tasks had a significant difference between the number of errors made by users of the different interface groups (Section 7.4.2.3.7) and therefore, where possible the errors will be divided among the different interface configurations.

8.2.1 Task 1: Open a document

The same trend was observed for both the users who did not complete the task successfully and those who did – namely that the most incorrectly chosen menu option and icon was that of New. Incorrect menu options accounted for 14% of the total errors. Of those the New menu option was chosen in 60% of the cases. The consolidated group who completed the task correctly had an error percentage of 76% for the New menu option for this task.

The **New** icon totalled 62% of the erroneous icon selections for the incorrect tasks, which was only slightly more than for the correct tasks where the **New** icon accounted for 58% of the incorrect icons chosen. Regardless of whether the task was completed successfully or not, users making errors consistently chose the **New** menu option or icon. Even users of the text buttons, where the words "New" and "Open" were printed on the buttons, also incorrectly chose the **New** button as the majority error selection in both the correctly and incorrectly completed tasks. This confirms the prior assertion that there appears to be some confusion between the actions of **Open** and **New**, a situation which is possibly further hampered by users misunderstanding the pictorial icons.

8.2.2 Task 2: Bold a single word

Users who could not complete the task correctly did not make many errors. No errors were forthcoming from users of interfaces 3, 4 and 5, indicating that these interface groups had a 100% success rate for this task. Users of both interfaces 1 and 2 each made 40% of the errors, whilst users of interfaces 6 and 7 made 7% and 13% of the errors respectively. Error selections were fairly widespread for these users with no single icon really emerging as the majority error selection.

8.2.3 Task 3: Centre align a word

Of the icon errors made by the users who could not complete the task correctly, 21% of those errors were caused by selection of the **Underline** icon. However, only users of the alternative interfaces and the text buttons erroneously selected the **Underline** icon. The remainder of the errors were spread fairly evenly among a number of different icons. Very few incorrect menu items were selected by this user group.

8.2.4 Task 4: Underline a word

The breakdown of the errors for incorrect answers for this task was as follows: standard icons, 80%, alternative icons, 7% and text buttons 13%. Overall, the **Font colour** and the **Bold** icons were the two icons chosen incorrectly the most times by these users while attempting to complete the task. However, the **Font colour** icon was not in the vast majority as with the users who completed the task correctly. The incorrect selections were reasonably widespread amongst a number of icons.

8.2.5 Task 5: Create a bulleted list

The errors were distributed fairly evenly amongst the interfaces (Chart 8.1), although alternative icon users did make the most errors in total. Once again, a number of different errors accounted for the errors made. The inability of these users to complete the task successfully, the number of errors made as well as the varied nature of the errors indicates the difficulty experienced by the users in completing this task successfully. Two possible explanations exist for the difficulties experienced by these users. The first possibility is that users did not know what a bullet was and as such struggled to complete the task and simply randomly selected icons until abandoning their efforts as futile. Secondly, bullets had to be created on multiple lines and users may not have realised that the bullet gets created on the line where the cursor is positioned and should a line already have a bullet, clicking the bullet option will remove the bullet from that line.

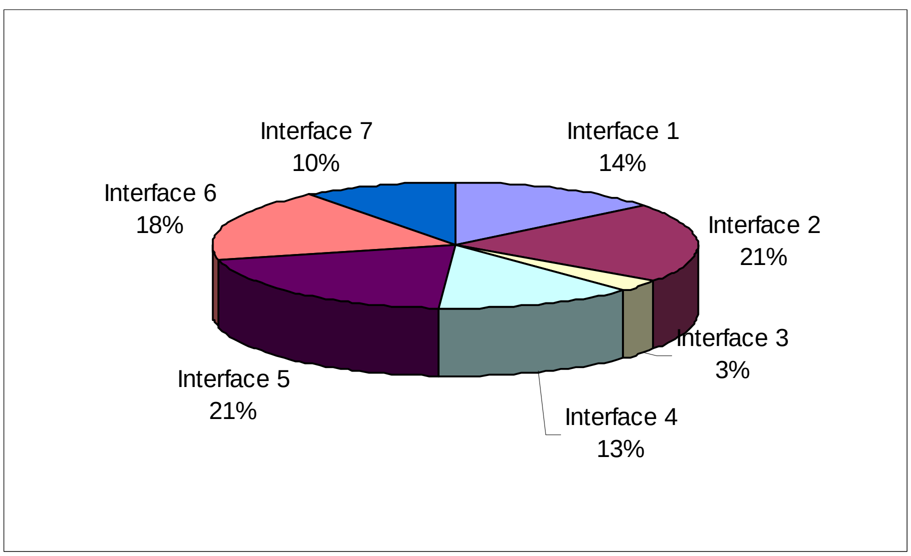


Chart 8.1: Task 5 – Interface incorrect error percentage

8.2.6 Task 6: Close a document

Incorrect menu selections made by these users for this task were negligible. The three error selections with the highest percentage for the consolidated group who completed the task correctly were: **Open** (30%), **Save** (19%) and **New** (13%). For the users who could not complete the task correctly, the three highest incorrect selections were: **New** (40%), **Open** (20%) and **Undo** (14%). As with the correct answers, these three with the highest selection percentage were by far in the majority, with the remainder of the errors being spread fairly evenly amongst a number of different icons. Both **New** and **Open** feature in the top three erroneous selections for these two groups. **New** also appeared to be a problem for both groups when attempting to open an existing document, further supporting the claim that there appears to be general confusion between the different document handling actions of **Open**, **Close**, **New** and **Save**.

8.2.7 Task 7: Create a new document

When creating a new document, the **Open** icon was the most incorrectly chosen menu item and icon for both the users who completed the task successfully and those who did not. The percentage breakdown for these two groups is shown below:

Table 8.3: Task 7 – Incorrect menu and icon selection percentages

	Correct answers	Incorrect answers
Open menu	94%	67%
Open icon	72%	60%

Once again, it would seem as though there was some confusion between **Open** and **New**. The task that required users to open a document had the **New** menu item and icon as the majority error selection for both the users who could complete the task and those who could not. Similarly, the task where a new document had to be created had the **Open** menu and icon as the majority error selection. Furthermore, for both incorrect and correct tasks the **Open** icon was also the most incorrectly chosen icon for each individual interface group.

8.2.8 Task 8: Underline a word

Task 8, together with task 9, had the highest success rate of all the tasks and there were also very few errors made by the users who could not manage to complete the task successfully, with no particular icon having a clear majority of erroneous selections and no incorrect menu selections being made.

8.2.9 Task 9: Italicise a phrase

This was the second task with the highest success rate (Section 7.4.2.3.8, Chart 7.10 and Table 7.17) and much the same as task 8, it also had no incorrect menu selections and very few incorrect icons chosen by the users who could not complete the task correctly. With so few incorrectly chosen icons, no particular icon emerged as the one chosen most often.

8.2.10 Task 10: Right align a paragraph

The error selections made by the users unable to complete the task were very widespread amongst the different icons, thereby not allowing for a meaningful comparison to be made. As with the users who completed the task successfully, the Left Align icon was chosen a number of times, but not as the majority icon and also not by users of all the interfaces. No menu items were incorrectly selected by these users.

8.2.11 Task 11: Save and close a document

Following the established inclination of selecting the incorrect document handling icon and confirming the results of incorrect error selections for the correctly completed tasks, the icons chosen incorrectly most often for the incorrect answers of this task were the New and Open icons, each with a 35% selection rate. Although there were very few incorrect menu selections, all were for the Open menu item.

8.2.12 Discussion

Many of the errors made by the users who were unable to complete the tasks successfully reflected those made by the users who could complete the tasks successfully. Therefore, there is nothing on the part of the interface that prevented these specific users from completing the task since their behaviour and selections echoed those of the users who could complete the tasks successfully. Two reasons are possible for their inability to complete the tasks correctly. Firstly, they may not have realised that they did not complete the task according to the specifications required and accepted that the task had been completed correctly and moved onto the next task. Secondly, they may not have persevered for as long as the users who had completed the task correctly, abandoning the task as impossible after only a few attempts at selecting the correct option.

The most valuable information obtained from the error analysis is the fact that users appear to be unable to distinguish between the document handling functions of **New**, **Open**, **Close** and **Save**. Secondly, the icons of **Underline** and **Font colour** were easily confused as were the icons of **Right align** and **Left align**.

8.3 Selection rate of menus and icons

The next step in the menu and icon analysis was to determine which interaction method was employed the most by the users during completion of the tasks. In order to do this, the actual selection rate of the menus was qualitatively compared to the actual selection rate of the icons. Only those interfaces which had a menu were included in this analysis, since inclusion of interfaces with no menus would cause a disproportional increase in icon selection without adequate representation of menu selection. Furthermore, the most popular interaction method could only be established when investigating which interaction method was used the most by the users who were given a choice between using icons or menus. The distribution of the users amongst the interfaces included in the analysis is shown in Table 8.4.

Table 8.4: Selection rate of menus and icons user distribution

	First-time users	Expert users	Total
Standard icons, Menu, Tooltips	50	43	**93**
Alternative icons, Menu, Tooltips	39	40	**79**
Total	**89**	**83**	**172**

Table 8.5 indicates the number of correct and incorrect answers per task for the users included in this analysis, as well as the percentage of correct answers per task for these users.

Table 8.5: Icon and menu selection task results

Task no	Total answers	Incorrect answers	Correct answers	Percentage correct answers
1	160	22	138	86.25
2	172	15	157	91.28
3	172	11	161	93.6
4	172	6	166	96.51
5	172	66	106	61.63
6	170	26	144	84.71
7	171	17	154	90.06
8	172	6	166	96.51
9	171	10	161	94.15
10	171	16	155	90.64
11	172	17	155	90.12

Table 8.6 shows the selection rates of the menu options and icons per expertise level and per interface group. All tasks, whether completed correctly or not, were included in the analysis. The selection rate for each interface group and each expertise level was calculated as a percentage of the total selections for each task. The percentages were calculated for all selections, including errors and duplicate selections. This would allow for a proper comparison between the two interaction methods, as well as for identification of the interaction method preferred by the users.

From Table 8.6 it can be deduced that menu selections between the different interfaces occurred on a comparable level. Icon selections were also fairly even among the different interfaces, although interface 1, which had standard icons, tended to have a slightly higher selection rate than interface 4 for most of the tasks for both first-time and expert users. The higher percentage of users in the interface 1 group, particularly with the first-time users, could account for the higher selection percentage for interface 1.

Table 8.6: Menu option and icon selection percentage per interface and per expertise level

	Menu selection					***Icon selection***				
	First-time users		**Expert users**			**First-time users**		**Expert users**		
Task no	**IF 1**	**IF 4**	**IF 1**	**IF 4**	**Total**	**IF 1**	**IF 4**	**IF 1**	**IF 4**	**Total**
1	14.93	12.50	14.58	12.85	**54.86**	15.97	6.94	18.75	3.47	**45.14**
2	0.00	0.00	0.00	0.38	**0.38**	40.77	23.46	19.23	16.15	**99.62**
3	0.00	0.53	0.00	0.00	**0.53**	31.38	22.34	23.94	21.81	**99.47**
4	0.00	0.44	0.00	0.44	**0.88**	39.82	20.35	21.68	17.26	**99.12**
5	5.19	5.06	7.59	5.73	**23.57**	30.36	16.25	14.25	15.58	**76.43**
6	16.93	12.70	14.81	18.52	**62.96**	10.58	10.58	10.58	5.29	**37.04**
7	11.71	13.96	17.12	11.71	**54.50**	14.86	8.11	13.51	9.01	**45.50**
8	0.00	0.45	0.00	0.90	**1.36**	28.05	24.43	27.15	19.00	**98.64**
9	0.00	1.01	0.51	0.51	**2.02**	33.33	19.19	24.24	21.21	**97.98**
10	0.00	2.78	0.00	0.00	**2.78**	31.11	20.00	25.00	21.11	**97.22**
11	18.48	15.62	15.81	17.71	**67.62**	14.29	5.71	7.81	4.57	**32.38**

IF 1: Interface configuration 1
IF 4: Interface configuration 4

Taking a closer look at the total menu selection percentage in Table 8.6, it can be seen that only four tasks have a selection of over 50%. These tasks were tasks 1, 6, 7 and 11 which required the users to (1) open a document, (6) close a document without saving any changes, (7) create a new document and (11) save and close a document respectively. These tasks were the only tasks which focused on document handling – actions that were discovered in the previous chapter and the previous section to have created problems for the users since many of them confused the actions. The even split of the selection percentages between icons and menus for these tasks and the fact that the other tasks had a higher selection rate for icons would indicate that the users eventually resorted to using the menus to complete the tasks in the event that they were unable to locate the correct icon. It is further interesting to note that the only other task with a menu selection percentage of over 10% is task 5, with a selection percentage of 23.57. This task required users to create a bulleted list and had the lowest success rate of all 11 tasks, with only 62% of the users being able to complete the task successfully.

Closer inspection of the nature of the selection by the users is required to determine the significance of this observed trend. Inspection of the manner in which the users select icons and menus might shed more light on how the users went about completing the task and whether they did in fact first attempt to locate the correct icon before using the menu to complete the task. Table 8.7 shows the percentage of the users who selected menus only and those who selected icons only during their attempt to complete the task. Once again, all tasks are included in the analysis, regardless of whether they were completed correctly or not. Selection percentages are also divided between the different interface groups and the different user expertise levels.

Table 8.7: Percentage of users who selected only a menu or an icon

	Menu selection					***Icon selection***				
	First-time users		**Expert users**			**First-time users**		**Expert users**		
Task no	**IF 1**	**IF 4**	**IF 1**	**IF 4**	**Total**	**IF 1**	**IF 4**	**IF 1**	**IF 4**	**Total**
1	16.56	12.58	13.25	18.54	**60.93**	9.93	5.96	5.96	4.64	**26.49**
2	0	0	0	0.62	**0.62**	30.25	19.75	26.54	22.84	**99.38**
3	0	0.63	0	0	**0.63**	29.38	20	25.63	24.38	**99.38**
4	0.59	0	0	0.59	**1.18**	29.41	22.35	25.29	21.76	**98.82**
5	5.59	4.35	4.97	6.83	**21.74**	19.25	16.77	16.77	15.53	**68.32**
6	17.53	13.64	16.88	21.43	**69.48**	11.04	9.09	6.49	1.95	**28.57**
7	11.9	13.69	10.71	12.5	**48.81**	16.67	8.33	11.31	10.71	**47.02**
8	0	0.58	0	0.58	**1.17**	29.24	22.22	25.15	21.64	**98.25**
9	0	1.19	0	0.6	**1.79**	28.57	21.43	25	22.62	**97.62**
10	0	0	0	0	**0**	28.57	21.12	26.09	23.6	**99.38**
11	18.24	14.71	13.53	18.82	**65.29**	8.24	3.53	4.12	2.35	**18.24**

IF 1: Interface configuration 1
IF 4: Interface configuration 4

The highest percentage of users who selected menus only are for tasks 1, 6, 7 and 11, which corresponds to the high percentages of menu items selected during completion of these tasks and as tabulated in Table 8.6.

Table 8.8 tabulates the percentage of users who selected both a menu and an icon during completion of each task.

Table 8.8: Percentage of users who selected both a menu and an icon

	First-time users		*Expert users*		
Task no	**IF 1**	**IF 4**	**IF 1**	**IF 4**	**Total**
1	4.64	1.32	5.96	0.66	**12.58**
2	0	0	0	0	**0**
3	0	0	0	0	**0**
4	0	0	0	0	**0**
5	3.73	0.62	3.73	1.86	**9.94**
6	1.3	0.65	0	0	**1.95**
7	0.6	0.6	2.98	0	**4.17**
8	0	0	0	0.58	**0.58**
9	0	0	0.6	0	**0.6**
10	0	0.62	0	0	**0.62**
11	2.94	4.12	7.65	1.76	**16.47**

IF 1: Interface configuration 1
IF 4: Interface configuration 4

The tasks having a high percentage of users who selected both menus and icons correspond to the tasks in Table 8.6 and 8.7 that indicated a comparable selection rate for menus and icons. Tasks 1, 5, 6, 7 and 11 have the highest percentage of users who selected both a menu option and an icon to complete the task. Task 6 has a relatively low percentage of users who selected both a menu and an icon but it is the task with the highest selection rate for menus only. This task required users to close a document without saving the changes made to the document. The selection rates for this task seem to indicate that many users resorted to use of a menu to complete a task and did not even attempt to select an icon once they realised that they could not identify the correct icon (Table 8.7). Tasks 1 and 11 have the highest selection rate for both menus and icons which seems to indicate that users may have attempted to first select the correct icon before eventually using the menu to locate the correct option.

Chart 8.2 puts the selection rate in perspective by clearly showing the comparison between the total selection rate of icons only, menus only and those users who selected both a menu and an icon while completing the task. The bottom stack indicates the percentage of users who selected only a menu option, the middle stack those who selected only icons and the top stack is the percentage of users who selected both an icon and a menu option while completing the tasks.

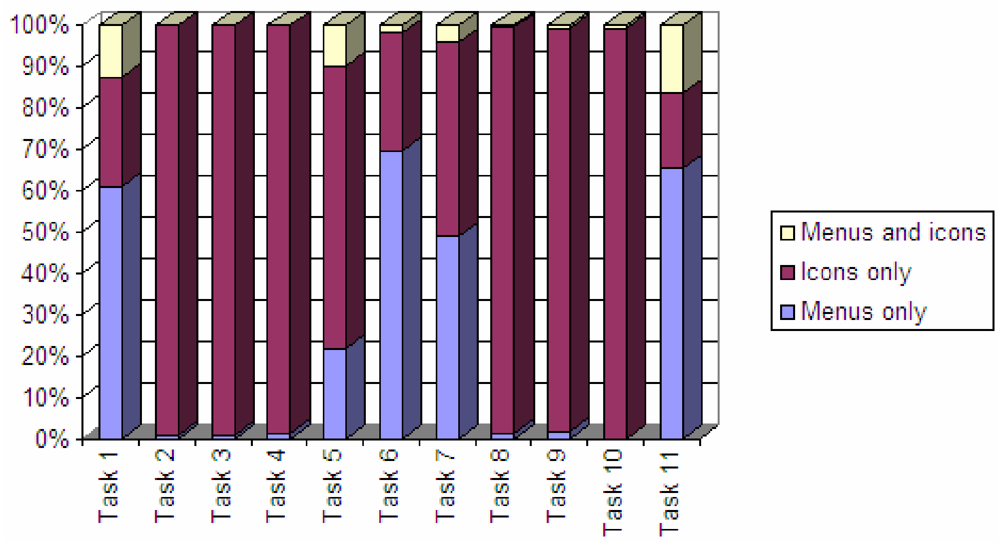


Chart 8.2: Selection rate of icons, menus and both

Although the percentage of users who selected both a menu and an icon during completion of the task only vaguely substantiates the assertion that menus are only used when the user fails to identify the correct icon, this does not entirely discount the theory either. Considering the following five facts concerning tasks 1, 6, 7 and 11:

1. All other tasks have a majority selection for the icons and the majority of users completed the tasks by selecting only icons.
2. Apart from task 5, these tasks (1, 6, 7 and 11) have selections which are evenly split between iconic and menu selections and even task 5 has meaningful representation of both iconic and menu selections.
3. The percentage of users who selected both menus and icons during completion of the tasks is highest for these five tasks.
4. Four out of five of the tasks with a menu selection rate of over 10%, appear in the bottom 5 of successfully completed tasks for all users (Section 7.4.2.3.8, Table 7.17). Of the four tasks that have a menu selection percentage of over 50%, three of the tasks feature in the five tasks having the lowest success rate for all the users (Section 7.4.2.3.8, Table 7.17). Task 5 is the task with the lowest success rate and has the fifth highest selection rate for menu options. These five tasks were also the five tasks with the lowest success rate for the users included in the section's analysis (Table 8.5).
5. Both the users who could complete the tasks correctly and those who could not struggled to differentiate between the different concepts of **Open**, **Close**, **Save**

and **New**. This was evident by the errors made during completion of these tasks.

Taking the five above-mentioned points into account, it would be fair to assume that users attempted to complete the task by first trying to select the correct icon and then, when failing to do so, completed the task by means of using the menu. Simply because the percentage of users who selected both a menu and an icon is not high does not mean that the reasoning behind the high menu selection percentage is an incorrect assessment. The fact that the percentage of users who selected both menu items and icons is low but still noticeable for these five tasks indicates that the users may have followed one of two different strategies to complete the task. First, they may have attempted to identify the correct icon without actually selecting an icon. When they were unable to do this they resorted to selection from a menu. Secondly, they may have attempted to identify the correct icon and even selected a few icons that they felt might be the appropriate icon. Eventually, when they were unable to identify and use the correct icon they made use of the menu instead. The former strategy can be considered to be a mental process of elimination method, while the latter is more of a combination of a physical and mental process of elimination.

8.4 Interaction preference

Table 8.9 shows the average number of menu and icon selections per task for each interface configuration and for each user expertise level.

Table 8.9: Average menu and icon selections

Menu selection					***Icon selection***				
First-time users		**Expert users**			**First-time users**		**Expert users**		
IF 1	**IF 4**	**IF 1**	**IF 4**	**Total**	**IF 1**	**IF 4**	**IF 1**	**IF 4**	**Total**
21.55	20.09	22.64	21.73	**86.00**	76.45	44.27	49.91	38.64	**209.27**

IF 1: Interface configuration 1

IF 4: Interface configuration 4

The average selection rate of menu options was fairly comparable between the different expertise levels and different interface configurations (Table 8.9). The average icon selection rate was higher than the average menu selections for all interface configurations but still comparable between the different expertise levels, with the exception of first-time users of interface 1 (Table 8.9). These users had a higher average selection rate for icons, quite possibly because this group was more prone to using the afore-mentioned physical process of elimination than the other groups. On average, a menu option was chosen by the users 86 times during completion of a task, whilst an icon was chosen, on average, 209.27 times during completion of a task. This shows that given the choice between using a menu or an icon, the majority of the users chose an icon. This observation supports the findings of Lane et al. (2005a), who found that users prefer employing the use of icons during their interaction with a graphical system, even to the exclusion of more efficient methods, such as keyboard shortcuts. Bickford (1997) made a similar assertion that, when given the choice between using a menu or an icon, users would tend to opt for using the icon. This finding indicates that the use of icons is not unwarranted and that they serve their purpose by providing shortcuts to the most frequently used functions within a word processor. Providing the users with shortcuts would not be worthwhile if the users continued to make use of other interaction methods which were possibly more inefficient. Whether menus or icons are the more efficient method of interaction has not been established and this remains an area of possible further research.

8.5 Summary

Overall, the error rates of the tasks were relatively low, with most tasks showing very few errors. However, the discoveries made during analysis of the errors reinforced the prior findings that there appears to be some misunderstanding between the different actions which can be carried out on a whole document. The **New**, **Open**, **Close** and **Save** commands often made up the bulk of the errors for other document handling tasks. As mentioned previously, the confusion caused by these commands may be due to their abstract nature – there is no physical document which can be manipulated. A second possibility is the interchangeable nature of the words within an office setting and the fact that not all words are commonly used within the metaphor used by a word processor (Section 9.3.2.1).

The high average selection rate of icons indicates that users prefer using icons when attempting to complete a task as opposed to selecting the required options from a menu structure. Furthermore, users only resorted to menu selection in the cases where the correct icon could not be ascertained. These assertions were substantiated when investigating the percentage of users who selected only menus, those who selected only icons and those who selected both menus and icons while completing the tasks. This validates the need for icons to be provided as shortcuts for commonly used functions but whether the icons provide the most efficient method of interaction for the users is an area for further research.

The next chapter will summarise and provide conclusions for the results as discussed in the previous three chapters, and will attempt to suggest an interface that will ensure high effectiveness and efficiency during interaction.

Chapter 9

Interpretation of results

9.1 Introduction

Once the user testing has been completed and the collected data analysed, the results must be interpreted and compared with previous studies and available literature. Conclusions based on the results must be drawn which contribute to the field of study and provide vindication for the elected research endeavour. The previous chapters discussed the pilot study conducted with the developed test instrument as well as the results of the user tests that were conducted with novice, first-time and expert users. The errors made during completion of the task and the selection rate of the different available interaction methods were compared. These chapters contained limited interpretation of the data analysis. This chapter will provide a more comprehensive and all-inclusive discussion on the results of all the user testing. The chapter will culminate with the recommendation of an interface which could possibly serve to increase usability of a word processor application. The recommendation will be based on the results of the individual analysis of each task.

9.2 Results

Results obtained from the user testing included results of the icon preference, the computer anxiety and computer attitude questionnaires. The objective measurements of user performance will also be discussed in terms of the interface components that were tested. Results gathered from the formative evaluation during the pilot study will also be discussed in a later section.

9.2.1 Icon preference

The icon preference questionnaire distributed amongst the school-going users showed the highest preference for the standard icons, followed by the text buttons, with the alternative icons as the least preferred of the icons. In particular, the **Bold** and **Underline** standard icons were a popular first choice amongst these users, with each of these icons getting a 65% selection as the first choice. The **Save** and **Italic** standard icons were also a popular first choice, getting 63% and 58% of the first choice respectively.

The icon preference questionnaire failed to confirm the findings of previous icon preference studies (Teklebrhan and Blignaut, 2005) or the findings of the first brainstorming session from which many of the alternative icons were drawn. The alternative icons for **Open**, **Close**, **Save**, **Cut**, **Copy** and **Paste** were all obtained from these icon preference studies. In the response to the questionnaire distributed to the participants of the pilot study the alternative **Open**, **Close**, **Save**, **Copy** and **Paste** icons were all the least preferred icon of the three, while the alternative **Cut** was the second least preferred icon. For all of these functions, the standard icon was the most preferred amongst the respondents.

