UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Towards Guidelines for Error Message Design in Digital Systems

by
Hein Pieterse
26227887

Submitted in fulfilment of the requirements for the degree
M IT (Information Systems)

in the

FACULTY OF ENGINEERING, BUILT ENVIRONMENT & IT

at the

UNIVERSITY OF PRETORIA

Study leader:
Prof. H Gelderblom

Date of submission
2016-04-30

Declaration regarding Plagiarism

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

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

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

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

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

I (full names & surname):	Hein Pieterse
Student number:	26227887

Declare the following:

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

Signature _____ Date _____

Contents

List of Figures

Towards Guidelines for Error Message Design in Digital Systems

Abstract

A part of many digital system is the display of error messages. The research aims to create a set of guidelines for error message design in digital systems. These guidelines will enable designers and developers to create better error messages that convey the right information at the right time and in the right way. In other words, to create error messages that are necessary and effective.

The first step in the process of generating this set of guidelines was to perform a literature review in order to find existing theory that is applicable to the design of error messages. The literature review also includes research on warning design theory. The assumption is that warnings are, to some extent, similar to error messages. Therefore, the research surrounding it is also expected to be applicable to error messages. The use of warning design theory stems from the lack of research on error messages and the comparable richness of the body of knowledge on warnings.

From this literature review it was possible to propose a set of guidelines for error message design. The initially proposed guidelines were evaluated by performing two usability studies on an existing Internet banking website. The first usability study involved a heuristic evaluation of some of the error messages in the website, using the guidelines as heuristics. The second usability study entailed conducting individual interviews with representative users where the same error messages used in the heuristic evaluation was evaluated.

The results of the heuristic evaluation were then used to determine whether the guidelines are effective. The effectiveness of a guideline is an indication of whether experts can easily use it to analyse error messages and detect possible usability problems. The results of the individual interviews were used to determine whether the proposed guidelines are valid. The validity of the guidelines is a measure of how well the guidelines and the suggestions raised by using them, reflect the pain points and concerns of users. The results of both these usability studies were also compared to one another to get a further indication of the effectiveness and validity of the guidelines.

From this analysis some changes and additions were made to the initially proposed guidelines. These updates are expected to increase the effectiveness and validity of the guidelines compared to the initially proposed guidelines. In other words, the updates are expected to

make the guidelines easier to use and to enable experts to find usability problems that are a closer match to the concerns of users.

The research followed the design science research methodology, but only completing one iteration of the process. Subsequent iterations that will further refine the proposed guidelines is left for future research.

Keywords: Error messages; warning messages; warning design; Internet banking; heuristic evaluation, design science research

1 Introduction

This dissertation includes all the details of research performed in fulfilment of the requirements of a master degree MIT (Information Systems). The research is concerned with the design of error messages in digital systems. Specifically, the research aims to create and validate a set of design guidelines that can be used to analyse existing error messages or to assist in the design of new error messages.

The sections below briefly discuss and frame the research in the larger body of knowledge; describes the overall problem that was addressed; lists the research questions answered; and describes the assumptions and the limitations of the research.

1.1 Background Information

The body of knowledge regarding *warning design* theory is extensive and since the mid 1980's, researchers have been investigating what constitutes an effective warning, how to design an effective warning, and when and why warnings work (Laughery 2006). The field is traditionally concerned with warnings in the physical world that include different components, such as: warning signs, printed messages on products, user manuals, product inserts and auditory warnings that include verbal warnings and alarms (Laughery 2006). Recently the research has been extended to also include warnings in digital systems. This research is mostly focused on security related warnings where the principles of traditional warning design theory is more directly applicable. Research shows that factors found in traditional warning design is also applicable to digital systems (Bravo-Lillo, Komanduri, Cranor, Reeder, Sleeper, Downs and Schechter 2013), which is expected because the principles of warning design is independent of the specific product or environment (Laughery 2006).