The questionnaire format did differ between the first brainstorming session and this study which could be the reason for the difference in icon preference. A more plausible reason for the perceived difference in icon preference is the different respondent demographic. Respondents of the first brainstorming session questionnaire were all non-computer literate, while respondents to the questionnaire in the pilot study had at least some knowledge of the task domain and of the interface domain, ranging from limited, or none, to more advanced. Respondents who answered the questionnaire with a patterned response were excluded from the analysis, so it could be assumed that the prior exposure to the interface domain did not significantly affect the questionnaire responses. From this it could be accepted that the more task-knowledgeable users could simply relate better to the concepts and as such made the preference choices using a different strategy to the non-computer literate users.

9.2.2 Computer anxiety

The computer anxiety of the task-knowledgeable users was measured using the Computer Anxiety Scale (Marcoulides, 1989) for the independent variables of word processor experience and interface group.

Word processor experience was found to have influenced the computer anxiety of the users included in the study and there was a significant difference between the users of the different interface groups for both general and overall anxiety. However, due to the presence of interaction effects between the expertise and the interface, the differences could be attributed to certain users within the different categories. Expert users of interface 5 had significantly lower anxiety than experts of interface 2, first-time users of interface 7 and first-time users of interface 6.

Computer anxiety was very loosely related to word processor expertise within certain interface groups. It was previously theorised that increased experience would increase the comfort level of the user when using a computer and therefore reduce the fear of the unexpected – thereby decreasing the anxiety of the user. Whether word processor experience would influence the computer anxiety could not be determined due to the lack of a fixed relationship between word processor and computer expertise. Word processor expertise does not necessarily reflect overall computer expertise and a user may have high computer expertise but relatively low word processor expertise. Therefore, these results could not really be compared to previous studies which provided evidence of a relationship between computer experience and computer anxiety (for example, Glass and Knight, 1988; Joncour, Sinclair and Bailey, 1992; Szajna, 1994).

Depending on the anxiety levels of the user, performance can be either adversely or positively affected by the anxiety of the user (Eysenck and Calvo, 1992; Rauh and Seccia, 2006). Since very few differences existed between the anxiety levels of the users, any differences found between the performances of the different expertise levels could be attributed solely to their experience and not to any latent effect of the anxiety present during interaction with the application. Similarly, since no significant difference was found between users of the different interfaces, it could be concluded that anxiety could not have unduly affected the performance of a single interface group. Therefore, any performance difference between the users of the different interface groups could solely be attributed to the effect of the interface on the usability of the product.

9.2.3 Computer attitude

The computer attitude of the task-knowledgeable users was measured using the Computer Attitude Scale originally developed by Loyd and Gressard (1984b) and later refined by Bandalos and Benson (1990). Word processor experience was found to influence the computer attitude of the users, with more experienced users having a more positive attitude towards computers. It would be expected that users with a high degree of word processor experience would also have fairly high computer experience, but users with low word processor experience do not necessarily have low computer experience. Conflicting results have been found concerning the effect of computer experience on computer attitude. The findings of this study support the theory that computer attitude is heavily influenced by computer experience (Busch, 1995; Loyd and Gressard, 1984a; Orr, Allen and Poindexter, 2002), since it could be assumed that participants with high word processor expertise necessarily have high computer expertise. This relationship was expected due to the influence of anxiety on attitude and the correlation between anxiety and computer experience. Performance can be affected by the attitude of the user towards computers (Galitz, 2002). Since no particular interface group had a significantly different attitude towards computers there should have been no interplay between the attitude of the computer user and the interface. Therefore, performance of a single interface group should not have been unduly affected by the attitude of the user and performance of the different interface groups could only be affected by the interface.

9.2.4 Word processor expertise

As would be expected the expert users were able to complete the tasks much faster. For some tasks experts were also able to complete the tasks with significantly fewer actions and errors than the first-time users. Therefore, first-time users took longer to identify the correct icon or menu option for use, but did not necessarily select any more icons or menu options than expert users did.

Word processor expertise was directly linked to the performance of the user, which substantiated the need to differentiate between users based on their level of experience with the application. It also indicated that expert users were able to move seamlessly between interfaces which made use of different icons. Experts also expressed a significantly higher level of satisfaction with the application than the first-time users.

9.2.5 Translation

It was expected that expert and first-time users would not be unduly affected by translation of the interface into their first language as they already had some degree of experience with the task domain. Possibly, the first-time users would gain some assistance from translation but expert users should be comfortable within the application domain in the language that was available and no effect should be experienced by translation of the interface. It was expected that novice users would benefit from translation of the interface due to their limited knowledge of both the task and interface domain. However, this study found no influencing effect of language on any of the expertise levels. These findings support the previous findings of studies conducted with bilingual users using both websites and software applications where it was found that translation was not necessary (De Wet et al., 2002; Stander, 1997) and that performance was not affected by translation into the first language of a bilingual user (McDonald and Blignaut, 2005; Stander, 1997). Performance was certainly also not impeded by translation into the user's L1 as has been prone to happen when websites are translated (De Wet et al., 2002). Since many of the Sesotho speaking users were unable to understand the terminology and phrasing used on the translated website, it is reasonable that their performance would be somewhat hampered by the translation. The current study did not have large amounts of text that had to be translated. Instead the text buttons mainly consisted of a single word and more verbal context was provided by the tooltips. Therefore, the vast differences between the styles of interaction could easily explain the contradictory findings between this study and that of De Wet et al. (2002).

Translation for novice users did not, as expected, ease the learning time required by the users nor did it increase their performance after interaction with and exposure to the application.

The task-knowledgeable group was more fluent and proficient in English, thus possibly accounting for the lack of a significant difference between L1 and L2 users.

However, since these results were emulated by the group of novice users, who could possibly be considered to be less fluent in English, it could be concluded that language did not influence the usability of an application for bilingual users. First language users did, however, display a significantly lower overall satisfaction than the L2 users. This could be attributed to the fact that they were task-knowledgeable users and therefore, may have become accustomed to using the application in English. The final finding of this study is that there is no need for translation into the L1 of a bilingual user.

9.2.6 Icons

No expectations or predictions could be made as to which pictorial icon set would provide better usability for the users. About half of the icons were chosen as the most preferred icon by a group of non-computer literate users and the remainder of the icons were designed to provide context and easy interpretation of the icons. These icons should then have been able to facilitate easy use and interpretation, but whether they were actually better than the standard icons could not be predicted. The findings of this study indicated that neither of the pictorial icon sets significantly influenced the overall performance of any of the user groups. Novice users performed equally well on both icon sets and performances of the more task-knowledgeable users were not really influenced by the use of the different pictorial icon sets.

Individual findings for each task did, however, provide evidence of some design traits which should be kept in mind when designing icons. The original idea for the **Bold**, **Italic** and **Underline** was to develop similar surrounding icons (Barr et al., 2002b; Piamonte, 2000) to provide more context and easier visualisation and subsequently interpretation of the icons. Therefore, these icons were each designed using the same single letter with the font style applied to the letter to signify the respective font style effect. These were, however, not as successful as the combination of the verbal and visual cues provided by the standard **Bold**, **Italic** and **Underline** icons. This finding provides proof that the best icons are those that resemble the underlying function (Ferreira et al., 2006). Both the standard and alternative icons resembled the underlying function, but only the standard icons significantly increased performance, therefore there was obviously some other aspect of the standard icon which provided increased usability. Ellis et al. (1995) attributed the increased intuitiveness of these icons to the combination of a visual representation and a verbal cue in a single icon. This previous study (Ellis et al., 1995) supported the superiority of the **Bold**, **Italic** and **Underline** icons, while the current study resulted in the standard **Bold** and **Italic** icons emerging as the better icons. By extension it could be assumed that the **Underline** icon would function equally as well as the **Bold** and **Italic** were it not for the presence of the **Font colour** icon which was easily confused with the **Underline** icon due to its high degree of similarity. Therefore, this study confirms the findings of Ellis et al. (1995) regarding the superiority of the combination of visual and verbal cues into a single icon. The standard icons for **Bold**, **Underline** and **Italic** were also the most preferred icons of the three sets. In this case, the performance reflected preference, with the more preferred icons delivering a better performance.

Icons which were not intuitive to users who received no training on the interface did not necessarily decrease the learnability of the interface. The alternative **Close** icon was proof of this assertion. This icon was unsuccessful when tested with the more task-knowledgeable users unless presented in an interface which included tooltips. However, this icon did not significantly decrease the usability of the interface for the novice users, thereby indicating that even icons which were not at first glance intuitive could be learnt and used without hampering the effectiveness and efficiency of the users. Previous studies have also found that an icon is accepted regardless of the depiction used and once the function behind the icon has been explained to the users (Heukelman, 2006).

Use of text buttons could either enhance or impede the use of the interface. Usability could be negatively impacted since images are easier and faster to process, store and recall than text. However, use of single function words promoted recognition above recall since all available functions were simultaneously visible to the user and it eliminated the need to interpret a picture to extract meaning. Neither of these possibilities materialised since there was not enough superiority of either icon type over the other to indicate that either is exclusively the better type of icon type to be used.

The novice users showed no increased performance when using the text buttons as opposed to the interface with pictorial icons on which they received their training. This shows that the users were able to adapt easily to changes in the interface once they had learnt the basic usage and terminology of the application and that the text buttons provided no extra usability to the interface. Neither did the text buttons detract from the usability of the interface since performance was also not degraded by switching to the text buttons. These results support the findings of previous studies which found no difference in the usability provided by the use of pictorial and textual icons (Benbasat and Todd, 1993; Rogers, 1989b; Zammit, 2000).

There was no significant difference between the satisfaction ratings of the users of the different interface groups.

There was not enough evidence amongst the users of the different interface groups to provide unequivocal support for the recommendation that the standard icons be replaced by either the alternative icons or the text buttons. However, a number of discoveries were made about individual and groups of icons which could be used to recommend a combination of icons which could increase the usability of a word processor. This recommended interface will be discussed in a following section.

9.2.7 Menus and tooltips

Menus only appeared to be used when the correct icon could not be identified and thus, they provided a valuable contribution to the overall usability of the interface, although this contribution did not significantly impact the usability of the interface.

In contrast to the findings of Zammit (2000), tooltips were found to increase the ability of the users to interpret the meanings of the icons. Icons which were otherwise difficult to identify when used in an interface with no tooltips, presented no significant problem to users of interfaces with the same icons but where tooltips were included in the interface. Zammit’s (2000) study focused on young children who may not have realised that tooltips were available for use, providing a possible reason for the differing results of the two studies.

As previously mentioned, the language of the menu and the tooltips did not affect the usability of the interface.

9.3 Recommendations

From the results found in Chapters 6 and 7 and the qualitative observations discussed in Chapters 5 and 6, a number of recommendations could be made to improve the interface of current word processing applications.

9.3.1 Icons

The high incidence of icon use allows it to be surmised that icons are the most popular form of interaction (Chapter 8), a finding which supports previous studies (Lane et al., 2005a) and available literature (Bickford, 1997). Icons therefore serve their purpose of providing shortcuts for most commonly used functions. Therefore, it is imperative that icons still be used in applications. Whether icons are the most efficient form of interaction is still unknown and further investigation is needed into whether users eventually switch to keyboard shortcuts as these are the most efficient method of interaction.

9.3.1.1 New

No significant relationships between the interfaces were found for the new task, but closer inspection of the average completion time, the average number of actions required and the mean number of errors made during completion of the task would seem to indicate that the users of the interfaces with alternative icons had the best performance. Therefore, the recommended interface should use the alternative New icon.

9.3.1.2 Open and close

The users of interface 4, which used alternative icons, were able to complete the open task in the fastest average time. Users of interface 4 also had the third lowest average number of actions and the least number of errors made during completion of the task. Therefore, the alternative icon is a prime candidate for the **Open** command.

The text button could possibly be the most successful icon for closing a document as these users were able to complete the tasks the fastest, although when taking into account the number of actions required and the number of errors made, users of the standard icons appeared to have performed better on this task. Therefore, the standard icon for **Close** is recommended.

Open and **Close** are closely related to each other and as such the two icons should form a single group in a manner that easily conveys to the user that the two actions are related. The alternative **Open** icon and the standard **Close** icon do not really manage to communicate that they are related to each other in the sense that the actions are opposites of each other, so either they must both be the alternative icons or both the standard icons. Although the alternative **Open** and **Close** did not really conform to the guideline of ensuring that related functions are conveyed as such, they were both chosen as the preferred icons by non-computer literate users and as such they could be included as a set in an interface.

However, the alternative icon for **Close** could not be considered to be a successful icon since users were unable to interpret the meaning of the icon without the added assistance of a tooltip and therefore it was certainly not an appropriate icon for the **Close** function (Hersh, as cited in Lodding, 1983) as it did not successfully represent the function in a recognisable way (Apple, 2006; Shneiderman, 1998). Users of the alternative icon with neither a menu nor tooltips took the longest on average to complete the task, required the most actions and made the most errors. There was also often a significant difference between the performance of these users and users of other interfaces. Therefore, the alternative icon for **Close** cannot even be considered for inclusion into the recommended interface as a good icon should be intuitive and easy to interpret. This means that the standard icons must be used for both the **Open** and **Close** functions.

9.3.1.3 Save

Inspection of the mean time required, the mean number of actions required and the mean number of errors made, together indicate that the alternative icon users fared the best when having to **Save** a document, followed by the standard icon users and then the text button users. The alternative icon for **Save** is, therefore, proposed for use in the recommended interface.

9.3.1.4 Cut, copy, paste and undo

Only the novice users were given a single task where they were required to **Cut** and **Paste**. Neither **Copy** nor **Undo** was tested with any of the users. Qualitative observations did uncover some problems with the copying of text but these had nothing to do with the icon itself and will be discussed in a following section.

Neither the standard nor the alternative icon significantly increased the usability of the **Cut** and **Paste** icons for the novice users, therefore it is immaterial which icon is used from an objective measurement point of view. Using available guidelines to assess which of the icons is the better one, it is recommended that the alternative **Cut** icon be used since it provides more context in the form that something is being **Cut** from the document. **Paste** presents somewhat of a problem since the standard icon would work nicely provided that the users were aware of the clipboard metaphor, which many of them weren't. However, should the changes be made as suggested below, where the clipboard becomes a more prominent form of action communication to the user, then the standard icon would fare nicely as a **Paste** icon. The standard **Copy** icon is recommended since it clearly indicates that there is an exact duplicate of something being made. The alternative **Cut** and standard **Copy** and **Paste** still manage to convey the fact they are a closely related group of functions.

Although at first glance the alternative icon appears to be the better icon for **Undo** since it depicts an action of erasing, it is recommended that the standard icon be used for the **Undo** function. This recommendation is based on the findings of a previous South African study which found that the two most popular icons for **Undo** were those that incorporated an arrow in some way. This was further substantiated when respondents were asked to depict the action of **Undo** and 40% of them drew some form of arrow (Heukelman, 2006). Unfortunately though, these users had had some exposure to Microsoft Office products (Heukelman, 2006) and this could have clouded their creative development and choice of icon. Nevertheless, it is still recommended that the standard icon for **Undo** be kept.

9.3.1.5 Font type and font size

As will be discussed in an upcoming section, the drop-down lists were not found to be suitable for use by the less experienced users and therefore, it is recommended that the **Font name** and **Font size** drop-downs on the toolbar be replaced with text buttons. Activation of the buttons will then result in a dialog box being displayed.

9.3.1.6 Font colour

The standard icons for **Font colour** and **Underline** (Section 7.4.2.3.7) did not conform to the guideline that each icon should be distinct from every other icon (Apple, 2006; Hersh as cited in Lodding, 1983; Shneiderman, 1998). Therefore, one of the icons should be redesigned or adjusted. Since the **Underline** icon has been shown to be one of the most successful icons of the standard icon set (Ellis et al., 1995) and it should function equally as well as the **Bold** and **Italic** icons functioned during this study, it is proposed that the **Font colour** icon should be adapted to ensure that it is no longer similar to that of **Underline**.

A number of possibilities could be proposed for a new **Font colour** icon.

- Of paramount importance is that the line underneath the letter “A” be removed from the icon as this is the chief reason why the **Font colour** and the **Underline** are so similar to each other. This small change will go a long way in ensuring that the icon is no longer so similar to the **Underline** icon. The colour of the letter “A” could then change according to the colour selected for the text. However, the letter “A” has no associated meaning, thereby not providing an additional verbal cue and therefore this icon was not recommended for use.
- The letter “F” (for font colour) would perhaps be more appropriate than the letter “A”, and then the colour of the letter could change according to the colour chosen. Although the letter F did not increase the usability of the **Bold**,

Italic and Underline icons, it may succeed in providing that required verbal cue of the letter F for Font colour. The fact that the colour of the letter will also change may also increase the intuitiveness of the icon. It will be similar to the alternative Font colour icon with the exception that the colour of the letter "F" will change according to the selected font colour. The fact that the icon will default to black may detract from its usability, particularly since it is recommended that the icon be moved adjacent to the Font size, which will place the icon directly to the left of the Bold icon. This icon was also not recommended for use.

- Another option is to use the Font colour icon used in the current Wordpad application which makes use of a colour palette lying slightly behind a capital letter "A" (). The problem with this icon is the same as that experienced with the original first iteration of alternative icons – that is, that there is a lot of detail on the icon with the result that the icon may appear cramped and the users might not be able to recognise the meaning behind the icon.

These are all possible alternative pictorial icons each with at least one obvious reason against using the icon. Since the Font colour text button did not result in any erroneous selections and text buttons were recommended for use with the Font type and Font size, it is recommended that the Font colour icon be replaced by the text button for Font colour. The Font colour should then also be moved next to the Font type and Font size on the left side of the toolbar instead of where it is currently placed at the far right of the toolbar.

9.3.1.7 Bold, italic and underline

Findings suggested that the best icons were those that resembled the underlying functions. Although both the standard and alternative font styling icons provide the visualisation of a group of related functions and the representation of the underlying function, the standard icons provided greater usability than the alternative icons. This finding supports the findings of Ellis et al. (1995) who found that the appropriateness of the Bold, Italic and Underline icons stemmed from the fact that they provided not only visual cues as to the font style effect but also verbal cues by employing the use of the first letter of the font style function name. Therefore, it is recommended that the standard icons for **Bold**, **Italic** and **Underline** be used.

9.3.1.8 Alignment

The left and right align icon were often confused due to the fact that the jagged edge seemed to attract the eye of the user resulting in an interpretation of the left align as right align and vice versa. Although the users of the interface 2 (standard icons with tooltips and no menu) had the fastest mean time, the least number of actions and the least number of errors for the right align task, the nature of the errors made during completion of the alignment task dictated that the standard icon should not be used. Although the alternative alignment icons provided more context by the inclusion of a border to signify the edges of a page, they still have the jagged edges which attract the eye of the user. Therefore, the final recommendation for an improved interface is to use the text buttons for left and right align, but to change the caption to **Left Align** and **Right Align** or alternatively to **Align Left** and **Align Right**. This could increase the usability of the text buttons by providing more verbal context. The use of **Align Left** on the text button allows the verb "align" to give an easier visualisation of an action being carried out to move text towards the left hand side of the document window. Therefore, it is recommended that the text buttons for left and right align contain the words **Align Left** and **Align Right** respectively. Tooltips could provide even more verbal context by providing text to the effect of "Align the current paragraph along the left side of the document window". Although no problems were experienced with either pictorial icon for **Centre Align**, to ensure consistency and to enable the alignment icons to be viewed as a family of icons, the **Centre Align** icon must also be changed to the text button.

9.3.1.9 Bullets

Most users seemed to struggle with creating a bulleted list, due more to the fact that they did not know what a bullet was, than because of an unsuccessful icon. The number of actions required and the number of errors made during completion of this task remained fairly consistent among the various interface groups. Users of the interfaces which made use of alternative icons had the lowest mean time to complete the task of creating bullets; therefore it is recommended that the alternative icon for bullets be used.

9.3.2 Menus

It was previously stated that the use of icons should continue as they proved to be the most popular form of interaction. The fact that users appeared to resort to the use of menus when they were unable to locate the correct icon, served as strong recommendation for the inclusion of menus in an interface. With the vast amount of functionality available in single applications, it is obviously impractical to consider providing all functionality via icons and toolbars for applications on the scale of a word processor. Therefore, most applications must in any case use menus to ensure that all functionality is equally accessible. Some smaller applications will be able to easily provide all functionality by means of representation as icons on a toolbar. Nevertheless, evidence does suggest that the inclusion of menus is vital as users employ menus when they are unable to locate and identify the correct icon. Therefore, even applications implemented on a smaller scale with limited functionality must still provide a menu even if it only serves to duplicate the options available on the toolbar. This will offset the negative impact should the icons not serve their purpose of intuitive interpretability.

It is imperative that menus provide descriptive labelling of the available commands which will essentially provide more verbal context, for example, “Open an existing document” as opposed to simply “Open”, or understandable terminology, for example, to reduce the lack of understanding evident between the different document handling functions.

Although it is essential that menus be provided for users, the language of the menus did not affect the usability for bilingual users. It is therefore of no consequence in which language the menu is provided. There are, however, some menu options which could be adapted to increase the usability of the interface.

9.3.2.1 New, open, close and save

The solution to the confusion between the document handling tasks of New, Open, Close and Save could possibly be found in the adaptation of the current menu items for these functions. There are two possible solutions to the problem of users' perceived confusion between New, Open, Close and Save.

The first solution is to provide more verbal context. Verbal context will lighten the load on the user and allow him/her to infer the correct action from the extended wording of the menu item. To provide more verbal context, the following changes could be made:

- New could become Create a new document or Start a new document
- Open could become Open an existing document
- Close could become Close and store the current document or Close and file the current document
- Save could become Save changes to current document.

Verbal context is equally applicable to tooltips and as such the same extended version of the document handling function should be reflected in both the menu options and the tooltips.

The second solution is to amend the existing terminology to reflect the proper terminology used in an office filing environment. The word processing application metaphor of a file documenting system can start to break down when it comes to the terminology of document handling. For example, it is easy to say "***Open*** a ***new*** file for customer ABC" – thereby referring to both open and new to encompass a single action. Remarks like these could obviously create confusion between the document handling actions of Open and New. Similarly, when working with physical files, once a change has been made it is enough to simply close the file and store it; there is no need to first save the changes in some manner before filing the document again. The following changes could then be made:

- New could become Create a new document or Start a new document
- Open could become Retrieve a document or Draw a document (from the file cabinet)

- **Close** could become **File the document** (in file cabinet)
- **Save** could become **Apply the changes to document.**

The **Close** and **Save** functions do not present such a problem since if the user were to close the document without saving, the application would ask whether the changes must be saved before closing. The terminology should still be consistent and should the **Open** and **New** change to different terminology, then so should the **Close** and **Save**.

Both of these solutions are viable although it is possible that provision of more verbal context will negate the need for better and different terminology.

9.3.2.2 Bullets

Users were unsure of what a bullet was, and therefore the menu option for bulleted lists should provide a more descriptive naming of the function. The wording of the menu option, and consequently the tooltip, could possibly be extended to “Create bulleted paragraph list” or “Create bullet list for paragraph”. The menu option could even be extended to “Paragraph bullet and numbered list options”.

9.3.3 Tooltips

Icons are recommended for use due to the popularity of their use during user interaction, therefore they serve their purpose of providing shortcuts to common functions. However, they are not always intuitive and as such the use of tooltips is highly recommended within an application interface. Tooltips are essential for a usable interface since even an icon which has been developed with the best intentions and following all the recommended guidelines can still fail to accurately convey the correct functionality to the user. Provision of tooltips did assist users with interpretation of icons before they had to resort to menu use to find the correct function.

Tooltips should be reworded in the same way as the menus to provide more verbal context or better terminology. In the cases where the toolbar shortcut is not available as an individual menu option, such as with **Bold**, **Italic** and **Underline**, which are only available as options in the **Font** dialog box, the tooltip should be expanded and should provide more than just the function name. For example, the tooltip for **Bold** could be "Make the text bold" instead of just "Bold". It could even supply a clue as to which menu option should be used to obtain the styling effect, such as "Change font style to bold".

Although tooltips are essential to a usable interface, the language of the tooltips can be decided upon for each development project. Of course, it is obviously highly recommended that the language of the menus and the tooltips be consistent with one another.

9.3.4 Recommended interface

In summary, the following recommendations are made concerning the icons (see Table 9.1), menus and tooltips for use by bilingual users.

Both menus and tooltips should always be present in an interface and both should be worded to provide more verbal context. Should provision of more verbal context still fail to dispel the confusion present with the use of the current document handling terminology then consideration should be given to changing the terminology to allow for easier understanding and differentiation of the functions. For bilingual users it does not matter whether the language of the menus and tooltips is in their L1 or L2.

The complete recommended interface is shown in Figure 9.1.

Figure 9.1: Recommended interface

Table 9.1: Recommended icons for each function

	Function	Recommended icon
Standard Toolbar	New	
	Open	
	Close	
	Save	
	Cut	
	Copy	
	Paste	
	Undo	
Formatting toolbar	Font Name	Font Name
	Font Size	Font Size
	Font Colour	Font Colour
	Bold	**B**
	Italic	*I*
	Underline	U
	Left Align	Align Left
	Centre Align	Align Centre
	Right Align	Align Right
	Bullets	

9.3.5 Qualitative observations

Proposed solutions to the qualitative observations made during the study and discussed in Chapters 5 and 6 could also increase the usability of not only a word processor interface, but also other interfaces which exhibit some of the behaviour and use the interface components that were found to create problems for the users.

The problems that were discovered during the pilot study and the user testing with the novice users were:

- Font;
- Font colour;
- Font size;
- Complex dialog boxes;
- Subtle changes;
- No visible feedback.

9.3.5.1 Font

Users were unable to understand the entire concept of font and did not realise that font encapsulated such aspects as font type, size, and colour. Therefore, the recommendation is to subdivide the font menu into three separate menus of font type, font size and font colour. This recommendation complements the recommendation to replace the font type and font size drop-downs with text buttons on the toolbar and the recommendation to split the dialog box into three separate and individual dialog boxes.

9.3.5.2 Font colour

The selection of the text causes a distortion of the true colour. Therefore, it is recommended that the selection be automatically cancelled after the font colour has been changed.

9.3.5.3 Font size

The ruler confuses the users when they see the integer numbers and assume that they correspond to the integer font sizes. By replacing the **Font size** drop-down with a text button that contains the function name, this problem may be eliminated.

9.3.5.4 Complex dialog box

The complex font dialog box caused numerous problems during both the pilot study and the user testing with the novice users. To recap, these problems were:

1. Users were eventually able to find the correct area in the dialog box but had difficulty in understanding that clicking on the down arrow next to the drop-down would cause the drop-down to expand and reveal more options.
2. Clicking on the small arrow created difficulty due to poor mouse dexterity.
3. Once the drop-down had been successfully expanded, users did not understand how to scroll upwards and downwards in order to reveal more available options.
4. Users were overwhelmed by the sheer complexity of the dialog boxes and the number of options and information available in the small area.

These observations led to the formulation of the following suggestions which could be kept in mind and tested at a later stage with other novice users. The suggestion is based on a progressive methodology which will ease users from the simple individual dialog boxes to eventual use of the current dialog box.

1. Shorten the initial lists presented to users, at least until the user becomes more confident and more secure in the workings of the application. So, instead of giving the users a long list of all possible variations of colours to choose from, rather give them a list of 5-10 commonly used colours.
2. Instead of using a drop-down, present these options to the user in the form of a list box. This will remove the problems experienced with the expansion of a drop-down. By using a list box, all the available options will be visible to the user at first glance. A second option would be to follow Nielsen’s (2004) recommendation of using radio buttons instead of drop-down lists since this allows for all options to be visible immediately and simultaneously. Since these users did experience problems with mouse dexterity and positioning the

mouse pointer on a small target, the design would have to allow for selection of the option by either clicking on the button itself or on the corresponding label (Nielsen, 2004). The only restriction to using radio buttons is the lack of scrolling capabilities, but should the initial list be shortened, then space would not be a problem. However, transition from a list box to a drop-down would be easier since the transition could be implemented systematically. The list could initially be shortened, thus eliminating the need for scrolling. Once users had become accustomed to the list boxes, the lists could be lengthened and require users to scroll upwards and downwards to reveal hidden options. Once the user has mastered the action of scrolling, the list can be converted to a drop-down and by then the user should be confident and capable enough to be able to click on the down arrow to expand the list. Therefore, the final recommendation for increasing the usability of dialog boxes is to use list boxes instead of drop-downs.

3. Separate the font dialog box into at least three different and distinct dialog boxes, namely, those of font type, colour and size. This will eliminate the large amount of information presented to the user in a single dialog box and will streamline the selection of the correct dialog box and subsequent selection of the desired font effect.

Following the three steps above would systematically allow users to become accustomed to the dialog boxes without overwhelming them with too much information. Once they had adjusted to the use of the dialog boxes, the lists could be lengthened to provide more options and then eventually the three dialogs could once again be combined into a single comprehensive font dialog box.

9.3.5.5 Subtle changes

Subtle changes that occur when creating a new document or saving a document for the first time go unnoticed by the users. To increase their confidence and to prevent them from repeatedly pressing the same icon in an attempt to elicit a reaction from the system, more obvious feedback must be given. Perhaps when creating a new document the background could be darker than the grey colour used in the test instrument of this study. The change from the dark grey to white would then be more noticeable to the users. Solutions for the lack of any visual feedback for the subsequent saves of a document are discussed in the following section and will also solve the problem of subtle changes to the environment when saving the document for the first time.

9.3.5.6 No visual feedback

Actions such as subsequent saves of an already named document or copying a piece of text have no visual feedback and cause confusion amongst the users who are then unsure of whether the action has been successfully completed or not.

Strictly speaking, subsequent saves of the document do have feedback in the form of a progress bar in the bottom left corner that indicates the save process. Some documents are, however, so small that the progress bar is never really visible. Similarly, a small disk appears in the status bar of the application but once again the disk disappears very quickly when saving small documents and will not be noticed by the user. A possible solution to this is to have an indicator icon in the status bar that indicates whether the document has been modified and must be saved. Once the document has been saved the indicator icon can change to one which indicates that the document has been saved. Since this icon will always be visible to the user, there will no longer be any confusion as to whether the save action has been successfully completed. The spell check uses the same concept by providing an indicator icon which allows users to determine whether there are any spelling errors in the currently opened document (Figure 9.2).

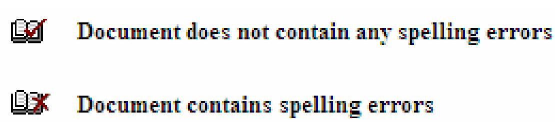


Figure 9.2: Microsoft Office spell check indicator icon

A second option to solve the problem of no visual feedback when saving a document, is to disable the save icon and menu option once the document has been saved and only enable it again when the document is modified.

To counteract the lack of visual feedback with copying of a piece of text, the clipboard tool as currently employed in Microsoft Word can be used but with the difference that it appears immediately as soon as any text is copied. The leading piece of the copied text can also be shown on the clipboard to provide the reminder to the user of what is currently on the clipboard. Another option is to create an indicator icon for the clipboard, similar to that proposed for the save action. The clipboard can then indicate how many copied pieces of text are currently available. By doing this it will be easy to determine whether the text has been successfully copied. Double-clicking the indicator icon can then open the clipboard in a panel alongside the current document window as is currently available in the Office applications.

Another option which is available for both saving and copying is to provide an animation similar to copying files in Windows Explorer. These animations will indicate the process of saving and copying text to a clipboard. The animations should be visible for long enough for the user to see even when saving a small document. The animation will provide more feedback than the progress bar and the appearance of the disk when saving, as it will be shown in the active window and not at the bottom of the screen in the status bar. The movement of the animation will also attract the eye of the user who will then be made aware that the action has been carried out.

9.4 Accommodating all users

As has been previously noted it is virtually impossible for a single interface to be able to cater for the diverse user base of such a complex application as a word processor (Sullivan, 1989). Many of the recommendations suggested here, such as splitting the font dialog box into three separate dialog boxes and immediately displaying the clipboard when text is copied, may assist the novice and first-time users, but will eventually slow down and possibly even cause annoyance for the expert users. Therefore, the final recommendation is that the interface is customisable in the sense that these options can be switched off and that, for instance, the font dialog box can be displayed as a single dialog and not as three separate dialogs. Although the users can gradually be eased into using the single dialog box as mentioned previously, the option must still be available to immediately switch to the currently used dialog box. Similarly, users who are aware of the nature of the copy action need not be inundated with the information on the clipboard or an animation each time they copy a piece of text.

9.5 Summary

A questionnaire distributed to gauge user preference indicated no need for the development of a new interface as the most popular icons were the standard set of icons. The standard icons were followed by the text buttons as the next most popular, and least popular were the alternatively designed icons. Even those alternative icons obtained from previous icon preference questionnaires failed to elicit a high preference from the respondents in this study. A common perception is that user preference and user performance are not always necessarily the same, therefore user testing still had to be conducted to determine which icons, if any, provided the most usable interface.

Results of the subsequent user testing showed no difference between user performance on either the pictorial icon set or the text buttons. Consequently, it could be concluded that the icons did not influence the usability of the interface to a great extent. The best icons appeared to be those that successfully combined a verbal cue and a visual one, such as in the case of the standard Bold, Italic and Underline. Icons which were not intuitive could still be learnt and used with no adverse effects, as demonstrated by the unaffected performance of the novice users with the alternative Close icon. There was no real difference in user performance when using an interface in the L1 or the L2 of the user, thus there is no need for translation of a word processor interface.

The results were not able to recommend the superiority of a single interface over another. However, by considering the tasks separately, the performance of the users could be analysed and an appropriate icon could be recommended for that function. Where possible, the objective usability measures were inspected and the decision was based on the combination of the mean time required by interface users to complete the task, the mean number of actions needed and the number of errors made during completion of the task. If no objective measures were available for that function or if the objective measures failed to pinpoint a single interface icon as the best, then established icon design guidelines were used to evaluate the available icons for that function and the best one was recommended for use. The result was a recommended interface which used a combination of standard and alternative icons and text buttons, while ensuring that groups of icons still formed a cohesive unit and the interface was consistent and usable.

It was recommended that menus and tooltips always be included in any interface but that there was no need to provide these menus and tooltip in any language other than English for bilingual South African users. Menus and tooltips should also be amended to provide more verbal context for the users since tooltips provide assistance when the function behind the icon cannot be identified, and menus are used when the user is unable to locate and use the correct icon. Should more verbal context still not solve the problem associated with the document handling functions, then a change in the terminology used was recommended.

The qualitative observations recorded during the pilot study and the interaction of the novice users with the application provided a means by which the usability of the interface could be improved. Of course, while these adjustments are required, it must still be kept in mind that many of the changes could hamper the interaction of expert users and as such the interface should be as customisable as possible.

The following chapter will summarise and provide final conclusions for the study.

Chapter 10

Conclusion

10.1 Introduction

The aim of this research undertaking was to determine whether there was a need for the development of an interface which could increase the usability of a word processor application. The interface components which were concentrated upon during the research study were icons and the images depicted on them, icons presented in the form of text buttons, and the language of the menus and tooltips of the interface. A set of different interface configurations was developed and user testing was conducted on them to determine whether any of the interfaces presented a more effective, efficient and satisfying interaction with the word processor. The previous chapters provided the analysis and interpretation of the user testing results. From this analysis and interpretation a recommended interface for use in a word processor application was proposed. This chapter will recap and conclude the research study.

10.2 Aims and motivation

The original aim of this research study was to determine whether there was a need for a word processor application interface to be designed for specific cultures. Attempts to design localised icons for the cultures included in this study proved to be an almost impossible task. Subsequently, the focus of the study shifted to determine whether the current interface used in word processor applications could be improved upon in any way. The establishment of the set of icons found in these applications as the standard icons for use in the majority of applications does not automatically guarantee that the icons are the best icons to depict the functions they represent. Investigation of some of the available literature (Chapter 2) discounted this perception that icons which are accepted as the industry standard can automatically be assumed to be the best icons for use in applications. Therefore, a set of icons obtained from previous studies and brainstorming sessions was included in this study as an alternative set of interface icons (Chapter 4).

The literature (Chapter 2) also indicated that nonverbal and verbal stimuli are processed, stored and recalled with varying proficiency, with the nonverbal, or imagery, stimuli emerging as superior with regard to the processing, storage and recollection of information. The display of text buttons as opposed to pictorial icons should have been just as advantageous due to the promotion of recognition over recall and the fact that users no longer had to interpret the meaning of an image. A number of studies found conflicting results when comparing icons, text and a combination of icons and text. There was no discernable difference in user performance between users of text and users of icons, when performance was measured in terms of time taken to complete a task and the number of errors made. Icons did, however, increase the recall abilities of the users while text increased the understandability of the interface. This study aimed to determine what the effect of text buttons and icons would be on a group of South African users. To this end, a set of text buttons was developed (Chapter 4) which displayed the function name as opposed to a pictorial representation of the function represented by the icon.

Bilingualism is also an important factor in application design, particularly in the South African and wider African context, due to the unique nature of language on the African continent. Accepted models of bilingual representation within the mind of the bilingual allowed for the conclusion to be drawn that second language users were placed at a distinct disadvantage from the outset when using interfaces not in their first language. This evidence has been disputed in South African studies, which in the majority of cases found no difference between the performances of bilingual users. User performance on translated websites was, in some cases, affected by the language of the user, but also by the amount of prior experience on the website. Only a single previous South African study was found which investigated a more traditional software application for use by bilingual users. This study was conducted in a different region from that of the current study, and therefore it was considered to be of paramount importance to ascertain whether the same results would be obtained. Translation of the interface was achieved through translation of the menu and tooltips into Afrikaans and Sesotho. Menus and tooltips were included in some interface configurations.

10.3 Research design

The opportunities presented for further research by the translation of the interface and the set of alternative icons and text buttons formed the foundation of the research focus. Experimentation, in the form of user testing, was identified as the best method of comparison between the different user interface configurations and the user testing procedure followed in this study was designed using stringent guidelines as laid out in the literature for successful user testing (Chapter 3). A word processor application provided a novel and convenient application with which to put all the theories to test. Word processors, or some form of text editor, have become very popular tools in everyday life. Therefore, a word processor application which was capable of limited functionality and which allowed for interchangeable interfaces was developed (Chapter 4) for use during the conduction of the user testing.

The different interface components that were tested were then combined into several different interface configurations. These interface configurations were tested using established user testing methods. First, a pilot study was conducted (Chapter 5), which managed to reveal certain problems with both of the designed icons and other problems with the interface, the solutions to which could lead to a better interface. Secondly, some of the interface configurations were tested with novice computer users (Chapter 6). Thirdly, all the available interface configurations were tested using more task-knowledgeable users, namely the first-time and expert users (Chapter 7). From the results of the user testing it was also possible to study the types of errors made (Chapters 7 and 8) as well as other problems experienced by the users which impeded their interaction with the application.

10.4 User testing

User testing was conducted by means of requiring representative users to complete a number of preset tasks which were identified as representative tasks indicative of common word processor use. The interaction of the users was captured and the usability of each interface configuration was measured using the usability measurements of (i) the score obtained for the test, (ii) the time taken to complete each task, (iii) the number of actions required to complete each task, (iv) the number of errors made by the users during completion of the task and (v) the success rate of each task. User satisfaction was also measured for the larger group of task-knowledgeable users.

By conducting user testing it was possible to test the general hypotheses identified and formulated in Section 1.3. These hypotheses were:

1. $H_{0,1}$: A set of alternatively designed icons does not influence the usability of a word processor.
2. $H_{0,2}$: A set of text buttons does not influence the usability of word processor.
3. $H_{0,3}$: An interface in the first language of the user does not influence the usability of a word processor.

Results of the user testing (Chapters 6 and 7) showed that translation of the interface into the L1 neither significantly increased nor decreased user performance and that no particular interface configuration included in the study was the single most usable interface for word processor applications.

Neither the novice nor the task-knowledgeable users were affected by translation of the interface. The novice users did not learn nor use the interface in their L1 any more easily than the users who were expected to use an English interface. Similarly, the task-knowledgeable users performed equally well on an interface in their L1 and an interface in English. Therefore, there was no cause or evidence to reject $H_{0,3}$ as listed above.

Concerning the icons, no single icon set as a whole could be identified as the best interface for use with a word processor. The text buttons neither hampered nor increased user performance for the task-knowledgeable users and the novice users were able to adapt easily to the text buttons once they become accustomed to the terminology and use of a word processor. Therefore, $H_{0,2}$ could not be rejected. The alternative icons also did not significantly influence the performance of either the novice or the task-knowledgeable users. Novice users were able to learn and use either interface with comparable performance and were even capable of successfully using an icon which could not be identified by the task-knowledgeable users. Therefore, users can learn to associate meaning with an icon even though it may not successfully convey the functionality with which it is associated. The most successful icons were identified as those which combined a visual and a verbal cue as these icons increased the measured usability of the icons. Since there was very little difference in performance between users of the different pictorial icon interfaces, $H_{0,1}$ could not be rejected.

Analysis of the errors made by the users (Chapters 7 and 8) revealed the confusion experienced by the users between the different document handling functions, which could be solved by the inclusion of verbal context or the adaptation of the terminology associated with these functions. Tooltips assisted the users in identifying the correct icon for use to achieve the desired effect and icons also proved to be the most popular form of interaction when users were given the choice between icons and menus (Chapter 8). Menus are still required since users seemed to employ the use of menus when they were unable to identify the correct icon for use (Chapter 8).

The overall conclusion of this study is that both the novice and the task-knowledgeable users performed well on all the interfaces included in their part of the study. Therefore, no single interface could be recommended as the most usable word processor interface. Users performed on a comparable level between the different interfaces and expert users were able to adjust easily to the alternative icons and text buttons without their effectiveness and efficiency suffering. Similarly, novice users were able to seamlessly make the transition from the pictorial icons to text buttons with no significant difference in their performance.

However, by analysing the results of the tasks individually, a recommended icon was identified for each function included in the study (Chapter 9). The recommendation was based on objective usability measurements and where this was not possible the recommendation was based on existing icon and interface design guidelines. It was also concluded that menus and tooltips should always be included in an interface. The language of the menus and tooltips can be either in English or the L1 of a bilingual user as performance is not affected by the language of the interface. Menus and tooltips should provide more verbal context and if that is unsuccessful in alleviating some of the problems experienced by the users then the terminology should be amended to more understandable phrases. Qualitative observations of the users also revealed certain problems, the solutions to which were included in the recommended interface.

10.5 Contribution to the field

This study, as undertaken within the field of HCI, resulted in a number of contributions being made to the field of HCI.

The study has shown that even though processing of L1 and L2 occurs at different levels of competence, there is no need for provision of a software application interface in any language other than English for bilingual South African users. Therefore, it provides further evidence in the debate opposing the need for translation. Secondly, whether the interface incorporates the use of pictorial icons or text buttons has no influence on the usability of the product, providing evidence that the unequal processing of verbal and nonverbal stimuli cannot be extended to include the components of a user interface. Thirdly, results of the study indicated that the depiction of the function can be learned and used without any negative influence on the usability and performance of the user and that the standard icons hold up equally as well as a set of icons designed by following established icon design guidelines.

Finally, although performance of the users was not unduly influenced by either the icon type or the actual depiction of the function, it was possible to evaluate tasks individually and in so doing recommend an interface for possible inclusion in future word processors. This discovery continues the established trend of adapting and changing the interface of a word processor to the ever-changing environment and in so doing, ensuring that the needs of the users are always met. This interface, in this form, could also possibly lead to the improvement of the user experience with a word processor or could simply provide a starting point for the development of a better and more usable word processor application.

10.6 Further research

This study provides many opportunities for further research. Firstly, further research is required to determine whether menus or icons are the more efficient form of interaction. Secondly, confirmation of the manner in which users decide whether to use a menu or an icon is needed. Thirdly, it should be determined whether users eventually migrate from the easier-to-use but less effective, to the more effective methods of interaction, such as keyboard shortcuts. Finally, the suggested interface should be tested to determine whether it does indeed increase the usability of the word processor.

10.7 Summary

The chapter summarised the path followed during this research study. The research was founded and based in existing literature and the research methodology was developed by following a set framework for research design. The experimental methodology was then based on the foundation built by examination of existing literature and using a previously determined research design. User tests were conducted using as many representative users as possible. The results of the user tests were then analysed. It was found that neither language nor a specific single interface significantly contributed to the usability of a word processor interface. Instead, by combining the standard and alternative icons and the text buttons as well as taking into account the recommendations for menus and tooltips, a single improved interface was recommended for possible use in a word processor.

In conclusion, this study was unable to design and test a single interface which provided increased usability for a word processor. It was, however, discovered that there is room for improvement on the current interface. By analysing user response to graphics, in the form of pictorial icons, text in the form of text buttons and language in the form of translation of the interface, this study was able to recommend a possibly improved word processor interface. This final recommended interface could serve as the first step towards the improvement of user interaction with the popular word processing application for South African users.

References

Abdelnour-Nocera, J. (2001). Situating the interface in context: User diversity and cultural intersections. Workshop on Multiple User Interfaces over the Internet: Engineering and Applications Trends, Lille, France.

Abran, A., Suryn, W., Khelifi, A., Rilling, J. and Seffah, A. (2003a). Consolidating the ISO usability models. *Proceedings of 11th International Software Quality Management Conference*, Glasgow, Scotland. Retrieved 4 April 2005 from the World Wide Web: http://www.gelog.etsmtl.ca/publications/pdf/768.pdf.

Abran, A., Khelifi, A., Suryn, W. and Seffah, A. (2003b). Usability meanings and interpretations in ISO standards. *Software Quality Journal*, 11: 325-338.

Adler, M.K. (1977). *Collective and individual bilingualism: A sociolinguistic study*. Hamburg: Buske.

Alonso, D.L., Rose, A., Plaisant, C. and Norman, K. (1998). Viewing personal history records: A comparison of tabular format and graphical presentation using LifeLines. *Behaviour and Information Technology*, 17(5): 249-262.

Allport, G.W. (1953). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology* (pp. 798-844). Massachusetts: Clark University Press.

Al-Qaimari, G. and McRostie, D. (2001). KALDI: A CAUsE tool for supporting testing and analysis of user interaction. In A. Blandford, J. Vanderdonckt and P. Gray (Eds), *People and Computers XV – Interaction Without Frontiers: Joint proceedings of HCI 2001 and IHM 2001* (pp. 153-169).

Anderson, R.C. and Kulhavy, R.W. (1972). Imagery and prose learning. *Journal of Educational Psychology*, 63: 242-243.

Apple. (2006). *Human interface guidelines: The Apple desktop interface*. Retrieved 25 August 2006 from the World Wide Web http://developer.apple.com/documentation/UserExperience/Conceptual/OSXHIGuidelines/OSXHIGuidelines.pdf.

Arend, U., Muthig, K. and Wandmacher, J. (1987). Evidence for global feature superiority in menu selection by icons. *Behaviour and Information Technology*, 6(4): 411-426.

Atwater, L. and Babaria, K. (2001). Controlled experiments. CHARM: Choosing human-computer interaction (HCI) appropriate research methods, University of Maryland. Retrieved 15 October 2005 from the World Wide Web: http://www.otal.umd.edu/hci-rm/cntlexp.html.

Baecker, R.M. and Buxton, W.A.S. (Eds). (1987). *Readings in human-computer interaction: A multidisciplinary approach*. California: Morgan Kaufan.

Bandalos, D. and Benson, J. (1990). Testing the factor structure invariance of a computer attitude scale over two grouping conditions. *Educational and Psychological Measurement*, 50(1): 49-60.

Barker, P.G., Najah, M. and Manji, K.A. (1987). Pictorial communication with computers. *International Journal of Man-Machine Studies*, 27: 315-336.

Barr, P., Biddle R. and Noble, J. (2002a). A taxonomy of user-interface metaphors. *SIGCHI New Zealand (SIGCHINZ) Workshop on Human-Computer Interaction*.

Barr, P., Noble, J. and Biddle, R. (2002b). Icons R Icons: User interface icons, metaphor and metonymy. Technical report CS-TR-02/20, Victoria University of Wellington.

Batibo, H.M. (2005). *Language decline and death in Africa: Causes, consequences and challenges*. Clevedon, Buffalo: Multilingual Matters.

Battle, L. and Degler, D. (2001). Around the world in 80 clicks. *Performance Improvement (EPSS Special Edition)*. ISPI, 40(7).

Benbasat, I. and Todd, P. (1993). An experimental investigation of interface design alternatives: Icon vs. text and direct manipulation vs. menus. *International Journal of Man-Machine Studies*, 38: 369-402.

Berenson, M.L. and Levine, D.M. (1979). *Basic business statistics: Concepts and applications*. New Jersey: Prentice Hall.

Bevan, N. and Macleod, M. (1994). Usability measurement in context. *Behaviour and Information Technology*, 13: 132-145.

Bickford, P. (1997). *Interface design: The art of developing easy-to-use software*. United States of America: Academic Press.

Blattner, M.M., Sumikawa, D.A. and Greenberg, R.M. (1989). Earcons and icons: Their structure and common design principles. *Human-computer interaction*, 4: 11-44.

Blignaut, P.J. (1999). Usability of a computerised patient record system in a busy township primary health care clinic: A feasibility study. PhD Thesis. University of the Free State, Republic of South Africa.

Blignaut, P.J. and McDonald, T. (2006). Die implikasies van lees- en skryfvoorkeurtaal met betrekking tot Internettoegang vir 'n veeltalige Suid-Afrika. *S.A. Tydskrif vir Natuurwetenskap en Tegnologie*, September 2006.

Bodley, G.J.H. (1993). Design of computer user interfaces for Third World users. M.Com. Dissertation. University of Port Elizabeth, Republic of South Africa.

Bohmann, K. (2000). User performance metrics. Retrieved 19 October 2005 from the World Wide Web: http://bohmann.dk/articles/user_performance_metrics.html.

Brinck, T., Gergle, D. and Wood, S.D. (2002). *Usability for the web*: *Designing web sites that work*. United States of America: Academic Press.

Brosnan, M.J. (1998). The impact of computer anxiety and self-efficacy upon performance. *Journal of Computer Assisted Learning*, 14: 223-234.

Busch, T. (1995). Attitudes towards computers. *Journal of Educational Computing Research*, 12(2): 147-158.

Caron, J.P., Jamieson, D.G. and Dewar, R.E. (1980). Evaluating pictographs using semantic differential and classification techniques. *Ergonomics*, 23(2): 137-146.

Cato, J. (2001). *User-centered web design*. Great Britain: Addison-Wesley.

Chen, H. C., and Ng, N. L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory and Cognition*, 17: 454–462.

Chi, M., Glaser, H.R. and Farr, M.J. (1988). *The nature of expertise*. Lawrence Erlbaum Associates.

Chin, J.P., Diehl, A and Norman, K.L. (1988). Development of a tool measuring user satisfaction of the human-computer interface. *Proceedings CHI '88 Conference Proceedings: Human Factors in Computing Systems*. pp. 213-218.

Choong, Y-Y. and Salvendy, G. (1998). Design of icons for use by Chinese in mainland China. *Interacting with Computers*, 9: 417-430.

Collins English Dictionary (4^{th} edition updated). (2000). England.

Collins, R. (2001). Software localization: Issues and methods. *The 9^{th} European Conference on Information Systems*, Bled, Slovenia.

Cox, K. and Walker, D. (1993). *User-interface design* (2^{nd} edition). Singapore: Prentice Hall.

Cyr, D. and Trevor-Smith, H. (2004). Localization of web design: An empirical comparison of German, Japanese, and U.S. website characteristics. *Journal of the American Society for Information Science and Technology*, 55(13): 1-10.

De Carolis, B., De Rosis, F. and Errore, S. (1995). A user-adapted iconic language for the medical domain. *International Journal of Human-Computer Studies*, 43: 561-577.

De Saussure, F. 1985. The linguistic sign. In R.E. Innis (Ed), *Semiotics: An introductory anthology* (pp. 24-46). Bloomington: Indiana University Press.

De Wet, L., Blignaut, P. and Burger, A. (2002). Comprehension and usability variances among multicultural web users in South Africa. *Proceedings of CHI 2002*, Minneapolis, Minnesota, United State of America.

Digital census atlas. (2001). Retrieved 4 May 2005 from the World Wide Web: www.statssa.gov.za.

Dillon, A. (2001). The evaluation of software usability. In W. Karwowski (Ed), *Encyclopedia of Human Factors and Ergonomics*. London: Taylor and Francis.

Dillon, A. and Song, M. (1997). An empirical comparison of the usability for novice and expert searchers of a textual and a graphic interface to an art-resource database. *Journal of Digital Information*, 1(1).

Dix, A., Finlay, J., Abowd, G. and Beale, R. (1993). *Human-computer interaction*. New Jersey: Prentice-Hall.

Dowdy, S. and Weardon, S. (1983). *Statistics for research*. New York: John Wiley and Sons.

Dufour, R. and Kroll, J. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory and Cognition*, 23: 166-180.

Dukes, R.L., Discenza, R. and Couger, J.D. (1989). Convergent validity of four computer anxiety scales. *Educational and Psychological Measurement*, 49: 195-203.

Dumas, J.S. and Redish, J.C. (1999). *A practical guide to usability testing*. Great Britain: Cromwell Press.

Duncan, V. and Fichter, D.M. (2004). What words and where? Applying usability testing techniques to name a new live reference service. *Journal of the Medical Library Association*, 92(2): 218-225.

Durgunoğlu, A.Y. and Roediger, H.L. (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language*, 26: 377-391.

Egido, C. and Patterson, J. (1988). Pictures and category labels as navigational aids for catalog browsing. *Proceeds of CHI '88: Human factors in computing systems*, pp. 127-132.

Eisenberg, D. (1992). Word Processing (History of). In *Encyclopedia of Library and Information Science* (Vol. 49, pp. 268-278). New York: Dekker.

Electronic statistics textbook. (2005). (Electronic Version): StatSoft, Inc. Tulsa, OK: StatSoft. Available on the World Wide Web: http://www.statsoft.com/textbook/stathome.html.

Ellis, J., Tran, C., Roo, J. and Shneiderman, B. (1995). *Buttons vs. menus: An exploratory study of pull-down menu selection as compared to button bars*. Technical report CAR-TR-764/CS-TR-3452, University of Maryland.

Esselink, B. (1998). *A practical guide to software localization*. Amsterdam: John Benjamins Publishing Co.

Evers, V. (2001). Cultural aspects of user interface understanding: An empirical evaluation of an e-learning website by international user groups. PhD Thesis. University of Amsterdam.

Eysenck, M.W. and Calvo, M.G. (1992). Anxiety and performance: The Processing Efficiency Theory. *Cognition and Emotion*, 6: 55-80.

Faulkner, C. (1998). *The essence of human-computer interaction*. Great Britain: Prentice-Hall.

Faulkner, X. (2000). *Usability engineering*. London: Macmillan Press.

Ferranti, M. (1999). From global to local. *Infoworld*, 21(41): 36-37.

Ferreira, J., Noble, J. and Biddle, R. (2006). A case for iconic icons. *Proceedings of the 7th Australian User Interface Conference*, Hobart, Australia.

Finch, H. (2005). Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology*, 1(1): 27-38.

Fitrianie, S. and Rothkrantz, L.J.M. (2005). An icon-based communication tool on a PDA. *Euromedia*.

Fowler, S.L. and Stanwick, V.R. (1995). *The GUI style guide*. United States of America: Academic Press.

French, R.M. and Jacquet, M. (2004). Understanding bilingual memory: Models and data. *TRENDS in Cognitive Science*, 8(2): 87-93.

Galitz, W.O. (2002). *The essential guide to user interface design: An introduction to GUI design principles and techniques* (2nd Edition). New York: Wiley.

Gambrell, L.B. (1982). Induced mental imagery and the text prediction performance of first and third graders. In J.A. Niles and L.A. Harris (Eds.), *New inquiries in* *reading research and instruction* (pp. 131-135). Rochester, NY: National Reading Conference.

Gambrell, L.B. and Jawitz, P.B. (1993). Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly*, 28(3): 264-276.

Gardiner, H.W. and Kosmitzki, C. (1998). *Lives across cultures: Cross-cultural human development* (2^{nd} Edition). Boston: Allyn and Bacon.

Gardner, D.G., Discenza, R. and Dukes, R.L. (1993). The measurement of computer attitudes: An empirical comparison of available scales. *Journal of Educational Computing Research*, 9(4): 487-507.

Gardner, E., Render, B., Ruth, S. and Ross, J. (1985). Human-oriented implementation cures cyberphobia. *Data management*, November, pp. 29-32.

Gee, Q. (2005). Review of script displays of African languages by current software. *New Review of Hypermedia and Multimedia*, 11(2): 247-255.

Giacoppo, A. (2001). The role of theory in HCI. CHARM: Choosing human-computer interaction (HCI) appropriate research methods, University of Maryland. Retrieved 15 October 2005 from the World Wide Web: http://www.otal.umd.edu/hci-rm/ theory.html.

Giles, J. and Middleton, T. (1999). *Studying culture: A practical introduction*. Massachusetts: Blackwell Publishers.

Gittins, D. (1986). Icon-based human-computer interaction. *International Journal of Man-Machine Studies*, 24: 519-543.

Glass, C.R. and Knight, L.A. (1988). Cognitive factors in computer anxiety. *Cognitive Therapy and Research*, 12(4): 351-366.

Goggin, J. and Wickens, D.D. (1971). Proactive interference and language change in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 10: 453-458.

Granić, A., Glavinić, V. and Stankov, S. (2004). *Proceedings of 8th ERCIM Workshop – "User Interfaces for All"*, Crete, Greece.

Green, P.A. (1979). Rational ways to increase pictographic symbol discriminability. PhD Thesis. University of Michigan.

Gribbons, W.W. (1997). Designing for the global community. *Proceedings of IEEE International Professional Communication Conference*, Salt Lake City, Utah, United States of America. pp. 261-273.

Griffiths, D., Heppel, S., Millwood, R. and Mladenova, G. (1994). Translating software: What it means and what it costs for small cultures and large cultures. *Computers in Education*, 22(1&2): 9-17.

Grudin, J. (1989). The case against user interface consistency. *Communications of the ACM*, 32(10): 1164-1173.

Hall, B. (2002). *Among Cultures: The Challenge of Communication*. Florida: Harcount, Inc.

Hall, E.T. (1959). *The Silent Language*. New York: Doubleday Press.

Hars, A. (1996). Localizing software is not just a multilingual issue; it's also multicultural. *BYTE*, March, 1996.

Harper, B.D. and Norman, K.L. (1993). Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5. *Proceedings of the First Mid-Atlantic Human Factors Conference*, Virginia Beach, Virginia, United States of America. pp. 224-228.

Hemenway, K. (1982). Psychological issues in the use of icons in command menus. *Proceedings of the ACM conference on Human Factors in Computer Systems*. pp. 20-24.

Heukelman, D. (2006). Can a user centred approach to designing a user interface for rural communities be successful? *Proceedings of CHI-SA 2006*, Cape Town, Republic of South Africa.

Hodge, R and Kress, G. (1988). *Social semiotics*. London: Priority Press.

Hofstede, G. (1991). *Cultures and Organizations: Software of the Mind*. McGraw-Hill Companies, Inc.

Hogaboam, T.W. and Pellegrino, J.W. (1978). Hunting for individual differences in cognitive processes: Verbal ability and semantic processing of pictures and words. *Memory and Cognition*, 6(2): 189-193.

Hugo, J. (2000). Human-computer interaction: A new challenge for the South African IT industry. Report based on the introductory speech at CHI-SA 2000. Retrieved 15 March 2005 from the World Wide Web: http://www.chi-sa.org.za/articles/ HCIchallenge.htm.

Hugo, J. (2002). HCI and multiculturalism in Southern Africa. Report on CHI-SA 2002. Retrieved 15 March 2005 from the World Wide Web: http://www.chi-sa.org.za/articles/ChiSAReport.htm.

Hunt, K. (1996). Reflections on an icon development process: Negotiating design issues. *Proceedings of the STC Annual Conference*.

Hutchison, A. (1997). Empty icons in the metaphor trap. *Proceedings of the Australian Society for Computers in Learning in Tertiary Education*, Perth, Australia.

ISO9241. (1998). ISO 9241-11: *Ergonomic requirements for office work with visual display terminals*. Berlin: Beuth.

ISO11581. (1999). ISO 11581-6: Information technology – User system interfaces and symbols – Icon symbols and functions – Part 6: Action icons.

Jackson, L.A., Biocca, F., Lim, L., Bradburn, K., Tang, M., Mou, W., Barbatsis, G., Von Eye, A., Zhao, Y. and Fitzgerald, H.E. (2003). Effects of culturally adapted interfaces on learning and attitudes: Findings from the HomeNetToo project. *Proceedings of the International Association for the Development of Information Society (IADIS) International Conference, WWW/Internet 2003*, Algarve, Portugal.

Jawahar, I.M. and Elango, B. (2002). Predicting end user performance. In M.A. Mahmood (Ed), *Advanced topics in end user computing* (pp. 177-187). United States of America: Idea Group Publishing.

Jay, T.B. (1981). Computerphobia: What to do about it. *Educational Technology*, 21(1): 47-48.

Jervell, H.R. and Olsen, K.A. (1985). Icons in man-machine communications. *Behaviour and Information Technology*, 4(3): 249-254.

Johassen, D.H. (1985). *State anxiety and exposure to microcomputers: Assessing computerphobia*. Paper presented at the Association for Educational Communications and Technology. Las Vegas, Nevada.

Johns, S.M. (1997). Colors, buttons, words and culture: Designing software for the global community. *1997 CODI Conference*. Mesa, AZ. Retrieved 24 February 2005 from the World Wide Web: http://www.sensi.org/~alec/locale/other/colore~1.htm.

Joncour, N., Sinclair, K.E. and Bailey, M. (1994). *Computer anxiety, computer experience and self-efficacy*. Paper presented at the Annual Conference of the Australian Association for Research in Education. Newcastle, New South Wales.

Jones, M.K. (1989). *Human-computer interaction: A design guide*. New Jersey: Educational Technology Publications.

Kacmar, C.J. and Carey, J.M. (1991). Assessing the usability of icons in user interfaces, *Behaviour and Information Technology*, 10(6): 443-457.

Kagan, J. and Havemann, E. (1980). *Psychology: An introduction* (4th Edition). New York: Jovanovich.

Kang, S.R. User experience: Beyond usability. (2003). *Bulletin of the 6th Asian Design conference: International Symposium on Design Conference*. p. 121.

Kang, K-S. and Corbitt, B. (2003). Effectiveness of graphical components in web site e-commerce application – A cultural perspective. *The Electronic Journal on Information Systems in Developing Countries*, 7(2): 1-6.

Keniston, K. (1997). Software localization: Notes on technology and culture. Working paper No. 26, Massachusetts Institute of Technology.

Kent, E.W. (1981). *The brains of men and machines*. New York: McGraw-Hill.

Kernan, M.C. and Howard, G.S. (1990). Computer anxiety and computer attitudes: An investigation of construct and predictive validity issues. *Educational and Psychological Measurement*, 50(3): 681-690.

Kim, J. and Lee, K. (2005). A study on relation between culture and mobile phone interface design with emphasis on recognition performance according to level of abstraction in icon style, Korea Society of Design Science. Seoul, Korea.

Kintsch, W. (1974). *The representation of meaning in memory*. New Jersey: Erlbaum.

Kintsch, W. (1977). *Memory and cognition*. New York: John Wiley and Sons.

Kluckhohn, C. (1949). *Mirror for Man*. New York: McGraw-Hill.

Kralisch, A. and Berendt, B. (2005). Language-sensitive search behaviour and the role of domain knowledge. *New Review of Multimedia and Hypermedia 2*.

Kroll, J.F. and Stewart, E. (1994). Category influences in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language*, 33: 149-174.

Kukulska-Hulme, A. (2000). Communication with users: Insight from second language acquisition. *Interacting with Computers*, 12: 587-599.

Kurniawan, S.H., Goonetilleke, R.S. and Shih, H.M. (2001). Involving Chinese users in analyzing the effects of languages and modalities on computer icons. In C. Stephanidis (Ed.), *Universal Access in HCI* (pp. 491-495). Mahwah, United States of America: Lawrence Erlbaum.

Kuter, U. and Yilmaz, C. (2001). Survey methods: Questionnaires and interviews. CHARM: Choosing human-computer interaction (HCI) appropriate research methods, University of Maryland. Retrieved 15 October 2005 from the World Wide Web: http://www.otal.umd.edu/hci-rm/survey.html.

Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Landauer, T.K. (1972). *Psychology: A brief overview*. New York: McGraw-Hill.

Lane, D.M., Napier, H.A., Peres, S.C. and Sándor, A. (2005a). Hidden costs of graphical user interfaces: failure to make the transition from menus and icon toolbars to keyboard shortcuts. *International Journal of Human-Computer Interaction*, 18(2): 133-144.

Lane, D.M., Peres, S.C., Sándor, A and Napier, H.A. (2005b). A process for anticipating and executing icon selection in graphical user interfaces. *International Journal of Human-Computer Interaction*, 19(2): 241-252.

Lansdale, M.W. (1988). On the memorability of icons in an information retrieval task. *Behaviour and Information Technology*, 7(2): 131-151.

Lansdale, M.W., Simpson, M. and Stroud, T.R.M. (1990). A comparison of words and icons as external memory aids in an information retrieval task. *Behaviour and Information Technology*, 9(2): 111-131.

Levi, M.D. and Conrad, F.G. (1998). *Usability testing of World Wide Web sites*. U.S. Department of Labor, Bureau of Labor Statistics, BLS Research Papers.

Lodding, K.N. (1983). Iconic interfacing. *IEEE Computer Graphics and Applications*, 24: pp. 11-20.

Loyd, B.H. and Gressard, C. (1984a). The effects of sex, age and computer experience on computer attitudes. *AEDS Journal*, 40: 67-77.

Loyd, B.H., Gressard, C. (1984b). Reliability and factorial validity of computer attitude scales. *Educational and Psychological Measurement*, 44(2): 501-505.

Luke, C. (1996). Reading gender and culture in media discourses and texts. In G. Bull and M. Anstey (Eds), *The literary lexicon*. Australia: Prentice Hall.

Lye, W.F. and Murray, C. (1980). *Transformations on the Highveld: The Tswana and Southern Sotho*. Cape Town: Creda Press.

Lynch, P.D., Kent, R.J. and Srinivasan, S.S. (2001). The global Internet shopper: Evidence from shopping tasks in twelve countries. *Journal of Advertising Research*, May/June: 15-23.

Mackey, W.F. (2005). The description of bilingualism. In L. Wei (Ed), *The bilingualism reader*. New York: Routledge.

Maguire, M.C. (1990). A review of human factors, guidelines and techniques for the design of graphical human-computer interfaces. In J. Preece and L. Keller (Eds.), *Human-computer interaction: Selected readings* (pp. 161-184). United Kingdom: Prentice Hall.

Maner, W. (1997). Internationalization of user interfaces. Paper based on T. Fernandes, *Global interface design: A guide to designing international user interfaces*. Retrieved 17 September 2004 from the World Wide Web: http://web.cs.bgsu.edu/maner/uiguides/internat.htm.

Marcoulides, G.A. (1989). Measuring computer anxiety: The computer anxiety scale. *Educational and Psychological Measurement*, 49(3): 733-739.

Marcus, A. (2001). *Cross-cultural user-interface design*. Retrieved 25 February 2005 from the World Wide Web: http://www.amanda.com/resources/HCII01/HCII01_PanXC_Marcus_160401.pdf.

Marcus, A. (2006). Culture: Wanted? Alive or Dead? *Journal of Usability Studies*, 2(1): 62-63.

Marcus, A., Smilonich, N. and Thompson. L. (1995). *The cross-GUI handbook for multiplatform user interface design*. Reading, Massachusetts: Addison-Wesley.

Martin, B.L. (1998). Computer anxiety levels of Virginia Cooperative Extension Field Personnel. PhD Thesis. Virginia Polytechnic Institute and State University.

Masoeu, A. and De Villiers, C. (2001). Web usability in a multicultural environment: A concern for young South African web users? *Proceedings of CHI-SA 2001*, Pretoria, Republic of South Africa.

Maurer, M.M. (1994). Computer anxiety correlates and what they tell us: A literature review. *Computers in Human Behavior*, 10(3): 369-376.

Mayhew, D. (1999). *The usability engineering lifecycle: A practitioner's handbook for user interface design*. San Francisco: Academic Press.

McDonald, T. and Blignaut, P. (2005). The effect of cultural differences on the efficiency of searches on a university website. *11th International Conference on Human-Computer Interaction*, Las Vegas, Nevada, United States of America.

McDougall, S. and Curry, M. (2004). More than just a picture: Icon interpretation in context. *Proceedings of first international workshop on coping with complexity*, University of Bath, Bath, United Kingdom. pp. 73-81.

McDougall, S., Forsythe, A. and Stares, L. (2005). Icon use by different language groups: Changes in icon perception in accordance with cue utility. In M. Costabile and F, Paternò (Eds), *Human-computer interaction – Interact 2005* (pp. 1083-1086).

McInerney, V., Marsh, H.W. and McInerney, D.M. (1999). The designing of the computer anxiety and learning measure (CALM): Validation of scores and on a multidimensional measure of anxiety and cognitions relating to adult learning of computer skills using structural equation modelling. *Educational and Psychological Measurement*, 59(3): 451-470.

Microsoft. (nd). You are not world-ready if... Your application contains localizable bitmaps. Retrieved 11 May 2005 from the World Wide Web: http://www.microsoft.com/globaldev/getWR/nwr/nwrpartV.mspx.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review*, 63: 81-97.

Miller, J. (2003). Technology roadmap for the human-computer interface (HCI): Preliminary overview. Department of Science and Technology, Pretoria.

Miller, L.C. (1994). Transborder tips and traps. *Byte*, 19(6): 93-102.

Mitchel, A. (2005). Do bilinguals access a shared or separate conceptual store? Creating false memories in a mixed-language paradigm. Psychology Honours Project, Macalester College.

Morris, L., Davis, D. and Hutchings, C. (1981). Cognitive and emotional components of anxiety: Literature review and revised worry-emotional scale. *Journal of Educational Psychology*, 75: 541-555.

Mouton, J. (2001). *How to succeed in your Master's and Doctoral studies: A South African guide and resource book*. Pretoria: Van Schaik Publishers.

Murrell. K.A. (nd). Human computer interface design in a multi-cultural multi-lingual environment. Retrieved 17 March 2005 from the World Wide Web: http://www.und.ac.za/users/murrell/classrm/paper1.html.

Muter, P. and Mayson, C. (1986). The role of graphics in item selection from menus. *Behaviour and Information Technology*, 5(1): 89-95.

Newman, W. and Lamming, M. (1995). *Interactive system design*. Massachusetts: Addison-Wesley.

Nielsen, J. (nd). International usability testing. Retrieved 28 February 2005 from the World Wide Web: http://www.useit.com/papers/international_usetest.html.

Nielsen, J. (1992). The usability engineering life cycle. *Computer*, March: 12-22.

Nielsen, J. (1994). *Usability engineering*. Boston: Academic Press Inc.

Nielsen, J. (1996a). International usability testing. In E. del Galdo and J. Nielsen (Eds), *International user interfaces* (pp. 1-19). New York: John Wiley and Sons.

Nielsen, J. (1996b). International Web Usability. Alertbox, August 1996. Retrieved 25 February 2005 from the World Wide Web: http://www.useit.com/alertbox/9608.html.

Nielsen, J. (1999). *Designing web usability*. Indianapolis: New Riders Publishing.

Nielsen, J. (2001a). Usability metrics. Alertbox, January, 2001. Retrieved 27 March 2006 from the World Wide Web: http://www.useit.com/alertbox/20010121.html.

Nielsen, J. (2001b). Success rate: The simplest usability metric. Alertbox, February, 2001. Retrieved 9 June 2006 from the World Wide Web: http://www. useit.com/alertbox/20010218.html.

Nielsen, J. (2003a). Recruiting test participants for usability studies. Alertbox, January, 2003. Retrieved 27 March 2006 from the World Wide Web: http://www.useit.com/alertbox/20030120.html.

Nielsen, J. (2003b). Usability 101: Introduction to usability. Alertbox, August, 2003. Retrieved 4 September 2006 from the World Wide Web: http://www. useit.com/alertbox/20030825.html.

Nielsen, J. (2004). Checkboxes vs. radio buttons. Alertbox, September, 2004. Retrieved 9 June 2006 from the World Wide Web: http://www.useit.com/ alertbox/20040927.html.

Nielsen, J. (2006). Outliers and luck in user performance. Alertbox. 6 March 2006. Retrieved 27 March 2006 from the World Wide Web: http://www.useit.com/ alertbox/outlier_ performance.html.

Nielsen, J. and Phillips, V.L. (1993). Estimating the relative usability of two interfaces: Heuristic, formal, and empirical methods compared. *Proceedings of the SIGCHI conference on Human factors in computing systems*, Amsterdam, The Netherlands. pp. 214- 221.

Norman, D.A. and Bobrow, D.G. (1976). Active memory processes in perception and cognition. In C.N. Cofer (Ed.), *The structure of human memory* (pp. 114-132). San Francisco: Freeman.

Norman, D. and Rumelhart, D. (1975). *Explorations in cognition*. San Francisco: Freeman.

Olivier, M. (2004). *Information technology research: A practical guide for Computer Science and Informatics* (2nd Edition). Pretoria: Van Schaik.

Onibere, E.A., Morgan, S., Busang, E.M. and Mpoeleng, D. (2001). Human-computer interface design issues for a multi-lingual English speaking country – Botswana. *Interacting with Computers*, 13: 497-512.

Orr, C., Allen. D. and Poindexter, S. (2001). The effect of individual differences on computer attitudes: An empirical study. *Journal of End User Computing*, April/June: 26-39.

O'Sullivan, P. (2003). Building cultural diversity into the development of multilingual applications. The LISA newsletter: Globalization Insider.

Ott, C. (1999). *Global solutions for multilingual applications: Real world techniques for developers and designers*. New York: John Wiley and Sons.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.

Pearrow, M. (2000). *Web site usability handbook*. Massachusetts: Charles River Media.

Peirce, C.S. (1985). Logic as semiotic: The theory of signs. In R.E. Innis (Ed.), *Semiotics: An introductory anthology* (pp. 1-23). Indiana University Press: Bloomington.

Pelc, J., Sebeok, T.A., Stankiewicz, E. and Winner, T.G. (1984). *Sign, system and function: Papers of the first and second Polish-American semiotics colloquia*. Berlin: Walter de Gruyter & Co.

Peurifoy, R.Z. (1992). *Anxiety, phobias and panic: Taking charge and conquering fear: A step-by-step program for regaining control of your life* (2nd Edition). New York: Warner Books.

Piamonte, D.P.T. (2000). Using multiple performance parameters in testing small graphical symbols. Doctoral Thesis. Lulea Tekniska University, Sweden.

Piamonte, D.P.T., Ohlsson, K. and Abeysekera, J. (1999). Across the seas: A user-based evaluation of candidate telecommunication icons. Research report 1999:06.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. and Carey, T. (1994). *Human-computer interaction*. England: Addison-Wesley.

Pressley, G.M. (1976). Mental imagery helps eight-year-olds remember what they read. *Journal of Educational Psychology*, 68: 355-359.

Pope-Davis, D.B. and Twing, J.S. (1991). The effects of age, gender, and experience on measures of attitude regarding computers. *Computers in Human Behavior*, 7: 333-339.

Popovich, P.M., Hyde, K, R. and Zakrajsek, T. (1987). The development of the attitudes towards computer usage scale. *Educational and Psychological Measurement*, 47(1): 261-269.

Potosnak, K. (1988). Do icons make interfaces easier to use? *IEEE Software*, 5(3), May/June: 97-99.

Potter, M.C., So, K-F., Von Eckardt, B., and Feldman, L.B. (1984). Lexical and conceptual representation in forward and backward translation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23: 23-38.

Raskin, J. (2001). *The humane interface: New directions for designing interactive systems*. Massachusetts: Addison-Wesley.

Rauh, M.T. and Seccia, G. (2006). Anxiety and performance: An endogenous learning-by-doing model. *International Economic Review*, 47(2): 583-609.

Read, J., MacFarlane, S. and Casey, C. (2001). Measuring the usability of text input methods for children. In A. Blandford, J. Vanderdonckt and P. Gray (Eds), *People and computers XV – Interaction without frontiers: Joint proceedings of HCI 2001 and IHM 2001* (pp. 559-572).

Redmond-Pyle, D. and Moore, A. (1995). *Graphical user interface design and evaluation (guide): A practical process*. Great Britain: Prentice Hall.

Reece, M.J. and Gable, R.K. (1982). The development and validation of a measure of general attitudes toward computers. *Educational and Psychological Measurement*, 42(3): 913-916.

Reingold, H. (1989). The interface of tomorrow, today. *Infoworld*.

Richards, S., Barker, P., Banerji, A., Lamont, C. and Manji, K. (1994). The use of metaphors in iconic interface design. *Intelligent Tutoring Media*, 5(2): 58-62.

Roberts, T.L. and Moran, T.P. (1983). The evaluation of text editors: Methodology and empirical results. *Communications of the ACM*, 26(4): 265-283.

Rogers, Y. and Oborne, D.J. (1987). Pictorial communication of abstract verbs in relation to human-computer interaction. *British Journal of Psychology*, 78: 99-112.

Rogers, Y. (1989a). Icon design for the user interface. In D.J. Oborne (Ed.) *International Reviews of Ergonomics* (pp. 129-155). London: Taylor and Francis.

Rogers, Y. (1989b). Icons at the interface: Their usefulness. *Interacting with Computers*, 1(1): 105-117.

Rosson, M.B. (1984a). Characterizing freeform editing behavior. IBM Research Report RC 10550, IBM T. J. Watson Research Center, Yorktown Heights, New York.

Rosson, M.B. (1984b). Effects of experience on learning, using, and evaluating a text editor. *Human Factors*, 26(4): 463-475.

Russo, P. and Boor, S. (1993). How fluent is your interface? Designing for international users. *Proceedings of the 4th Conference on human factors in computing systems*. pp. 342-347.

Sadoski, M., Goetz, E.T., Olivarez, A., Lee, S. and Roberts, N.M. (1990). Imagination in story reading: The role of imagery, verbal recall, story analysis, and processing levels. *Journal of Reading Behavior*, 22: 55-70.

Sadoski, M. and Paivio, A. (2004). A dual coding theoretical model of reading. In R. B. Ruddell and N. J. Unrau (Eds), *Theoretical models and processes of reading* (5th edition) (pp. 1329-1362). Newark: International Reading Association.

Schild, W., Power, L.R. and Karnaugh, M. (1980). Pictureworld: A concept for future office system. RC 8384 (#36518), IBM, Thomas J. Watson Research Center.

Scholtz, J. (2004). *Usability evaluation*. Publication #545, National Institute of Standards and Technology.

Sen, R.N. and Chen, I. (2002). Some studies on ergonomic design of icons and related graphics for users in Malaysia. *Proceedings for the 3rd International Conference on Ergonomics*, CybErg' 2002, Wits University, Republic of South Africa.

Shepard, R.N. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6: 156-163.

Sholl, A., Sankaranarayanan, A. and Kroll, J. (1995). Transfer between picture naming and translation: Test of asymmetries in bilingual memory. *Psychological Science*, 6: 45-49.

Shneiderman, B. (1980). *Software psychology: Human factors in computer and information systems*. Massachusetts: Winthrop Publishers.

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3rd Edition). Massachusetts: Addison-Wesley.

Shneiderman, B. (2003). *Leonardo's laptop: Human needs and the new computing technology*. Massachusetts: The MIT Press.

Singh, N. (2002). Analyzing cultural sensitivity of websites. *Journal of Practical Global Business*, April.

Sittig D.F., Kuperman G.J. and Fiskio J. (1999). Evaluating physician satisfaction regarding user interactions with an electronic medical record system. *Proceedings of the AMIA 1999 Annual Symposium*.

Slaughter, L., Harper, B. and Norman, K. (1994). Assessing the equivalence of the paper and on-line formats of the QUIS 5.5. *Proceedings of the Second annual Mid-Atlantic Human Factors Conference*. pp. 87-91.

Smith, D.C., Irby, C., Kimball, R. and Verplank, B. (1982). Designing the STAR user interface. *Byte*, 7: 242-282.

Staggers, N. (2000). Comparing response time, errors, and satisfaction between text-based and graphical user interfaces during nursing order tasks. *Journal of the American Medical Informatics Association*, 7(2): 164-176.

Stander, A. (1997). Computer user interfaces in a multicultural society. M.Sc. dissertation. Cape Technikon.

Stander, A. (1998). Bridging the gap: Issues in the design of computer user interfaces for multicultural communities. In C. Ess and F. Sudweeks (Eds), *Proceedings cultural attitudes towards communication and technology '98* (pp. 255-260).

Statistica Electronic Manual. (2005).

Sullivan, P. (1989). Human-computer interaction perspectives on word-processing issues. *Computers and Composition*, 6(3): 11-33.

Sylaiou, S., Almosawi, A., Mania, K. and White, M. (2004). Preliminary evaluation of the augmented representation of cultural objects system. *Proceedings of Virtual Systems and Multimedia 2004* (VSMM 2004), Japan.

Szajna, B. (1994). An investigation of the predictive validity of computer anxiety and computer aptitude. *Educational and Psychological Measurement*, 54(4): 926-934.

Taylor, D. (1992). *Global software*. New York: Springer-Verlag.

Taylor, V. and De Villiers, R. (2005). A pilot study into localised interface elements for the indigenous African adult. *Proceedings CHI-SA 2005*, Pretoria, Republic of South Africa.

Teklebrhan, R. and Blignaut, P. (2005). A study on the effect of Western designed metaphors in some culture groups in South Africa. Technical Report 2005/02, Department of Computer Science and Informatics: University of the Free State.

Translate.org.za. (nd). Accessed 13 June 2006 on the World Wide Web: www.translate.org.za.

Trochim, W.M. (2002). *Research methods knowledge base*. Retrieved 31 August 2006 from the World Wide Web: http://trochim.human.cornell.edu/kb/.

Trompenaars, F. (2004). Meeting of the minds: Gaining competitive advantage through reconciling cultural differences. Retrieved 27 February 2005 from the World Wide Web: www.eurofinas.org/database/SorrentoPresentations/Trompenaars.pdf.

Tuck, M. (2001). The real history of the GUI. Retrieved 28 April 2005 from the World Wide Web: http://www.sitepoint.com/print/real-history-gui.

Tudor, L.G. (1994). Growing an icon set: User acceptance of abstract and concrete icon styles. In B. Blumenthal, J. Gornostaev and C. Unger (Eds), *Human-computer interaction: 4th international conference, EWCHI '94*, St Petersburg, Russia.

Urdan, T.C. (2005). *Statistics in plain English* (2nd Edition). New Jersey: Lawrence Erlbaum Associates.

Usability glossary. (nd). Available on the World Wide Web: http://www. usabilityfirst.com.

Van Belle, J.P., Fellstad, R., Steele, C. and Van Bakel, W. (2003). Multi-cultural websites in a multi-cultural country: A South African perspective. *Proceedings of the Third International Conference on Electronic Business (ICEB 2003)*, Singapore.

Wallace, D.F., Norman, K.L. and Plaisant, C. (1998). The American voice and robotics "Guardian" system: A case study in user interface usability evaluation. Technical Report CAR-TR-392. University of Maryland.

Walsh, A. and Ollenburger, J.C. (2001). *Essential statistics for the social and behavioural sciences: A conceptual approach*. New Jersey: Prentice-Hall.

Wei, L. (2005). Dimensions of bilingualism. In L. Wei (Ed.), *The bilingualism reader*. New York: Routledge.

Whiteside, J., Jones, S., Levy, P.S. and Wixon, D. (1985). User performance with command, menu and iconic interfaces. *Proceedings of CHI '85 Conference*, San Francisco, United States of America. pp. 185-191.

Wikipedia. (2006). Word processor. Retrieved 12 September 2006 from the World Wide Web: http://en.wikipedia.org/wiki/Word_processor.

Wine, J.D. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, 76: 92-104.

Woodman, T and Hardy, L. (2001). Stress and anxiety. In R.N. Singer, H.A. Hausenblas and C.M. Janelle (Eds). *Handbook of Sport Psychology* (2nd Edition) (pp. 290-318). New York: Wiley and Sons.

Woodrow, J.E. (1991). A comparison of four computer attitude scales. *Journal of Educational Computing Research*, 7(2): 165-187.

Zahedi, F., Van Pelt, W. and Song, J. (2001). A conceptual framework for international web design. *IEEE Transactions on Professional Communication*, 44(2): 83-103.

Zammit, K. (2000). Computer icons: A picture says a thousand words, or does it? *Journal of Educational Computing Research*, 23(2): 217-231.

Zayas, E. (1996). Usability and Pedagogical issues in user interface design. Doctoral paper. Retrieved 13 February 2006 from the World Wide Web: http://www.technosphere.net/usability_in_uid.htm.

Bibliography

Amorim, L. (2001). Intercultural Learning: A few awareness tips for US and European Fellows & Host Community Foundations. European Foundation Centre: Community Foundation Transatlantic Fellowship Orientation Session.

Badre, A. (2000). The Effects of Cross Cultural Interface Design Orientation on World Wide Web User Performance. Technical Report Number: GIT-GVU-01-03, Georgia Institute of Technology. Retrieved 25 February 2005 from the World Wide Web: ftp://ftp.cc.gatech.edu/pub/gvu/tr/2001/01-03.pdf.

Barber, W. and Badre, A. (1998). Culturability: The merging of culture and usability. *Proceedings of the 4th Conference on Human Factors and the Web*, New Jersey, United States of America.

Bohmann, K. (2000). Calculating usability effects. Retrieved 19 October 2005 from the World Wide Web: http://bohmann.dk/articles/calculating_usability_effects.html.

Bourgess-Waldegg, P. and Scrivener, S.A.R. (1998). Meaning, the central issue in cross-cultural HCI design. *Interacting with computers*, 9: 287-309.

Brems, D.J. and Whitten, W.B. (1987). Learning and preference for icon-based interface. *Proceedings of the Human Factors Society 31st Annual Meeting*. pp. 125-129.

Casson, R.W. (1981). *Language, Cognition and Culture*. New York: Macmillan Publishing Co, Inc.

Chau, P.Y.K., Cole, M., Massey, A.P., Montoya-Weiss, M. and O'Keefe, R. (2002). Cultural differences in the online behavior of consumers. *Communications of the ACM*, 45(10): 138-143.

Cleary, Y. (2000). An examination of the impact of subjective cultural issues on the usability of a localized web site - The Louvre Museum Web Site. *Museums and the Web 2000: Conference Proceedings*, Pittsburgh, Pennsylvania.

Cushner, K. and Brislin, R.W. (1986). *Cross-cultural research and methodology (Vol. 9). Intercultural interactions: A practical guide* (2nd Edition). London: Sage Publications.

Evers, V. (1997a). Human-Computer Interfaces: Designing for Culture. PhD thesis. University of Amsterdam.

Evers, V. (1997b). The role of culture in interface acceptance. In S. Howard, J. Hammond and G. Lindegaard (Eds), *Human computer interaction*, Interact '97, London: Chapman and Hall.

Evers, V. (1998). Cross-cultural understanding of metaphors in interface design. In C. Ess and F. Sudweeks (Eds), *Proceedings Cultural Attitudes Towards Technology and Communication (CATAC) '98*, University of Sydney, Australia.

Evers, V. (2002). Cross-cultural applicability of user evaluation methods: A case study amongst Japanese, North-American, English and Dutch users. *Proceedings of CHI 2002*, Minneapolis, Minnesota.

Ford, G. and Gelderblom, H. (2003). The effects of culture on performance achieved through the use of human computer interaction. In J. Eloff., P. Kotzé, A. Engelbrecht and M. Eloff (Eds), *IT Research in developing countries – Proceeds of SAICSIT 2003*. pp. 218-230.

Ford, G. and Kotzé, P. (2005). Designing usable interfaces with cultural dimensions. *Interact*, 2005: 713-726.

Greenwood, T. G. (1993). International cultural differences in software. *Digital Technical Journal,* 5(3): 8-20. Retrieved 24 February 2005 from the World Wide Web: http://research.compaq.com/wrl/DECarchives/DTJ/DTJB01/DTJB01SC.TXT.

Goodman, M.E. (1967). *The Individual and Culture*. Illinois: The Dorsey Press.

Hall, E.T. (1976). *Beyond Culture*. New York: Doubleday Press.

Hall, E.T. (1966). *The Hidden Dimension*. New York: Doubleday Press.

Hoft, N. (1996). Developing a Cultural Model. In E. del Galdo and J. Nielsen (Eds). *International User Interfaces*. New York: John Wiley and Sons.

Kampurri, M., Tedre, M. and Tukiainen, M. (2006). Towards the sixth level in interface design: Understanding culture. *Proceedings of CHI-SA 2006*, Cape Town, Republic of South Africa.

Karat, J. and Karat, C-M. (1996). World-wide CHI: Perspectives on design and internationalization. *SIGCHI*, 28(1).

Kaschula, R.H. and Anthonissen, C. (1995). *Communicating across Cultures in South Africa: Toward a critical language awareness*. Johannesburg: Hodder and Stoughton.

Komlodi, A. and Carlin, M. (2004). Identifying cultural variables in information-seeking. *Proceedings of the 10th Americas Conference on Information Sciences*, New York, United States of America.

Luna, D., Peracchio, L.A. and de Juan, M.D. (2003). The impact of language and congruity on persuasion in multicultural E-marketing. *Journal of Consumer Psychology*, 13(1&2): 41-50.

Marcus, A., Armitage, J., Frank, V. and Guttman, E. (1999). Globalization of user-interface design for the web. *Proceedings of the 5th conference on human factors and the web*, Maryland, United States of America. Retrieved 14 March 2005 from the World Wide Web: http://www.ischool.utexas.edu/~i385ef04/readings/Marcus-globalization.htm.

Marcus, A. (2001). Cultural Dimensions and Global Web Design: What? So What? Now What? *Proceedings of the South African Human-Computer Interaction Conference 2000.*

Mayes, J.T., Draper, S.W., McGregor, A.M. and Oatley, K. (1988). Information flow in a user interface: The effect of experience and context on the recall of MacWrite screens. In D. Jones and R. Winder (Eds), *People and Computers IV*, Cambridge: Cambridge University Press.

Petre, M. (1995). Why looking isn't always seeing: Readership skills and graphical programming. *Communications of the ACM*, 38(6): 33-44.

Purvis, M., Hwang, P., Purvis, M., Madhavji, N. and Cranefield, S. (2001). A practical look at software internationalisation. *Transactions of the SDPS*, 5(3): 79-90.

Sacher, H. and Margolis, M. (2000). The culture of interaction: About foreign and not-so-foreign languages. *ACM Interactions*, 7(1): 39-45.

Sasse, M.A. (1997). Eliciting and Describing Users' Models of Computer Systems. PhD Thesis. University of Birmingham, England.

Schein, E.H. (1985). *Organizational Culture and Leadership*. San Francisco, California: Jossey-Bass Inc.

Shackel, B. (1990). Human factors and usability. In J. Preece and L. Keller (Eds), *Human-computer interaction: Selected readings* (pp. 27-41). United Kingdom: Prentice Hall.

Shen, J. (2000). User Interface Internationalization. Retrieved 12 February 2005 from the World Wide Web: http://eies.njit.edu/~turoff/coursenotes/CIS732/samplepro/user_interface_internationalizat.htm.

Sheppard, C. and Scholtz, J. (1999). The effects of cultural markers on web site use. *Proceedings of the 5th Conference on Human Factors & the Web*.

Singh, N. and Baack, D. (2004). Web site adaptation: A cross-cultural comparison of U.S. and Mexican web sites. *Journal of Computer-Mediated Communication*, 9(4).

Slaughter, L., Norman, K.L. and Shneiderman, B. (1995). Assessing users' subjective satisfaction with the Information System for Youth Services (ISYS). Technical Report CS-TR-3463, University of Maryland.

Stengers, H., De Troyer, O., Baetens, M. Boers, F. and Mushtaha, A.N. (2004). Localization of web sites: Is there still a need for it? *International Workshop on Web Engineering*. Santa Cruz, United States of America.

Steyn, M.E. and Motshabi, K.B. (Eds). (1996). *Cultural Synergy in South Africa: Weaving Strands of Africa and Europe*. Randburg: Knowledge Resources.

Sturm, C. (2005). Approaches for a successful product localization. *Mensch und Computer*. Linz, Austria.

Syarief, A., Giard, J. R., Detrie, T. and Mcbeath, M. K. (2003). An initial cross-cultural survey of user perception on web icon design for travel websites, *6th Asian Design Conference*.

The IEEE, I. (1990). *Standard glossary of software engineering terminology*. New York: IEEE.

Uden, L. and Dix, A. (2000). Iconic interfaces for kids on the Internet. *IFIP World Computer Congress*. Beijing. pp. 279-286.

Ulijn, J.M. and St. Amant, K. (2000). Mutual intercultural perception: How does it affect technical communication? Some data from China, the Netherlands, Germany, France, and Italy. *Technical Communication Online*, 47(2). Retrieved 16 February 2005, from the World Wide Web http://www.techcomm-online.org/issues/v47n2/full/0400.html.

Van Dam, N., Evers, V. and Arts, F.A. (2005). Cultural user experience issues in e-government: Designing for a multi-cultural society. In P. van den Besselaar and S. Koizumi (Eds.), *Digital Cities 3: Information technologies for social capital – cross cultural perspectives* (pp. 310-324). Berlin: Springer-Verlag.

Van Greunen, D. and Wesson, J. (2001a). Formal usability testing – Informing design. *South African Institute of Computer Scientists and Information Technologists Annual Conference*.

Van Greunen, D. and Wesson, J. (2001b). Usability testing: From theory to practice. *Proceedings of CHI SA 2001*, Pretoria, Republic of South Africa.

Vöhringer-Kuhnt, T. (2002). The influence of culture on usability. Masters Thesis. Freie Universitat, Berlin.

Walton, M., Vukovic, V. and Marsden, G. (2002). 'Visual literacy' as challenge to the internationalisation of interfaces: A study of South African student web users. *Proceedings of CHI 2002*, Minneapolis.

Wu, J. (2000). Accommodating both experts and novices in one interface. Retrieved 25 July 2006 from the World Wide Web: http://www.otal.umd.edu/UUGuide/jingwu/.

Yeo, A. (1996). World-wide CHI: Cultural user interfaces, a silver lining in cultural diversity. *SIGCHI*, 28(3): 4-6.

Appendix A

Computer anxiety scale

Key:
1=Strongly disagree, 2=Disagree, 3=Mildly disagree, 4=Mildly agree, 5=Agree, 6=Strongly agree

I feel anxious when …	**SD**	**D**	**MD**	**MA**	**A**	**SA**
1. ... thinking about taking a class in a programming language (e.g. Basic, Pascal, C++, etc).	1	2	3	4	5	6
2. … applying for a job that requires some training in computers.	1	2	3	4	5	6
3. … sitting in front of a home computer.	1	2	3	4	5	6
4. ... being around people who are "into" computers.	1	2	3	4	5	6
5. … watching a movie about an intelligent computer.	1	2	3	4	5	6
6. … looking at a computer print-out.	1	2	3	4	5	6
7. … getting error messages from the computer.	1	2	3	4	5	6
8. … watching or listening to news programs about the increasing role of computers in society.	1	2	3	4	5	6
9. … watching someone working at a computer terminal.	1	2	3	4	5	6
10. … being refused information because a terminal is "down".	1	2	3	4	5	6
11. ... talking to a computer programmer.	1	2	3	4	5	6
12. … learning to write computer programmes.	1	2	3	4	5	6
13. … using a typewriter.	1	2	3	4	5	6
14. … visiting a computer shop.	1	2	3	4	5	6
15. ... attending a workshop on the uses of computers.	1	2	3	4	5	6
16. ... erasing or deleting material from a computer.	1	2	3	4	5	6
17. … thinking about programmes for computers.	1	2	3	4	5	6
18. … taking a class about the uses of computers.	1	2	3	4	5	6
19. … learning computer technology.	1	2	3	4	5	6
20. … looking at a high speed computer printer.	1	2	3	4	5	6

Appendix B

Computer attitude scale

Key:
1=Strongly disagree, 2=Disagree, 3=Mildly disagree, 4=Mildly agree, 5=Agree, 6=Strongly agree

	SD	D	MD	MA	A	SA
1. The challenge of solving problems with computers does not appeal to me.	1	2	3	4	5	6
2. Figuring out computer problems does not appeal to me.	1	2	3	4	5	6
3. I think working with computers would be enjoyable and stimulating.	1	2	3	4	5	6
4. I would like working with computers.	1	2	3	4	5	6
5. I do not enjoy talking with others about computers.	1	2	3	4	5	6
6. I don't understand how some people can spend so much time with computers and seem to enjoy it.	1	2	3	4	5	6
7. Once I start working with computers, I would find it hard to stop.	1	2	3	4	5	6
8. I will do as little work with computers as possible.	1	2	3	4	5	6
9. Computers do not scare me at all.	1	2	3	4	5	6
10. I have lots of self-confidence when it comes to working with computers.	1	2	3	4	5	6
11. I get a sinking feeling when trying to use a computer.	1	2	3	4	5	6
12. I would feel comfortable working with a computer.	1	2	3	4	5	6
13. I think using a computer would be very hard for me.	1	2	3	4	5	6
14. Generally, I would feel OK about trying a new problem on a computer.	1	2	3	4	5	6
15. I am no good with computers.	1	2	3	4	5	6
16. I'm not the type to do well with computers.	1	2	3	4	5	6
17. I do not feel threatened when others talk about computers.	1	2	3	4	5	6
18. It will not bother me at all to take a computer class.	1	2	3	4	5	6
19. I will feel at ease in a computer class.	1	2	3	4	5	6
20. I will get good marks in a computer course.	1	2	3	4	5	6
21. I do not think I can handle a computer course.	1	2	3	4	5	6
22. I am sure I can work with computers.	1	2	3	4	5	6
23. I am sure that I can learn an application program.	1	2	3	4	5	6

Appendix C

Questionnaire for user interaction satisfaction

Adapted from Shneiderman (1998). *Designing the User Interface*. p 136 – 143.

PART 3: Overall User Reactions

3.1 Overall reaction to the system:	**Terrible**							**Wonderful**		
	1	2	3	4	5	6	7	8	9	NA
3.2	**Frustrating**							**Satisfying**		
	1	2	3	4	5	6	7	8	9	NA
3.3	**Dull**							**Stimulating**		
	1	2	3	4	5	6	7	8	9	NA
3.4	**Difficult**							**Easy**		
	1	2	3	4	5	6	7	8	9	NA
3.5	**Inadequate**							**Adequate**		
	1	2	3	4	5	6	7	8	9	NA
3.6	**Rigid**							**Flexible**		
	1	2	3	4	5	6	7	8	9	NA

PART 6: Learning

6.1 Learning to operate the system	**Difficult**							**Easy**		
	1	2	3	4	5	6	7	8	9	NA
6.1.1 Getting started	**Difficult**							**Easy**		
	1	2	3	4	5	6	7	8	9	NA
6.1.2 Learning advanced features	**Difficult**							**Easy**		
	1	2	3	4	5	6	7	8	9	NA
6.1.3 Time to learn to use the system	**Slow**							**Fast**		
	1	2	3	4	5	6	7	8	9	NA

PART 7: System capabilities

7.4 Correcting your mistakes	**Difficult**							**Easy**		
	1	2	3	4	5	6	7	8	9	NA
7.4.1 Correcting typos	**Complex**							**Easy**		
	1	2	3	4	5	6	7	8	9	NA
7.4.2 Ability to undo operations	**Inadequate**							**Adequate**		
	1	2	3	4	5	6	7	8	9	NA
7.5 Ease of operation depends on your	**Never**							**Always**		

level of experience

	1	2	3	4	5	6	7	8	9	NA
7.5.1 You can accomplish tasks knowing only a few commands	**With difficulty**								**Easily**	
	1	2	3	4	5	6	7	8	9	NA
7.5.2 You can use features/shortcuts	**With difficulty**								**Easily**	
	1	2	3	4	5	6	7	8	9	NA

Appendix D

Icon preference questionnaire

Nommer: ____________________

Number the icons listed below 1, 2, 3 according to your order of preference for each function.

New						New
Open						Open
Cut						Cut
Undo						Undo
Bold		**B**		**F**		Bold
Left Align						Left
Save						Save
Italic		*I*		*F*		Italic
Copy						Copy
Right Align						Right
Close						Close
Bullet						Bullet
Font						Font
Underline		U		F		Under line
Paste						Paste
Centre Align						Centre

Appendix E

Published papers

Two papers based on the research of this study have been accepted for publication. These are as follows:

- **Appendix E-1**

  Beelders, T., Blignaut, P. and McDonald, T. 2005. A Proposed Study to Determine the Effect of Culture on the Usability of Word Processors. *Proceedings of the 4th Conference on Human-Computer Interaction in Southern Africa (CHI-SA)*, Johannesburg, Republic of South Africa. pp. 29-33.

- **Appendix E-2**

  Beelders, T., Blignaut, P. and McDonald, T. 2005. A word processor prototype to assess the effect of culture on usability. *Proceedings of the South African Telecommunications Networks and Applications Conference (SATNAC)*, Drakensberg, Republic of South Africa.

A further two papers based on the research of this study have been submitted for review either for publication in a journal or for presentation at a conference. These are as follows:

- **Appendix E-3**

  Beelders, T., Blignaut, P., McDonald, T. and Dednam, E. Measuring user performance for different interfaces using a word processor prototype. Submitted to the International Journal of Human-Computer Interaction.

- **Appendix E-4**

  Beelders, T., Blignaut, P., McDonald, T. and Dednam, E. The impact of different icon sets on the usability of a word processor. Submitted to the 12th International Conference on Human-Computer Interaction, Beijing, People's Republic of China.

A Proposed Study to Determine the Effect of Culture on the Usability of Word Processors

Tanya Beelders **Theo McDonald** **Pieter Blignaut**
University of the Free State
Department of Computer Science and Informatics
P.O. Box 339, Bloemfontein, South Africa, 9300
beelderstr.sci@mail.uovs.ac.za / theo.sci@mail.uovs.ac.za / pieterb.sci@mail.uovs.ac.za

ABSTRACT

This paper covers a proposed research topic in the field of Human Computer Interaction (HCI), specifically a methodology to investigate the usability of a number of culturally adapted user interfaces within a word processor environment. A number of concepts, such as culture and usability are also discussed as background to the research problem.

Keywords

Culture, usability

INTRODUCTION

South Africa is truly a multicultural nation whose cultural diversity is protected under the South African constitution. However, despite these protective measures very little has been done in an effort to recognise this cultural diversity in technology. Very few South African websites are available in any other language besides English, even though English as a first language is in the minority [22].

As South Africa attempts to establish equilibrium between the many cultures it is important that technology does not get left behind. To prevent this from materialising it is the responsibility of computer professionals to ensure that the cultural diversity of South Africa is fully represented in software applications. This research study will attempt to provide evidence of the importance of this cultural inclusion.

Any reference to a cultural interface in this paper extends beyond a simple translation of an interface and incorporates the imagery, metaphors, colours and any other culturally sensitive component of an interface.

PROPOSED RESEARCH PROBLEM

As mentioned previously, despite the multicultural diversity that exists in the present day South Africa, very little effort has been focused on accommodating these diversities into software applications. This research project will investigate the effect different South African cultures have on the usability of applications. Unfortunately, only a subset of South African cultures can be included in the study and as such we aim to include only two or three different cultures, depending on availability.

The application that was chosen to be used in the study is a word processor. These types of applications have managed to become an every-day tool for the vast majority of the population. Existing word processors have, if any, a Western or Americanised, culturally biased interface. What this study proposes is to test a number of word processor interfaces designed specifically for the participating South Africa cultures and to eventually determine and propose the most effective and usable interfaces for each culture tested.

LITERATURE REVIEW

Internationalisation of an interface refers to designing the interface in such a way that a single design can be used internationally or world-wide [16]. At the other end of the scale is localisation which refers to adapting the interface for a country in such a way that it is specific for that locale [16]. In order to localise a product Shneiderman [19] recommends paying careful attention to, amongst others, the following characteristics of an interface with regard to cultural and international diversity

- Left-to-right vs. right-to-left
- Date and time formats
- Numeric and currency formats
- Icons, buttons and colours
- Etiquette, policy, tone, formality, metaphors

These listed characteristics correspond closely to the concepts of objective and subjective culture where objective culture are the visible parts of culture such as the dress and language of the culture and subjective culture consists of those elements of culture that are hard to examine such as values and norms [4].

Localising software extends far beyond translation [11] and in contrast to such authors as Nielsen [16] and

Esselink [5], who advocate that localisation is concerned with translation of the interface, it has been determined that translation must be carefully considered as not all users prefer carrying out tasks, such as information searches, in their mother tongue [3].

Most research in the field of culture and usability has been focused on the World Wide Web (WWW). This attention stands to reason due to the inherent global nature of the Internet. From this research it has become evident that there are cultural markers present in web sites and more importantly that these cultural markers are influenced by the country of origin and the genre of the web site [1]. It was also discovered that whilst users may indicate a preference for an interface designed with another culture in mind, measured user performance, such as task completion time and number of mouse clicks, contradicts the user satisfaction opinions as performance was consistently better when using a product with an interface designed for their specific culture [1]. Users also showed markedly improved performance when presented with a range of tasks to be completed on a predetermined web site, which was specifically designed for their culture [18].

From a usability perspective studies have shown that the attitude of different cultures varied significantly towards usability of a product and in particular when usability was measured according to individualism/collectivism [23]. Cultural differences also have a profound effect on user preferences and acceptance of a product [8]. These findings have been corroborated with the discovery that there exist cultural preferences for interface components such as language, navigation, symbols and colour use [2] and that emphasis must be placed on catering for various cultures when designing icons and menu spacing, using colours or date and time conventions etc [14].

It is evident that cross-cultural analysis must play an increasingly important part in the design process [15] as, although there are many aspects of a software package that must be adapted for the targeted culture, it is the user interface which serves as the point of contact between the user and the computer. As such it is most likely to create a potential culture clash [15]. Because of this fact, the interface must be modified to the expectations of the user and in order for this to be accomplished successfully the designer should pay special attention to components such as geometric layout, images, sound, symbols and colours [9].

Developers should also keep in mind that when doing usability testing, commonly used observation methods may not be appropriate for all cultures [6].

CULTURE

Hofstede [12] defines culture as "the collective programming of the mind which distinguishes the members of one group or category of people from one another" whilst Hall [10] defines it as "a historically shared system of symbolic resources through which we make our world meaningful".

A cultural model is considered to be the most effective method to analyse culture and to determine cultural sensitivity of a user interface [4].

Following is a summary of cultural models reviewed by Del Galdo and Nielsen [4]:

1. Edward T. Hall determined what releases the right response rather than what sends the right message. In so doing Hall identified cultural variables such as speed of messages which measures how fast a person decodes and reacts to a received message as well as action chains which consist of a series of events which eventually lead to the accomplishment of a goal.

2. David A. Victor developed the LESCANT cultural model which was useful in determining the aspects of culture most likely to affect communication specifically in a business setting. One of the variables identified in the model is social organisation which studies how social, religious and economic systems affect intercultural communication.

3. Geert Hofstede centred his research on determining the patterns of thinking, feeling, and acting that form a culture's mental programming. This research lead to the discovery of five national cultural variables including power distance which is concerned with power structure within a society and uncertainty avoidance which measures the extent to which new or unfamiliar situations threaten members of the culture.

4. Finally, Fons Trompenaars concentrated on the way in which a group of people solves problems which lead to the definition of seven variables including neutral vs. affective which determines whether members of the culture are comfortable with displaying their emotions and achievement vs. ascription which looks at the cultural society's opinions on status – whether it is a birthright or if it must be earned.

By using the cultural variables to build a cultural profile of study participants and measuring the usability of the product for each included culture, conclusions can be drawn as to which interface components complement each culture to the greatest extent.

USABILITY

According to the International Standards Organisation (ISO) standard 9126-1 usability is "the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions" or as defined by ISO 9241-11 usability is "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use".
The three levels of internationalisation of an interface are defined as follows [4]:

Comprehensibility: A computer interface that is capable of displaying the user's native language, character set, and notations, such as currency symbols.
Usability: A computer interface that is understandable and usable in the user's native language.
Desirability: A system that is able to produce systems that *accommodate users' cultural characteristics*.
The following diagram has been proposed as an illustration of internationalisation issues and levels of a software product [17].

Objectivity Levels	Internationalisation Issues	Example	Current Research Examples
Comprehensibility	Language	Product language localisation	Unicode Machine Translation Microsoft knowledge base for common computer translation
Usability	Institutional matters	Time zone, date format, currency, measurement	
	Environmental factors	Aesthetics, icons and symbols	ISO symbols for interface
			Microsoft knowledge base for international colour use
Desirability	Social Conventions	Forms and Values	Culture model

Figure 3: Internationalisation Issues and Levels

By measuring usability factors on numerous interfaces, this study aims to determine which interface is best suited to each participating culture. The interface which will be deemed the most suitable will be that one which shows the greatest usability in terms of efficiency, effectiveness, learnabilty and satisfaction for that particular culture.

PROPOSED METHODOLOGY

A sample will be selected of which participants must have roughly the same computer experience. Cultural profiles of all participants will be established. Most likely, a questionnaire will be used to determine the cultural profile of a participant. The questionnaire will be composed of questions specifically designed to extract the values system of the respondent. Each question will be categorised as corresponding to one of the pre-selected cultural variables. Available questionnaires include the Value Survey Model (VSM) [21] developed and tested by Hofstede and the Value Orientations Method (VOM) developed and implemented by the Florence R. Kluckhohn Center for the Study of Values [20]. Cultural variables listed by Del Galdo and Nielsen [4] will be considered for inclusion.

A simple word processor will then be developed. The processor will have minimal functionality while still ensuring that there is a wide variety of user actions available, for example, menu selection, font and paragraph formatting and dialog box options.

Once the template for the word processor has been completed a selection of user interfaces will be developed. These interfaces will be culturally sensitive. To achieve the cultural sensitivity of interfaces the five main components of a user interface, namely [15] metaphors, mental models, navigation, interaction and appearance will be re-evaluated. The following subsets of these interface components will play a particularly important role in the sensitising process [19]:

- Words and icons
  - Language
  - Terminology
  - Fonts, font sizes and font styles (bold, italic and underline)
  - Icons and graphics
  - Use of colour and backgrounds
- Screen layout issues
  - Menu selection, form-filling and dialog box formats
  - Wording of prompts, feedback and error messages

Participants will be divided into groups. Groups can either be composed along cultural lines or irrespective of cultural affiliation. Either way, the division will allow participants to be tested on culturally sensitive as well as culturally insensitive interfaces. This means that some participants will be tested on an interface developed specifically for their culture group whilst others will be tested on an interface developed for a cultural group other than their own.

The cultural profile to which the user belongs will be established on commencement of the test, where participants will be classified according to either African culture, for example Sotho, Xhosa and Zulu participants, or Western (English and Afrikaans speaking participants). A cultural profile of a society or group defines the values of the culture within the context of the selected cultural variables. For example, the Japanese are considered to be a high power distance, high uncertainty avoidance, collectivist, masculine and diffuse society [7]. Other user characteristics, which will also be captured on commencement of the test, are the age and gender of the user as well as proficiency with a word processor. The experiment will incorporate a number of tasks that must be completed by users of the word processor. These tasks will be displayed as part of the current application, for example along the bottom of the screen. Measurements will then be taken to establish the usability of the interface. These measurements will be captured per user, per task. Measurements include:

- Number of actions to complete the task, e.g. menu selections, icon selections, keystrokes.
- Time taken to complete the task
- Whether the task was completed successfully or not.

Statistical analysis will be performed on the measurements captured to determine which interface, if any, is the most suitable for each culture tested.

After usability testing participants will be requested to complete a questionnaire aimed at determining user satisfaction. The questionnaire will most likely consist of close-ended questions that must be ranked according to a scale. Open-ended questions can also be included. By administering the questionnaire it will afford the users a chance to voice their opinion of the application with regards to general usability and feelings experienced whilst using the system. These opinions can then be cross-referenced with the measures obtained during usability testing.

The proposed methodology is an adaptation of a 4-phase cross-cultural interface design strategy [13]:

1. Investigation
   a. Learn to understand the processes of the target culture.
   b. Identify cultural markers which may be included in the interface – a foraging study is an excellent way in which to identify cultural markers.
   c. Assess the user attitudes.
2. Translation
   a. Develop a cultural model to enable the developers to determine the similarities and differences present in the targeted cultures.
3. Implementation
   a. Develop a localised application, being sure to employ the culture model established in phase 2 as well as the information obtained in phase 1.
   b. Implement a prototype and perform usability testing with targeted cultures.
4. Evaluation
   a. Modify the application based on the results of the usability testing.

SUMMARY

As South Africa moves forward to a more unified nation, it is imperative that computer applications start to reflect the cultural diversity present in the country. This proposed research project aims to investigate the need for a cultural interface that extends beyond the traditional translated interface. By making use of cultural models and exhaustive usability testing, the most effective culturally sensitive interface can be discovered and recommended for use with a word processor application.

REFERENCES

1. Badre, A. 2000. The Effects of Cross Cultural Interface Design Orientation on World Wide Web User Performance. Retrieved 25 February 2005 from ftp://ftp.cc.gatech.edu/pub/gvu/tr/2001/01-03.pdf.
2. Cyr, D. and Trevor-Smith, H. Localization of Web Design: An Empirical Comparison of German, Japanese, and U.S. Website Characteristics. Retrieved 24 February 2005 from www.eloyalty.ca/docs/ Localization_of_ Web_Design.pdf
3. De Wet, L. Blignaut, P. and Burger, A. Comprehension and Usability Variances among Multicultural Web Users in South Africa. CHI 2002. Minneapolis. April 2002.
4. Del Galdo, E. and Nielsen, J (Eds). 1996. International User Interfaces. New York: John Wiley and Sons Inc.
5. Esselink, B. 1998. A Practical Guide to Software Localization. Amsterdam, The Netherlands: John Bejamins Publishing Co.
6. Evers, V. 2002. Cross-Cultural Applicability of User Evaluation Methods: A Case Study amongst Japanese, North-American, English and Dutch Users in Proceedings of CHI ’2002. Minneapolis. April 2002.
7. Evers, V. 2001. Cultural Aspects of User Interface Understanding: An Empirical Evaluation of an E-Learning Website by International User Groups. PhD Thesis. Institute of Educational Technology, The Open University.
8. Evers, V. 1997. Human-Computer Interfaces: Designing for Culture. Thesis submitted to University of Amsterdam.
9. Greenwood, T.G. International Cultural Differences in Software. Retrieved 24 February 2005 from http:// research.compaq.com/wrl/DECarchives/DTJ/DTJB01 /DTJB01SC.TXT.
10. Hall, B. 2002. Among Cultures: The Challenge of Communication. Florida: Harcount, Inc.
11. Hars, A. Localizing Software is not just a multilingual issue; it’s also multicultural. BYTE. March 1996.
12. Hofstede, G. 1997. Cultures and Organizations: Software of the Mind. McGraw-Hill Companies, Inc.
13. Jagne, J., Smith, S.G., Duncker, E. and Curzon, P. 2004. Cross-cultural Interface Design Strategy. Technical Report. Interaction design Centre. Middlesex University.
14. Johns, S.M. Colors, Buttons, Words and Culture: Designing Software for the Global Community. Retrieved 24 February 2005 from http://www.sensi.org/ ~alec/locale/other/colore~1.htm.
15. Marcus, A. 2001. Cross-Cultural User-Interface Design.
16. Nielsen, J. 1999. Designing Web Usability. United States of America: New Riders Publishing.

17. Shen, J. 2000. User Interface Internalization. Retrieved 12 February 2005 from http://eies.njit.edu/~turoff/ coursenotes/CIS732/samplepro/user_interface_internationalizat.htm.

18. Sheppard, C. and Scholtz, J. 1999. The Effects of Cultural Markers on Web Site Use. Retrieved 15 February 2005 from http://zing.ncsl.nist.gov/hfweb/proceedings/ sheppard.

19. Shneiderman, B. 1998. Designing the User Interface. United States of America: Addison-Wesley.

20. The Value Orientations Method. Retrieved 18 March 2005 from http://members.aol.com/_ht_a/frkvalues/ method.htm.

21. Value Survey Model. Retrieved 12 February 2005 from http://feweb.uvt.nl/center/hofstede/english.html.

22. Van Belle, J.P., Fellstad, R., Steele, C. and van Bakel, W. Multi-cultural Websites in a Multi-cultural Country: a South African Perspective. Retrieved 12 February 2005 from http://www.commerce.uct.ac.za/ InformationSystems/Staff/PersonalPages/jvbelle/pubs/ f-VanBelleJeanPaul1.pdf.

23. Vöhringer-Kuhnt, T. The Influence of Culture on Usability. Retrieved 15 March 2005 from http://userpage. fu-berlin.de/~kuhnt/thesis/results.pdf

A Word Processor Prototype to Assess the Effect of Culture on Usability

Tanya Beelders **Pieter Blignaut** **Theo McDonald**
University of the Free State
Department of Computer Science and Informatics
P.O. Box 339, Bloemfontein, South Africa, 9300
beelderstr.sci@mail.uovs.ac.za / pieterb.sci@mail.uovs.ac.za / theo.sci@mail.uovs.ac.za

Abstract—This paper discusses a proposed research topic in the field of Human Computer Interaction (HCI), specifically investigating the usability of a number of culturally adapted user interfaces by means of a word processor prototype. A short discussion of culture and usability will serve as background to the problem.

Index Terms—Culture, usability, word processor

I. Introduction

Although the diverse cultures of South Africa are guaranteed protection and equal rights under the South African constitution, very little of this cultural diversity is evident in the technology field. Evidence of this is found in that the majority of South African websites are only available in English, despite the fact that English is not the predominate language of South Africa [8].

The greatest potential culture clash of a software package is the user interface as it serves as the point of contact between the user and the computer [6]. This is why localisation plays an important part in software design, where localisation refers to adapting the interface for a country in such a way that it is specific for that locale [7].

Users show marked preferences for interface components such as language, navigation, symbols and colour use [1], indicating that care should be taken when using icons, menu spacing, and date and time conventions [5].

II. CULTURE AND USABILITY

Culture is defined as “the collective programming of the mind which distinguishes the members of one group or category of people from one another” [3]. Cultural models, considered to be the most effective method to test cultural sensitivity of a user interface, consist of cultural variables which are useful in building the cultural profile of a group of culturally similar individuals [2]. A cultural profile of a society or group defines the values of the culture within the context of the selected cultural variables.

According to the International Standards Organisation (ISO) standard 9241-11 usability is “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [4].

III. WORD PROCESSOR PROTOTYPE

Cultural profiles will be established for potential participating cultural groups using a selection of the available cultural models. This will allow participants to be classified according to a particular cultural affiliation where two cultures will be identified, namely an African and a Western culture. Test participants will be classified according to their home language where the African culture will comprise of, for example, Sotho, Xhosa and Zulu speaking participants, while the Western culture will be the English and Afrikaans speaking participants.

A simple word processor application has been developed which incorporates minimal functionality while still ensuring that it is fully representative of a commercially accepted word processor. Functions available in the prototype include font and paragraph formatting, text editing and document handling.

The structure of the word processor will allow the user to incorporate the use of many common interface components such as:

- Menu selection;
- Toolbar shortcut selection;
- Dialog boxes, such as font formatting;
- Tab sheets which include components such as checkboxes and radio buttons.

Using the already established word processor as a basis, a selection of localised interfaces will be developed and overlaid on the functionality of the word processor. To achieve this localisation the following interface components, amongst others, will be reviewed:

- Language, colour and menu layout;

- Icons and metaphors.

Information which will be required from the user on commencement of their first test will be their home language, for classification into a culture as well as their gender and identity number, from which their age will be determined, and finally an indication of their proficiency, in terms of both computer and word processor use. This proficiency will be based on how long they have been using a computer and word processor as well as how often they make use of these tools. Another possibility which exists is to test the proficiency of the participants beforehand using established testing methods rather than to rely on the word of the user.

Some participants will be tested on interfaces developed specifically for their culture group whilst others will be tested on an interface developed for a cultural group other than their own. This will allow researchers to determine which interface is most suited to the expectations of each culture and whether localisation does in fact play a role in the perceived usability of a word processor.

The test will consist of a number of small tasks that are considered to be common to everyday usage of a word processor, for example opening a document, or changing the style of the font or paragraph.

As shown in Figure 1, each task will be displayed individually in the lower section of the application, while the user will complete the task in the top section of the word processor. Once a task is completed the next task will be displayed to the user.

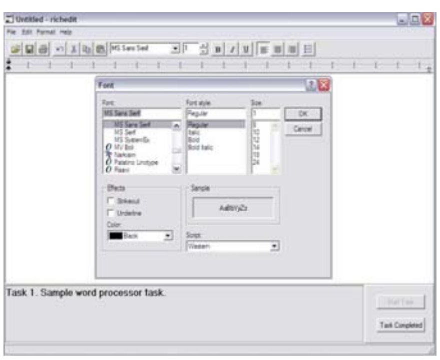


Figure 4: Prototype Word Processor

The measurements and actions that are captured by the application are as follows:

- Starting and ending time and date of the task.
- Time taken to complete the task.
- Total number of actions to complete the task divided into number of
  - Keystrokes;
  - Mouse clicks;
  - Toolbar shortcuts selected;
  - Menu options selected.
- Whether the task was completed successfully or not;
- The sequence of menu options selected;
- The sequence of toolbar shortcuts selected.

A second option which will be explored is to make use of a software package which captures all user actions. This package not only captures all the above-mentioned measurements but also stores a video of all activity on the computer for an allotted time. Audio and video of the user can also be captured thus allowing users to "think-aloud" without being directly observed by an outsider. The video can then be analysed for each user to determine precisely how each user completed each task and what difficulties were experienced by the user.

A learning curve will be established for each interface by giving users the opportunity to complete the test a number of times.

After completion of the test on the word processor participants will be requested to complete a usability questionnaire aimed at determining user satisfaction. The answers to this questionnaire will be analysed together with the captured usability measurements, thus allowing conclusions to be drawn as to which interface components complement each culture to the greatest extent – in other words which is the most usable interface for each culture.

IV. Conclusion

This research project intends to determine whether the need exists for a localised interface for a word processor application. By conducting usability testing on a word processor prototype with a variety of localised interfaces, the most effective localised interface for each culture group can be discovered and recommended for use with a word processor application.

REFERENCES

24. Cyr, D. and Trevor-Smith, H. Localization of Web Design: An Empirical Comparison of German, Japanese, and U.S. Website Characteristics. Retrieved 24 February 2005 from www.eloyalty.ca/docs/Localization_of_Web_Design.pdf
25. Del Galdo, E. and Nielsen, J (Eds). 1996. International User Interfaces. New York: John Wiley and Sons Inc.
26. Hofstede, G. 1997. Cultures and Organizations: Software of the Mind. McGraw-Hill Companies, Inc.

27. ISO9241. 1997. ISO 9241-11: Ergonomic Requirements for Office Work With Visual Display Terminals. Berlin: Beuth.

28. Johns, S.M. Colors, Buttons, Words and Culture: Designing Software for the Global Community. Retrieved 24 February 2005 from http://www.sensi.org/~alec/locale/other/colore~1.htm.

29. Marcus, A. 2001. Cross-Cultural User-Interface Design.

30. Nielsen, J. 1999. Designing Web Usability. United States of America: New Riders Publishing.

31. Van Belle, J.P., Fellstad, R., Steele, C. and van Bakel, W. Multi-cultural Websites in a Multi-cultural Country: a South African Perspective. Retrieved 12 February 2005 from http://www.commerce.uct.ac.za/InformationSystems/Staff/ PersonalPages/jvbelle/pubs/f-VanBelleJeanPaul1.pdf.

Biography: Tanya Beelders. Masters student in the Department of Computer Sciences and Informatics at the University of the Free State, Bloemfontein.

Measuring user performance for different interfaces using a word processor prototype

T.R. Beelders, P.J. Blignaut, T. McDonald, E.H. Dednam

Department of Computer Science and Informatics, University of the Free State, South Africa

Abstract

Usability tests were conducted in order to establish the effect on user performance of different icon sets in a word processor environment. Both an alternative and a textual set of icons were developed for a subset of word processor functions and compared to the standard icons. In order to accommodate users in their home language the interface of the word processor was available in English, Afrikaans and Sotho. This was done in an effort to determine whether usability of a product is increased when the users are allowed to interact with the product in their mother tongue rather than having to use the commonly available English interface. The scores obtained for completed tests as well as the time taken to complete tasks successfully were evaluated. Results indicate that neither icons nor language play a significant part in the usability of a product. In fact, the only significant contributor to user performance was the word processor expertise of the user.

Introduction

The advent of the graphical user interface (GUI) resulted in an increase in the use of icons within computer applications (Kacmar and Carey, 1991). More than 50% of the sales revenue for most computer companies is generated from markets outside of the country of development (Nielsen, 1996). Users have exhibited distinguishable preferences for interface components such as language, navigation, symbols and color

(Cyr, 2004). These facts motivate the need for careful consideration of translation and icon development, amongst other things, in user interfaces (Johns, 1997).

By means of empirical testing, this research first tested the usability of a set of alternative icons which were developed during a brainstorming session, together with those chosen by non-computer literate users. By comparing the alternative set of icons to the standard word processor icons (currently found in the Microsoft Office software package), it could be determined whether the alternatives are better suited to South African users, and it could also be established whether or not there is a need to develop new word processor icons for a South African audience.

Secondly, the research aimed to investigate how text-based icons and pictorial icons affect the usability of a word processor product. In this respect it reproduced a variation of the study undertaken by Benbasat and Todd (1993), with the notable exception that it was based in a South African context. With this in mind, the issue of translation was also tested in order to determine whether there is a need for translation of the interface into the first language of the user.

This paper will summarize some of the available literature on usability, icons and language processing. An outline of the research methodology that was used will be given, followed by a detailed discussion of the experiment results. Finally, a conclusion, based on the analysis of the results, will be drawn.

Usability

According to the International Standards Organisation (ISO) standard 9126-1 usability is "the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions". This definition is further expanded upon in ISO 9241-11 where usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO9241).

In terms of the above definitions, four distinct components of usability can be identified, namely effectiveness, efficiency, satisfaction and learnability. These components are defined as follows:

- Effectiveness is how well the user is able to achieve that which must be done by using the system (Abran, Khelfi, Surya, Suryn and Seffah, 2003) and can be measured in terms of accuracy and completeness (Cato, 2001).
- Efficiency refers to the resources which are required to complete the desired task (Abran et al, 2003).
- Satisfaction is the subjective feeling the user has about using the system (Abran et al, 2003).
- Learnability measures not only the time taken for a user to become familiarized with the system but also how well the user is able to remember system functionality (Cato, 2001).

Shneiderman (1998) lists a set of five measurable objectives that can be measured in order to determine the usability of a product. These measurements of time to learn, speed of performance, rate of errors by users, retention over time and subjective satisfaction allow for specific and controlled evaluation of a software application (Shneiderman, 1998). Several usability models are available that provide a number of measurements which can be used by developers to comprehensively test the usability of a product (Abran et al, 2003).

Icons

Icons are common interface components that employ images to represent an object or an action that can be carried out by the user (Benbasat and Todd, 1993). Continued use of icons has been attributed to the fact that they are easier for users to learn (Kacmar and Carey, 1991; Schild, Power and Karnaugh, 1980) and to use (Kacmar and Carey, 1991). Their use also increases the productivity of the user since recognition is generally faster for a picture than for text (Kacmar and Carey, 1991; Benbasat and Todd, 1993). Users' accuracy also improves with the inclusion of graphics in an interface (Muter and Mayson, 1986).

One disadvantage of icons is that they may be misinterpreted by users if the chosen image invokes unintended associations (Benbasat and Todd, 1993) – the picture that "speaks a thousand words may say a thousand different words to different viewers" (Zammit, 2000).

No visible advantages have been detected when using pictorial icons rather than a text based interface, but direct manipulation methods appear to increase the rate at which users complete tasks when compared with a strictly menu-based interface (Benbasat and Todd, 1993). Zammit (2000) investigated the effectiveness of pictorial and text icons with a group of 11 and 12-year-olds and found that neither the pictorial nor the text icons were always immediately recognizable to the users. The tooltips that appear below the icons as an expanded explanation for the icon also did not always serve to assist the user in determining what the icon represents (Zammit, 2000). User accuracy has been shown to be highest when selecting from a mixed format of text and graphics in a menu structure as opposed to a graphics only or text only menu structure (Kacmar and Carey, 1991). However, no discernable difference in the time taken to make a selection was detected between the three formats (Kacmar and Carey, 1991).

Any attention that is devoted to the interface detracts from the concentration of the user and constitutes interference with the primary task (Benbasat and Todd, 1993). Shneiderman (1998) notes that word processor applications generally make use of textual menus, whilst painting programs usually incorporate the use of icons, although nowadays most applications make use of both menus and icons. This distinctive design style is motivated by the fact that users are assisted by textual menus when concentrating on textual tasks and the more visual icons are better suited to a visually sensitive task, such as painting (Shneiderman, 1998). This prevents the user from having to switch between differing cognitive styles (Shneiderman, 1998). In contrast, it has been proposed that text and pictures require different amounts of inference in order to process the information (Benbasat and Todd, 1993). Since processing of text is considered to be a cognitive task, and the user is typically focused on some cognitive task when using a computer, more interference will be caused when using a text-based interface as it draws on the same cognitive resources as those required during completion of the task (Benbasat and Todd, 1993).

Language

In describing how individuals store and process language, the Revised Hierarchical Model (RHM) proposes that there is a two-level representation for language within the mind of a bilingual individual (Dufour and Kroll, 1995). The spoken languages are stored separately from one another on the lexical level but possess a common semantic system which spans all languages (Dufour and Kroll, 1995). The connections between the lexical and semantic are stronger for the first language (L1) than for the second language (L2) and even though increased proficiency of L2 results in a reduction of reliance on L1, there remains an unbalanced quality to these connections (Dufour and Kroll, 1995). This leads to the conclusion that semantic processing of an L2 message requires more cognition than that of an L1 message (Luna, Peracchio and De Juan, 2003). This suggestion has been supported by empirical testing which has found that processing of L1 stimuli is easier than processing of equivalent L2 stimuli (Sholl, Sankaranarayanan and Kroll, 1995). This would seem to support the call for localization of software applications, where the first and most important and, some would say, the only step of localization, is translation of the interface (Nielsen, 1999; Esselink, 1998).

The issue of translation into the home language of the user has, however, proven to be a fairly contentious one, with many researchers determining that translation increases usability of the product (Bodley, 1993), whilst others advocate caution when considering translation, as not all users show a preference for carrying out tasks in their mother tongue (De Wet, Blignaut and Burger, 2002; Blignaut and McDonald, 2006) and performance is often hampered by translation (De Wet et al, 2002). These studies did however focus on translation of web content which typically contains large amounts of text which must be read by the user, much more so than the single commands found in a word processing environment. Another consideration is that translation may also not be feasible due to technical or economic reasons (Kukulska-Hulme, 2000).

Users of an interface that is not in their first language do however encounter a number of inherent problems, one of which is verbal context – where surrounding words serve to place a word in context, allowing users to identify the actual meaning of the word rather than the potential meaning thereof (Kukulska-Hulme, 2000). Many interfaces do not include verbal context in menus, toolbars or buttons, which is clearly disadvantageous to second language users (Kukulska-Hulme, 2000) and could also lead to difficulty for novice or first time users who are unfamiliar with the domain terminology and concepts.

Methodology

By making use of a simple word processor application, the effect that different sets of icons and different languages have on the usability of a product was tested. Two sets of abstract pictorial icons were used in order to determine whether the icons influence product usability. A single set of textual word icons was also developed so that it could be established whether a difference exists between pictorial and text-based interface icons. The effect of interface language on product usability was also explored via implementation of the menus, tooltips and textual icons in English, Afrikaans and Sotho.

A small word processor application was developed in order to test the subjects. The word processor possessed minimal capabilities, while still ensuring that it was representative of a fully-fledged word processor or advanced text editor. Functions which were incorporated into the word processor prototype included document handling (e.g. open and close), text formatting (e.g. font size and style) and text manipulation (e.g. copy, cut and paste).

Users were required to complete a number of simple tasks, representative of common word processor tasks, such as font formatting. The tasks were displayed sequentially and individually at the bottom of the word processor window (Figure 1) and could be completed solely by making use of either a toolbar shortcut (icon) or a menu option.

The prototype allowed for real-time evaluation of the tasks, that is, once the user had completed a task the application immediately determined whether or not the task had been completed successfully. A number of measurements were also captured for each task. These included the number of menu options and toolbar shortcuts selected by the user, together with the number of keystrokes and mouse clicks and the time required to complete the task. The prototype also allowed for capturing of the users' demographic information.

Each task was assigned a difficulty index based on the number of actions or inferences which had to be carried out by the user in order to complete the task successfully. For example the task "Make the word 'One' bold" would consist of the following actions and inferences:

1. Understand the concept "bold"
2. Select the word "One"
3. Find the correct icon or menu option
4. Click the correct icon or menu option

This results in the above-mentioned task being assigned a difficulty index of 4. Tasks had difficulty indices ranging from 3 to 8.

Subjects

The test subjects consisted of first year university students who were taking a basic computer literacy course. Test subjects spoke a variety of languages, including

English, Afrikaans, Sotho, Tswana, Xhosa and Zulu. All subjects were conversant in either English or Afrikaans as these are the tuition languages of the university. The participants provided for different levels of word processor expertise. Of the participants 403 were female and 283 male.

Languages

As mentioned above there was a wide range of different languages spoken amongst the test subjects. Since the interface was only available in English, Afrikaans and Sotho, the participants were divided into one of these three groups according to their L1. Afrikaans users completed the test on either an Afrikaans (L1), or an English (L2) interface. Sotho users completed the test either in Sotho (L1), or in English (L2). For the purposes of this research Tswana speaking users completed the test either in Sotho (L1), or in English (L2), as Tswana is widely regarded as being very similar to Sotho and does in fact belong to the Sotho language group (Lye and Murray, 1980). The remainder of the users completed the test in English, where English would either be their L1 or L2.

Icon sets

Three sets of icons were used in the different interfaces, namely (i) the standard icons found in the Microsoft Office package, (ii) an alternative set of icons obtained from previous studies (Teklebrhan and Blignaut, 2005) and via two brainstorming sessions (see Figure 1), and (iii) text based icons.

The set of icons obtained during the first brainstorming session were distributed amongst potential word processor users. Respondents were required to indicate which icon they would choose for each of a number of listed word processor functions. Alternative icons for Open, Close, Save, Cut, Copy and Paste were determined in this way. Two icons were chosen by the same number of respondents for the Close function, so the icon for Close was selected by a process of elimination, as one of the icons was also chosen by a large margin as the preferred Save function.

The remainder of the icons were developed during a second brainstorming session, and these were included in the design without confirmation by non-computer literate users. The icons were developed to provide more context for novice users in the hope that these users would easily be able to relate to the concepts depicted by the icons. For example, the icons used for Bold, Italic and Underline consisted of a bold, italic or underlined capital letter “F” respectively. This was done in an effort to convey to the user the font changes that would occur if the function were invoked. By using the same letter throughout and by placing them adjacent to one another on the toolbar, easier visualization of the font styling (Figure 2) was ensured.

The textual word icons had no images; instead they displayed the name of the function they represented and were available in English, Afrikaans and Sotho.

Menus and tooltips

The menu structure, when available in the interface, was the same as the standard menu found in the Microsoft Office package. The toolbar situated at the top of the screen was divided into the standard and formatting toolbars as found in the Office package.

To enable the effect of the icons to be tested without interference from other interface components, each pictorial icon set was included in an interface with neither menus nor tooltips. This ensured that the user had to rely entirely on interpretation of the icon when using this interface. The next group of interfaces to be added to the test interfaces used the same pictorial icon sets but tooltips were added in English, Afrikaans or Sotho. To complete the set of interfaces for testing, the afore-mentioned interfaces were used as a base to which a menu was added in the same language as the tooltips for that particular interface.

The interface using the text-based icons had no menu although the tooltips were still used. This was to compensate for the fact that very often the entire function name did not fit on the button; in particular the Sotho translations were too long. To ensure legibility of the icons, a shortened version of the function was placed on the button and the full-length version was displayed in the tooltip to provide verbal context (Figure 3).

Interfaces

Taking all of the above-mentioned considerations into account, there were seven possible interface configurations:

1. Standard icons with a menu and tooltips
2. Standard icons with tooltips but no menu
3. Standard icons with neither a menu nor tooltips
4. Alternative icons with a menu and tooltips
5. Alternative icons with tooltips but no menu
6. Alternative icons with neither a menu nor tooltips

7. Text icons with tooltips but no menu

While interfaces 3 and 6 have no language component to speak of, since they have neither a menu nor tooltips, the remainder of the interfaces were available in either the users' L1 or L2, resulting in a total of 12 different interfaces.

Testing environment

The test was conducted during the first practical session of the course. This was before the users had received any instruction in word processor packages. Up to that point they had only been taught basic Windows concepts.

The application was loaded onto all the computers in the laboratory prior to the practical sessions. Each application defaulted to one of the available interfaces in such a manner that no two adjacent computers had identical interfaces. Users were assigned to the interface groups at random while ensuring that the number of L1 and L2 users for each interface was balanced.

After conclusion of all the practical sessions there were 686 tests that had been completed that met the requirements for the language division.

Analysis

The two interfaces without a language component (3 and 6) were removed from the initial analysis to be evaluated separately from the remainder of the interfaces. The subjects who participated in the study and completed the test on the interfaces that contained a language component are designated as group A, and the rest of the subjects are categorized as group B.

The independent variables used in the analysis were the interface employed during the test, the word processor expertise of the user and for group A, whether or not the user completed the test in their L1 or in L2 (English for all users).

Each user was classified as being a first time, intermediate or an expert word processor user based on their level of experience with a word processor application and the frequency with which they had made use of such an application prior to the test. The level of frequency and experience were rated on a scale of 0 to 4 and 0 to 5 respectively. These individual ratings were then cross-multiplied to obtain a scale consisting of fourteen distinct expertise ratings. The cut-off points between the user groups were chosen in such a way as to give the best distribution between the users.

In order to eliminate the effects of an individual's uncertainty regarding expertise, the intermediate group was not included in the analysis of the results. The final distribution of users is shown in Table 1.

The usability measures that were analyzed were (a) an effectiveness measurement of overall score (discussed below) and (b) an efficiency measurement of time taken to complete a task.

Each user was assigned a weighted score which was calculated as the sum of the cognitive loads of all the tasks completed successfully by that user. The time taken to complete each task was measured in seconds and then converted to 1/time for further analysis.

Analysis of group A

Group A consisted of those users who used any one of the interfaces 1, 2, 4, 5 or 7. Furthermore, these interfaces could either be in the L1 or the L2 of the user.

The weighted score for members of group A was therefore analyzed using a 5 (Interface) × 2 (Expertise) × 2 (Language) factorial ANOVA. The following hypotheses were formulated:

1. $H_{0,1}$: The word processor expertise of the user has no effect on the test score.
2. $H_{0,2}$: The interface used has no effect on the test score.
3. $H_{0,3}$: Whether the test is completed in the user's L1 or in L2 has no effect on the test score.

$H_{0,1}$ was rejected ($F_{Expertise}(1, 425) = 27.73$, $p = 0 < 0.001$), indicating that the word processor expertise of the user did indeed have an effect on the score achieved by the user. Neither $H_{0,2}$ ($F_{Interface}(4, 425) = 1.10$, $p = 0.356 > 0.05$) nor $H_{0,3}$ ($F_{Language}(1, 425) = 0$, $p = 0.947 > 0.05$) could be rejected, leading to the conclusion that neither the interface nor the language had an effect on the score achieved by the user.

A 5 (Interface) × 2 (Expertise) × 2 (Language) factorial ANOVA was performed for each task to evaluate the time taken to complete that task. Following the example of previous studies (Nielsen, 2006), only those tasks that were completed successfully were included in the analysis. The following hypotheses could be formulated for each task:

1. $H_{0,1}$: The word processor expertise of the user has no effect on the time taken to complete the task.
2. $H_{0,2}$: The interface used has no effect on the time taken to complete the task.
3. $H_{0,3}$: Whether the test is conducted in the user's L1 or L2 has no effect on the time taken to complete the task.

The results of the ANOVA are summarized in Table 2.

As would be expected, expert users performed significantly better than first time users with hypothesis 1 being rejected for all of the tasks. An α level of 0.05 was used throughout to distinguish between significant and non-significant differences.

Hypothesis 2 could be rejected for the tasks that require a single word to be made bold ($p = 0.002$, task 2) and a phrase to be italicized ($p = 0.041$, task 9).

Hypothesis 3 could not be rejected for any of the tasks at an α level of 0.05, therefore it could be concluded that the interface language had no effect on the time needed by the user to complete the task successfully.

As discussed previously, the group B users were removed from the analysis since the interfaces had neither a menu nor tooltips, thus containing no language component. Users of this group had to rely entirely on interpretation of the icon and as such could not be included in the initial analysis. Since it was shown that language had no effect on either the score achieved by the user or the time taken to complete the tasks, the need to separate the groups no longer existed and the two groups could be consolidated into a single group for the remaining analysis.

Analysis of the consolidated group

Groups A and B were amalgamated into a single user group in which language no longer played a role. The analysis of this group included users of all seven interfaces listed previously.

The weighted score for this user group was therefore analyzed using a 7 (Interface) × 2 (Expertise) factorial ANOVA. The following hypotheses were formulated:

1. $H_{0,1}$: The word processor expertise of the user has no effect on the test score.
2. $H_{0,2}$: The interface used has no effect on the test score.

$H_{0,1}$ was rejected ($F_{Expertise}(1, 503) = 26.47$, $p = 0 < 0.001$) indicating that the word processor expertise of the user does indeed have an effect on the score achieved by the user. The interface used had no effect on the score achieved by the users, so $H_{0,2}$ ($F_{Interface}(6, 503) = 2.01$, $p = 0.063 > 0.05$) could not be rejected.

For each task, a 7 (Interface) × 2 (Expertise) factorial ANOVA was performed to evaluate the time taken to complete that task. Once again, only those tasks that were completed successfully were included in the analysis. The following hypotheses could be formulated for each task:

1. $H_{0,1}$: The word processor expertise of the user has no effect on the time taken to complete the task.
2. $H_{0,2}$: The interface used has no effect on the time taken to complete the task.

The results of the ANOVA are summarized in Table 3.

Once again, hypothesis 1 could be rejected for all of the completed tasks since expert users performed significantly better than first time users in all of these tasks. Hypothesis 2 could be rejected for three of the eleven tasks. Possible reasons for these observations are discussed below.

- <u>Task 2</u>: Change font style of a single word to bold ($p = 0.001$)

A significant difference was found between the users of interface 2 and interface 7 as well as between users of interface 2 and interface 6. This indicates that the icons contribute significantly to the performance of the user. In both cases, users of the standard icon had a shorter completion time than users of the other two interfaces. This indicates that the standard icon for Bold is extremely intuitive and succeeds in conveying the concept of bold to the user, even more so than does the word "Bold" on the button. Since only those users of the alternative icons where no tooltips were used showed a significantly longer completion time than those of the standard icons, it would seem that the tooltips assisted the users in deciphering the functions linked to the icons for the remainder of the alternative icon users.

- Task 6: Close a document ($p = 0.001$)

With an average completion time of 53 seconds, users of the alternative icons with no menu and no tooltips (interface 6) took longer to close a document than all other users, who completed the task in times ranging from 18 seconds to just marginally longer than 20 seconds. The number of correct answers to this task was also the second lowest of all the tasks. Although the alternative icon for the Close function was chosen by questionnaire respondents, the results of this task show that it does not successfully communicate the concept of Close when used in an interface without any tooltips or menus to assist the user. In fact, the icon chosen by the respondents was actually designed as an alternative for an electronic mail interface. The icon appears to be acceptable when used in conjunction with a tooltip which leads the user to the correct icon required for the task at hand. Taking into consideration that choices of non-computer literate users were split evenly between the icon eventually used for Save and the one used for Close, it may be pointed out that perhaps the entire concept of closing a document needs to be explained more clearly to novice or first time users.

- Task 9: Italicize a phrase ($p = 0.006$)

Post-hoc tests indicated that the most significant difference occurred between users of interface 6 and interface 2 as well as between users of interface 6 and interface 4. Alternative icon users with no tooltips and no menu (interface 6) had a significantly longer average completion time than the users of the other two mentioned interfaces. These results indicate that the alternative italic icon does not succeed in conveying the function to the user. However, once again, inclusion of a tooltip to indicate the icon purpose assisted the users in determining the functionality linked to the icon.

Discussion

Overall, the most significant contributing factor to the performance of the users is that of word processor expertise. The interface used appears to have minimal effect on user performance. The only difference between the pictorial and text icons occurred in the task which required users to bold a single word. Thus, there is very little performance difference between users of pictorial icons and those using text icons, a finding which supports those of Benbasat and Todd (1993). The majority of performance differences that were detected existed between users of the alternatively designed icons where tooltips were absent from the interface and one of the other interfaces. The attempt to place the set of styling icons in a concrete context by using the same lettering and simply changing the styling effect had mixed results. Users of these icons with no tooltips showed a remarkably slower completion rate than users of other interfaces. The icons did, however, not seem to impede user performance when used in combination with tooltips or a menu. These results indicate that the actual icons that are used have little effect on the usability of the product, but that tooltips do in fact assist the user in determining the purpose of the icon. The only other icon that was unsuccessfully implemented was that of Close. Even though this icon was chosen as the preferred icon by non-computer literate users, it did not succeed in conveying the function concept to the user. This finding motivates the need for usability testing of interfaces even in the case where the interface is designed with the assistance of end users, since preferred interface choices do not always increase the proficiency of the users. Apart from these three icons, the other alternative icons did not hamper the proficiency of the user in any way.

Whether users work in their home language or not also has no effect on their productivity, contrary to what would be expected based on the unequal cognitive loads of L1 and L2 processing (Luna et al, 2003). These findings show that although users may not prefer to work on an interface in their L1 (De Wet et al, 2002; Blignaut and McDonald, 2006), a translated interface does not hamper their performance. In so doing the findings fail to corroborate the assertions that translation does increase user productivity (Bodley, 1993) and that translation may adversely affect user performance (De Wet et al, 2002). The failure to confirm the results of previous studies may be attributed to the fact that the mentioned studies tested user performance on a translated website which contained large amounts of text (De Wet et al, 2002), as opposed to single words or short phrases such as those used in this study. Also, where appropriate, word processing commands were placed in context, for example, the Sotho for Close and Open were translated as “close document” and “open document” respectively.

Conclusion

All indications are that user performance is not adversely affected by different interfaces, be they textual or pictorial, or different languages. Rather it is the experience of the user which dictates the effectiveness and efficiency of user performance. Differences between users of the standard and alternative interfaces were minimal, indicating that there is no need for the development of an alternative interface. From these results it appears evident that once the user has been provided with enough training and has gained enough experience to be confident within the task and the application domain, they will easily adapt to changes in the interface.

References

Abran, A., Khelfi, A., Surya, W., Suryn, W. and Seffah, A. (2003). Usability meanings and interpretations in ISO standards. *Software Quality Journal*, 11:325-338.

Benbasat, I. and Todd, P. (1993). An experimental investigation of interface design alternatives: icon vs. text and direct manipulation vs. menus. *International Journal of Man-Machine Studies*, 38:369-402.

Blignaut, P.J. and McDonald, T. (2006). Die implikasies van lees- en skryfvoorkeurtaal met betrekking tot Internettoegang vir 'n veeltalige Suid-Afrika. *S.A. Tydskrif vir Natuurwetenskap en Tegnologie*, September 2006.

Bodley, G.J.H. (1993). *Design of computer user interfaces for Third World users*. M.Com. Dissertation, University of Port Elizabeth, Port Elizabeth, South Africa.

Cato, J. (2001). *User-Centered Web Design*. Addison-Wesley, Great Britain.

Cyr, D. and Trevor-Smith, H. (2004). Localization of Web Design: An Empirical Comparison of German, Japanese, and U.S. Website Characteristics. *Journal of the American Society for Information Science and Technology*, 55(13):1-10.

De Wet, L., Blignaut, P. and Burger, A. (2002). Comprehension and usability variances among multicultural web users in South Africa. *Proceedings of CHI 2002*, Minneapolis, April 2002.

Dufour, R. and Kroll, J. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation, *Memory and Cognition*, 23:166-180.

Esselink, B. (1998). *A Practical Guide to Software Localization*. Amsterdam, John Benjamins Publishing Co, The Netherlands.

ISO9241. (1997). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals*. Beuth, Berlin.

Johns, S.M. (1997). Colors, buttons, words and culture: Designing software for the global community. *1997 CODI Conference*, April 9-11, Mesa, AZ. Retrieved 24 February 2005 from http://www.sensi.org/~alec/locale/ other/colore~1.htm.

Kacmar, C.J. and Carey, J.M. (1991). Assessing the usability of icons in user interfaces, *Behaviour and Information Technology*, 10(6):443-457.

Kukulska-Hulme, A. (2000). Communication with users: insights from second language acquisition, *Interacting with Computers*, 12:587-599.

Luna, D., Peracchio, L.A. and De Juan, M.D. (2003). The impact of language and congruity on persuasion in multicultural E-marketing, *Journal of Consumer Psychology*, 13(1&2):41-50.

Lye, W.F. and Murray, C. (1980). *Transformations on the Highveld: The Tswana and Southern Sotho*. Creda Press, Cape Town, South Africa.

Muter, P. and Mayson, C. (1986). The role of graphics in item selection from menus. *Behaviour and Information Technology*, 5(1):89-95.

Nielsen, J. (1996). *International Web Usability*, Alertbox, August 1996. http://www.useit.com/ alertbox/9608.html.

Nielsen, J. (1999). *Designing Web Usability*. Indianapolis: New Riders Publishing.

Nielsen, J. (2006). *Outliers and luck in user performance*, Alertbox, 6 March 2006. http://www.useit.com/alertbox/outlier_performance.html.

Schild, W., Power, L.R. and Karnaugh, M. (1980). *Pictureworld: a concept for future office system*. IBM, Thomas J. Watson Research Center, Yorktown Heights, N.Y. RC 8384 (#36518).

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3rd Edition). Addison-Wesley, Reading, Massachusetts.

Sholl, A. Sankaranarayanan, A. and Kroll, J. 1995. Transfer between picture naming and translation: Test of asymmetries in bilingual memory. *Psychological Science*, 6:45-49.

Teklebrhan, R. and Blignaut, P. (2005). *A study on the effect of Western designed metaphors in some culture groups in South Africa*. Technical Report, 2005/02, University of the Free State, Bloemfontein, South Africa.

Zammit, K. (2000). Computer icons: a picture says a thousand words or does it? *Journal of Educational Computing Research*, 23(2):217-231

Table 1: User distribution

User Group	Interface		Language	First time	Expert	Group Total
Group A	1	Standard icons, menu, tooltips	L1	24	20	
			L2	26	23	
	2	Standard icons, no menu, tooltips	L1	22	26	
			L2	24	13	
	4	Alternative icons, menu, tooltips	L1	13	17	
			L2	26	23	
	5	Alternative icons, no menu, tooltips	L1	21	23	
			L2	25	26	
	7	Text icons, no menu, tooltips	L1	15	20	
			L2	33	25	445
Group B	3	Standard icons, no menu, no tooltips		19	15	
	6	Alternative icons, no menu, no tooltips		17	21	72
Total				265	252	

Table 2: Group A: 1/Time ANOVA results

Task	n	Expertise	Interface	Language	Interface * Expertise	Interface * Language	Expertise * Language	Interface * Expertise * Language
1	364	0.004*	0.897	0.577	0.570	0.214	0.321	0.930
2	395	0.000*	0.002*	0.940	0.379	0.181	0.759	0.743
3	406	0.000*	0.989	0.743	0.134	0.344	0.392	0.080
4	419	0.000*	0.779	0.886	0.377	0.873	0.441	0.835
5	270	0.000*	0.621	0.867	0.541	0.926	0.166	0.156
6	360	0.000*	0.402	0.468	0.280	0.413	0.171	0.973
7	396	0.001*	0.147	0.697	0.894	0.017*	0.962	0.876
8	425	0.000*	0.225	0.802	0.880	0.201	0.418	0.831
9	424	0.000*	0.041*	0.624	0.510	0.268	0.527	0.673
10	405	0.002*	0.755	0.567	0.895	0.794	0.386	0.625
11	409	0.002*	0.848	0.579	0.801	0.558	0.088	0.475

* $p < 0.05$

Table 3: Consolidated user group: 1/Time ANOVA results

Task	n	Expertise	Interface	Interface * Expertise
1	418	0.022*	0.774	0.305
2	457	0.000*	0.001*	0.375
3	467	0.000*	0.845	0.087
4	486	0.000*	0.339	0.121
5	312	0.029*	0.781	0.133
6	396	0.002*	0.001*	0.279
7	457	0.000*	0.256	0.891
8	493	0.000*	0.281	0.830
9	492	0.000*	0.006*	0.318
10	464	0.002*	0.419	0.530
11	467	0.004*	0.120	0.555

* $p < 0.05$

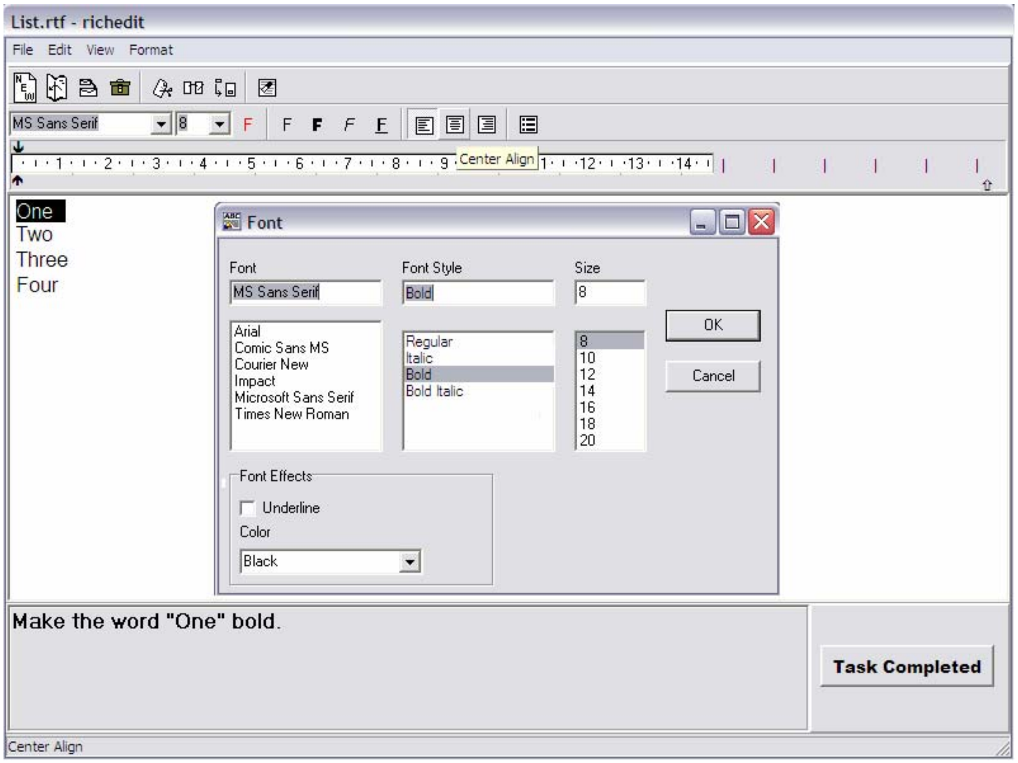


Figure 1: Word processor prototype with alternative icons and English menu

Figure 2: Font styling icons

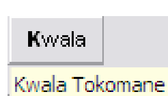


Figure 3: Sotho text icon for Close with expanded tooltip

The impact of different icon sets on the usability of a word processor

T.R. Beelders, P.J. Blignaut, T. McDonald, E. Dednam

Abstract—This paper discusses the results of usability tests obtained when testing different sets of icons in a word processor environment. An alternative set of icons was developed for a subset of word processor functions and compared to the standard icons. The score obtained for completed tasks as well as the time taken to complete tasks successfully were evaluated. Results indicate that the score is not affected by the icons used in the interface. It was noted that word processor expertise and the icons used have a significant effect on the time taken to complete some tasks. However, each of these factors exhibits an effect in only a single task completed in the prototype. Possible reasons for the significant difference are discussed.

Index Terms—Usability, icons, interface

INTRODUCTION

THE advent of the graphical user interface (GUI) resulted in an increase in the use of icons within computer applications [7]. Users have exhibited distinguishable preferences for interface components such as language, navigation, symbols and colour use [4]. These facts motivate the need for careful consideration of, amongst others, translation and icon development in user interfaces [6] – factors which could have an impact on product usability.

This paper will discuss the available literature on icons and usability. An outline of the research methodology that was used will be given, followed by a detailed discussion of the experiment results. Finally, a conclusion, based on the analysis of the results, will be drawn.

Usability

According to the International Standards Organisation (ISO) standard 9126-1 usability is "the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions". This definition is further expanded upon in ISO 9241-11 where usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [5].

Using the above definitions, four distinct components of usability can be identified, namely effectiveness, efficiency, satisfaction and learnability. These components are defined as follows:

- Effectiveness is how well the user is able to achieve that which must be done by using the system [5] and can be measured in terms of accuracy and completeness [3].
- Efficiency is the amount of resources required to complete the desired task [5].
- Satisfaction is the subjective feeling the user has about using the system [5].
- Learnability measures not only the time taken for a user to become familiarised with the system but also how well the user is able to remember system functionality [3].

Shneiderman [10] lists a set of five measurable objectives that can be measured in order to determine the usability of a product. These measurements of time to learn, speed of performance, rate of errors by users, retention over time and subjective satisfaction allow for specific and controlled evaluation of a software application [10]. The available usability models also provide a number of measurements which can be used by developers to comprehensively test the usability of a product [1].

Icons

Icons are a common interface component that employ images to represent an object or an action that can be carried out by the user [2]. Continued use of icons has been attributed to the fact that they are easier for users to learn [7], [9] and to use [7]. Their use also increases the productivity of the user since recognition is generally faster for a picture than for text [2], [7].

One disadvantage of icons is that they can easily be misinterpreted by users if the chosen image invokes unintended associations[2] – the picture that "speaks a thousand words may say a thousand different words to different viewers" [12].

Zammit [12] investigated the effectiveness of pictorial and text icons with a group of 11 and 12 year olds and found that neither the pictorial nor the text icons were always immediately recognisable to the users. Users' accuracy has been shown to be the highest when selecting from a mixed format of text and graphics in a menu structure as opposed to a graphics only or text only menu structure [7]. However, no discernable difference in the time taken to make a selection was detected between the three formats [7].

This research undertook to test the usability of a set of preferred icons chosen by non-computer literate users by means of empirical testing. To complete the set of chosen functions, the remainder of the icons were developed by means of a brainstorming session. By comparing the alternative set of icons to the standard word processor icons, it can be determined whether the alternatives are better suited to South African users and in so doing, establish whether or not there is a need to develop new word processor icons for a South African audience.

RESEARCH PROCEDURE

Method

By making use of a simple word processor application the effect of different sets of icons on the usability of a product was tested. Two sets of abstract pictorial icons [12] were used in order to determine if the icons used influence product usability.

Word processor prototype

A small word processor application was developed in order to test the subjects. The word processor possessed minimal capabilities, while still ensuring that it was representative of a fully-fledged word processor or advanced text editor. Functions which were incorporated into the word processor prototype included document handling (e.g. open and close), text formatting (e.g. font size and style) and text manipulation (e.g. copy, cut and paste).

The prototype also allowed for capturing of the users' demographic information, such as age, gender and language.

Users were required to complete a number of small tasks, representative of common word processor tasks. The tasks were displayed sequentially and individually at the bottom of the word processor window (Figure 1) and could be completed solely by making use of either a toolbar shortcut (icon) or a menu option.

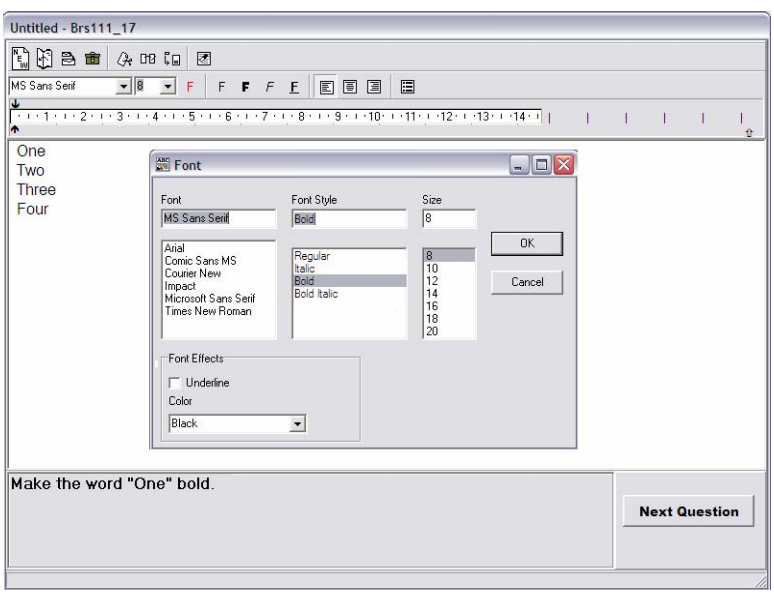


Figure 1: Word processor prototype with alternative icons

The prototype allowed for real-time evaluation of the tasks as the user completed each one. A number of measurements were captured for each task. These include the number of menu options and toolbar shortcuts selected by the user, together with the number of keystrokes and mouse clicks and the time required to complete the task.

Interface

Two sets of icons were used in the different interfaces, namely the standard icons found in the Microsoft Office packages and an alternative set of icons obtained from previous studies [11] and via two brainstorming sessions (see Figure 1).

The set of icons obtained during the first brainstorming session were distributed amongst potential word processor users. Given the complete set of icons, respondents were required to indicate which icon they would choose for each of a number of listed word processor functions. All the icons were available to be chosen for each function and icons could be chosen more than once. The alternative icons used in this study for the functions of Open, Close, Save, Cut, Copy and Paste, were chosen as the preferred icon by the group of non-computer literate respondents. There were two icons chosen by the same number of respondents for the Close function. Therefore, the icon for Close was selected by a process of elimination since one of the icons was also chosen as the preferred Save function by a large margin. A number of these icons were confirmed in the same manner in an independent study undertaken by Teklebrhan and Blignaut [11].

The remainder of the icons were developed during a second brainstorming session and included in the design without confirmation by non-computer literate users. The icons were developed to provide more context for novice and first time users. For example, the icons used for Bold, Italic and Underline consisted of a bold, italic or underlined capital letter "F" respectively. This was done in an effort to convey to the user the font changes that would occur if the function were invoked. By using the same letter throughout and placing them adjacent to one another on the toolbar, it allows for easier visualisation of the font styling (Figure 2). It was hoped that by developing in such a manner that novice and first time users would easily relate to the concepts depicted by the icons.

Figure 5: Font styling icons

To ensure that the effect of the icons could be tested without interference from other interface components, the interface had neither menus nor pop-up hints. The user had to rely entirely on interpretation of the icon.

In summary, the two interfaces tested were:

- Standard icons with no menu and no hints
- Alternative icons with no menu and no hints

Subjects

The test subjects consisted of first year university students that were taking a basic computer literacy course. Test subjects spoke a variety of languages, including English, Afrikaans, Sesotho and isiZulu. All subjects were conversant in either English or Afrikaans as these are the tuition languages of the university. The participants provided for different levels of word processor expertise. There were 61 female and 37 male participants who completed the test on either a standard interface or an alternative interface.

Testing Environment

The test was conducted during the first practical session of the course. This was before the users had received any instruction in word processor packages. Up until that point they had only been taught basic Windows usage.

Each participant was randomly assigned to one of the interface groups.

After all the practical sessions had been completed there were 98 tests that had been completed, of which 47 were completed on the interface with standard icons and 51 using the alternative icon interface.

ANALYSIS

The usability measures that were analysed were (a) an overall score (discussed below), (b) time taken to complete a task, (c) number of actions required to complete a task, (d) number of errors incurred whilst completing a task, (e) user satisfaction. The number of correct and incorrect answers was also compared for each task. Of these only the effectiveness measurement of score and the efficiency measurement of time [1] will be discussed in this paper.

Independent variables

The independent variables used in the analysis were the interface employed during the test and the word processor expertise of the user.

Each user was classified as either a first time or an expert [10] word processor user based on their level of experience with a word processor application, together with the frequency with which they make use of such an application.

Dependent variables

The two dependent variables discussed in this paper are the score for each user and the time taken to complete each task. To calculate the score each task was assigned a difficulty index based on the minimum number of actions and inferences required to complete the task successfully. This allowed for a weighted score to be computed for each user. The time taken to complete each task was measured in seconds and then converted to 1/time for further analysis.

Analysis of score

The evaluation of the score was done by means of a 2 x 2 between subjects factorial ANOVA. The following hypotheses were formulated for the score:

1. $H_{0,1}$: Word processor expertise has no effect on the score achieved .
2. $H_{0,2:}$ The interface used has no effect on the score achieved.

The word processor expertise of the user had no effect on the achieved score since $H_{0,1}$ could not be rejected ($F_{Expertise}(1, 94)= 0.989$, $p = 0.322$). $H_{0,2}$ could not be rejected ($F_{Interface}(1, 94)= 1.192$, $p = 0.278$), leading to the conclusion that the interface used during the test did not have a significant effect on the achieved score of the user.

Analysis of time

The time was evaluated individually for each task by means of a 2 x 2 between subjects factorial ANOVA. Only those tasks that were completed successfully were included in the analysis [8]. The following hypotheses were formulated for the time variable:

1. $H_{0,1}$: Word processor expertise has no effect on the time taken to complete a task successfully .
2. $H_{0,2:}$ The interface used has no effect on the time taken to complete a task successfully.

$H_{0,1}$ could be rejected for only one task ($F_{Expertise}(1, 81) = 4.302$, $p < 0.05$), where expert users performed significantly better than first time users. The task required users to change the font colour of a word. Two possible explanations could be offered for this difference. Firstly, it was observed during the test that many users experienced difficulty in grasping the concept that the drop-down box containing the font colour can be expanded to reveal a wider selection of colours. Secondly, after changing the font colour, the selection distorts the actual colour of the word. For example, green coloured font appears to be purple when selected. This phenomenon confused users not familiar with the effects that highlighting has on the font appearance. These two observations could possibly have caused some hesitation and confusion on the part of first time users, thus leading to a longer completion time for these users. Evaluation of the number of actions required and the number of errors incurred during completion of the task could provide more information on the cause of the difference.

A second task, which appears slightly later in the test, required users to change the colour of a whole sentence. There was no significant difference exhibited between the users for this second task. This seems to indicate that users retained the knowledge obtained in the previous task and did not experience the same problems again.

Users of the standard icons performed significantly better on the task that required users to close the text document ($F_{Interface}(1, 45) = 9.797$, $p < 0.05$), allowing $H_{0,2}$ to be rejected for that task. The alternative icon for the **Close** function was chosen by questionnaire respondents, but the results of this task show that it did not communicate the concept of **Close** as clearly as the standard icon. In fact, the icon chosen by the respondents was actually designed as an alternative for an electronic mail interface. Taking into consideration that choices of non-computer literate users were split evenly between the icon eventually used for **Save** and the one used for **Close**, it may be pointed out that perhaps the entire concept of closing a document needs to be explained more clearly to novice or first time users.

To place these icons in perspective, they are shown below in Table 1.

Table 2: Standard and alternative icons

	Standard	Alternative
Save		
Close		

The obtained result indicates that although users show a preference for a certain icon, it does not necessarily improve the usability of the product. Icons that are used should be chosen with care and developers should ensure that the icon does indeed convey the intended meaning or concept.

CONCLUSION

The interface had very little effect on the usability of the word processor, a finding which corroborates those of Kacmar and Carey [7] where time is concerned. The only significant difference between the users of the different interface occurred when using an icon that potentially did not convey the meaning of the function clearly to the user. This supports the assertion that careful consideration should be given to the development of icons [6]. The fact that the icon in question was chosen as the preferential icon by questionnaire respondents could indicate a distinct lack of understanding for the concept portrayed by the icon.

Word processor experience only had a significant effect on the task that required use of a complex dialog box, a situation in which it is understandable that a first time user would show some hesitancy or uncertainty. Subsequent tasks using the same dialog box showed no significant performance difference between users – an indication that users do retain the learned concepts, at least for a short period of time. It would be interesting to test whether users are able to retain this knowledge over a longer period of time than simply between two tasks.

Results would indicate that there is no need for development of an alternate set of icons for South African users. Standard icons appear to be intuitive enough that they correctly convey that which they attempt to represent. Rather, proper explanation of word processing concepts and functions is needed. Given enough time and practice, it appears that most users will be able to master the usage of a word processor application.

REFERENCES

1. Abran, A., Khelfi, A., Surya, W., Suryn, W. and Seffah, A., Usability meanings and interpretations in ISO standards. Software Quality Journal, vol. 11, pp 325–338, 2003.

2. Benbasat, I., Todd, P., An experimental investigation of interface design alternatives: icon vs. text and direct manipulation vs. menus, International Journal of Man-Machine Studies, vol. 38, pp 369–402, 1993.

3. Cato, J., User-Centered Web Design. Great Britain: Addison-Wesley, 2001.

4. Cyr, D. and Trevor-Smith, H. Localization of Web Design: An Empirical Comparison of German, Japanese, and U.S. Website Characteristics. Retrieved 24 February 2005 from www. eloyalty.ca/docs/Localization_of_Web_Design.pdf

5. ISO9241. ISO 9241-11: Ergonomic requirements for office work with visual display terminals. Berlin: Beuth, 1998.

6. Johns, S.M., Colors, Buttons, Words and Culture: Designing Software for the Global Community. Retrieved 24 February 2005 from http://www. sensi.org/~alec/locale/other/colore~1.htm.

7. Kacmar, C.J. and Carey, J.M., Assessing the usability of icons in user interfaces, Behaviour and Information Technology, vol. 10, no. 6, pp 443–457, 1991.

8. Nielsen, J., Outliers and luck in user performance, Alertbox, 6 March 2006. http://www.useit.com/ alertbox/outlier_performance.html, 2006.

9. Schild, W., Power, L.R. and Karnaugh, M., Pictureworld: a concept for future office system, IBM, Thomas J. Watson Research Center, Yorktown Heights, N.Y. RC 8384 (#36518), 1980.

10. Shneiderman, B. Designing the User Interface: Strategies for Effective Human-Computer Interaction. 3rd Edition. United States of America: Addison-Wesley, 1998

11. Teklebrhan, R and Blignaut, P. A study on the effect of Western designed metaphors in some culture groups in South Africa. Department of Computer Science and Informatics: University of the Free State, Technical Report 2005/02, 2005.

12. Zammit, K., Computer icons: a picture says a thousand words or does it?, Journal of Educational Computing Research, vol. 23, no. 2, pp 217–231, 2000.

Summary

The word processor or some form of editor-based application has become an integral tool for the many people who rely on computers on a daily basis. As such it has a wide and varied user base and must cater for a very diverse user group. Due to the heavy reliance on the word processor it is essential that it delivers pleasurable and efficient interaction to its users. Since its inception, the word processor has displayed the ability to evolve to continually exploit the increasing capabilities of technology. This study focused on furthering the improvement of the word processor usability for a subset of South African word processor users. Specifically, it concentrated on the impact of graphics, text and language on the usability of a word processor.

Graphics were incorporated into the interface by means of inclusion of the icons currently found in the Microsoft Office package, which have been accepted as the industry standard, and the development of an alternative set of icons whose usability could be compared to that of the standard icons.

Text was included in the interfaces in the form of menus and tooltips as well as text buttons which replaced the afore-mentioned pictorial icons and contained no graphical depiction of the associated function.

The impact of language on the usability of a word processor was viewed strictly in terms of bilingual users and was achieved through translation of the text buttons, menus and tooltips into the predominant languages of the area.

Comparative user testing was conducted through implementation of a scaled-down word processor application which could accommodate interchangeable interfaces and easy administration of preset tasks. Representative users were then required to complete a series of tasks on their respective pre-assigned interface, which conformed to one of the following general interface configurations:

a. An interface using either set of pictorial icons and excluding both menus and tooltips, thus containing no language component.

b. An interface in their first language, achieved through use of the text buttons, menus and/or tooltips.
c. An English interface, where English was not their first language.

A set of usability measures was identified which allowed for the effectiveness, efficiency and satisfaction of the users to be compared between the different user interface configurations. These measurements were:

a. the score achieved for the test, based on a built-in difficulty index assigned to each task;
b. the satisfaction experienced during interaction with the application; and
c. for each task, the:
    i. time,
    ii. number of actions,
    iii. number of errors and
    iv. ratio of correct and incorrect answers

Analysis of the user testing found that no particular interface configuration exhibited increased efficiency, effectiveness, learnability or satisfaction and that users were able to adapt to a changed interface with ease once they had become accustomed to the word processor environment. Therefore, the final finding of the study was that provision of an interface in a bilingual user's first language neither significantly contributed nor detracted from the application's usability. Similarly, neither of the pictorial icon sets nor the text buttons exhibited a significantly heightened level of usability. Therefore, none of the interface configurations could be recommended as the most usable.

However, a number of recommendations concerning the usability of a word processor were proposed based on both the analysis of the tasks and observation of user interaction. Finally, based on user performance for each individual task, an icon was identified which appeared to be the best and most applicable for that function. The final recommended interface, the usability of which must still be empirically established, consisted of a combination of standard icons, alternative icons and text buttons.

Keywords: Usability
User interface
Icons
Text
Bilingualism
Word processor

Opsomming

Die woordverwerker of enige soortgelyke program het ’n integrale stuk gereedskap geword in die lewe van baie mense wat rekenaars op ’n daaglikse basis gebruik. As sulks word dit deur ’n uiteenlopende en diverse groep gebruikers gebruik. Gevolglik is dit noodsaaklik dat die woordverwerker genotvolle en effektiewe interaksie aan gebruikers bied. Oor die jare het die woordverwerker die vermoë getoon om die groeiende tegnologiese vooruitgang tot voordeel te gebruik. Hierdie studie het op die verbetering van die bruikbaarheid van die woordverwerker vir ’n gedeelte van Suid-Afrikaanse woordverwerker-gebruikers gefokus. Sodoende het dit spesifiek op die impak wat grafika, teks en taal op die bruikbaarheid van woordverwerkers het, gekonsentreer.

Grafiese elemente is in die vorm van ikone in die koppelvlak ingesluit. Die ikone, soos dit tans in die Microsoft Office pakket gevind word, is ingesluit in die studie as gevolg van hulle aanvaarding as die standaard ikone vir gebruik in die industrie. ’n Alternatiewe stel ikone is deur middel van vraelyste en aanvaarde riglyne ontwerp. Die bruikbaarheid van die alternatiewe ikone kon dan met dié van die standaard ikone vergelyk word.

Teks is in die vorm van keuselyste en wenke (tooltips) in die koppelvlak ingesluit, sowel as knoppies wat die funksienaam, in plaas van ’n grafiese afbeelding van die funksie vertoon.

Die impak van taal op die bruikbaarheid van ’n woordverwerker is streng ten opsigte van tweetalige gebruikers bestudeer en is bereik deur die teksknoppies, keuselyste en wenke in die mees algemene tale van die area te vertaal.

Vergelykende gebruikerstoetse is deur middel van die implementasie van ’n afgeskaalde woordverwerkingsprogram gedoen. Die program kon die wisseling van koppelvlakke sowel as die maklike aflê van voorgeskrewe take hanteer.

Verteenwoordigende gebruikers het die take op een van die volgende algemene koppelvlakke voltooi:

a. ŉ Koppelvlak wat een van die twee stelle van prentjie-ikone gebruik en geen keuselyste of wenke gebruik nie.
b. ŉ Koppelvlak wat in hul moedertaal verskaf word deur gebruik te maak van die teksknoppies, keuselyste en/of die wenke.
c. ŉ Engelse koppelvlak, waar Engels nie die moedertaal van die gebruiker is nie.

ŉ Stel bruikbaarheidsmetings is geïdentifiseer waardeur die effektiwiteit, produktiwiteit en tevredenheid van die gebruikers van die verskillende koppelvlak-konfigurasies vergelyk kon word. Hierdie metings was:

a. die punt vir die toets, gebaseer op ŉ ingeboude moeilikheidsgraad wat aan elke taak toegeken is;
b. die tevredenheid ondervind tydens interaksie met die woordverwerker; en
c. vir elke taak:
    i. tyd,
    ii. aantal aksies,
    iii. aantal foute en
    iv. die verhouding tussen die aantal korrekte en verkeerde antwoorde.

Ontleding van die gebruikerstoetse het getoon dat nie een van die koppelvlak-konfigurasies groter effektiwiteit, produktiwiteit, leerbaarheid of tevredenheid verleen nie. Gebruikers kon ook maklik by ŉ veranderde koppelvlak aanpas indien hulle alreeds aan ŉ woordverwerkingsprogram gewoond was. Dus, die finale bevinding van die studie is dat die vertaling van die koppelvlak na die moedertaal van ŉ tweetalige gebruiker nie ŉ positiewe of ŉ negatiewe uitwerking op die bruikbaarheid van ŉ woordverwerker het nie. Net so, het nie een van die ikoonstelle beter bruikbaarheid tot gevolg gehad nie en sodoende kon nie een van die koppelvlakke as die mees bruikbare koppelvlak vir ŉ woordverwerker aanbeveel word nie.

Sekere aanbevelings, gebaseer op beide die ontleding van die take en die waarneming van die gebruikers-interaksie met die woordverwerker is egter oor die bruikbaarheid van ŉ woordverwerker voorgestel. Laastens, gebaseer op die ontleding van die gebruikers se werkverrigting vir elke taak, is die beste en mees aanvaarbare ikoon vir elke funksie geïdentifiseer. Die bruikbaarheid van die finale aanbevole koppelvlak, bestaande uit 'n kombinasie van standaard ikone, alternatiewe ikone en teksknoppies, moet nog empiries getoets word.

Sleutelwoorde: Bruikbaarheid
Koppelvlak
Ikone
Teks
Tweetaligheid
Woordverwerker