Different factors that contribute to the effectiveness of both physical and digital warnings, include include size (Barlow and Wogalter 1991, Young and Wogalter 1990), colour (Kline, Braun, Peterson and Silver 1993), location (Laughery, Young, Vaubel and Brelsford, Jr 1993, Wogalter, Godfrey, Fontenelle, Desaulniers, Rothstein and Laughery 1987, Barlow and Wogalter 1991), signal words (Adams, Bochner and Bilik 1998, Wogalter and Silver 1995), linguistics of the warning text (Harbach, Fahl, Yakovleva and Smith 2013, Duarte, Rebelo, Teles and Noriega 2012), and pictorials or icons (Davies, Haines, Norris and Wilson 1998, Jaynes and Boles 1990, Wogalter, Begley, Scancorelli and Brelsford 1997). Apart from these, research into warnings in digital systems also include factors specific to software systems including mode of display (Rainer and Stefan 2010, Maurer, Luca and Hussmann 2011, Krol, Moroz and Sasse 2012, Kelkar, Gadepalli and Indurkhya 2013), level of com-

puter expertise citeBravo-Lillo2011, and contextualization of warnings (Bartsch, Volkamer, Theuerling and Karayumak 2013).

Apart from the characteristics of the actual warning, the target audience also plays a role in the effectiveness of a warning. Factors that are present in the literature include habituation (Kim and Wogalter 2009), social influence and role modelling (DeTurck, Chih and Hsu 1999, Edworthy and Dale 2000, Wogalter, Allison and Mckenna 1989, Adams et al. 1998), how familiar a user is with the product or environment (Harrell 2003, Wogalter, Barlow and Murphy 1995), and the cost and benefits of compliance or non-compliance (Hunn and Dingus 1992).

A theoretical model proposed by Wogalter, DeJoy and Laughery (1999) called the communication human information processing (C-HIP) model, describes the different factors that contribute to warning effectiveness and how information flows and is processed by a human receiver. This model is described in more detail in section 2.2.1. According to the C-HIP model the factors of a target user that contribute to the effectiveness of a warning include: attention, comprehension, attitudes and beliefs, and motivation. From the C-HIP model it can be said that for a warning to be effective it must first be noticed and the information must be taken in (attention). Thereafter the content must be understood correctly (comprehension). Finally, the ability of the warning to affect the correct behaviour is dependent on the attitudes and beliefs as well as the user's motivation to comply.

In contrast to the large amount of detailed research on warning design, the design of error messages in digital systems have not received the same amount of attention. In this case error messages refer to messages displayed in a digital system that indicate that an error has occurred. Such errors are usually a result of an invalid user interaction or problems encountered while processing data. Some research on error message design is found in the literature and includes general guidelines for error and warning dialogs (Kelkar et al. 2013), and the display of input validation and error messages in web forms (Seckler, Heinz, Bargas-Avila, Opwis and Tuch 2014, Al-Saleh, Al-Wabil, Al-Attas, Al-Abdulkarim, Chaurasia and Alfaifi 2012, Bargas-Avila, Brenzikofer, Roth, Tuch, Opwis and Orsini 2010, Bargas-Avila, Oberholzer, Schmutz, de Vito and Opwis 2007). Guidelines for error messages are also included as part of general guidelines for the design of user interfaces (Molich and Nielsen 1990), but the research does not focus specifically on error messages.

The more general design guidelines for warning messages found in the literature can be applied to error messages as well. The effectiveness of an error message can be determined on the same grounds as warnings, in other words, whether the warning (or error message) draws and holds attention, whether information provided to the user is understood, and if all the information required is supplied to the user to result in the intended behavioural changes (Laughery 2006). The C-HIP model can also be applied to error messages because

the same steps are followed by the user when processing information from an error and a warning. The overall purpose of a warning is to make the world safer, provide information, influence behaviour, and remind people of risks (Laughery 2006). Similarly, it could be said that the purpose of error messages is also to make the world safer, provide information, and influence behaviour.

1.2 Purpose of the Study/Research Objectives

With the lack of research on the design of error messages in digital systems and the assumption that warning design has very similar concerns to error message design, it should be possible to combine existing warning design theory with existing error message design research into a set of design guidelines for error messages in digital systems. The objective of this study is, therefore, to create a set of guidelines for error message design.

Such a set of guidelines will typically be used in two ways: firstly, a set of guidelines will assist during the design and development of digital systems. If the designers and developers know what constitutes effective error messages, they are more likely to create effective error messages. Secondly, a set of guidelines could be used in the evaluation of user interfaces. One popular method that is used to evaluate user interfaces is a heuristic evaluation (Hollingsed and Novick 2007). In this evaluation method multiple experts evaluate a given user interface based on set of usability principles, sometimes referred to as heuristics (Nielsen 1992). A set of guidelines for error message design could then be used in a heuristic evaluation of the error messages in a user interface.

With such a set of guidelines that specify how to design error messages, and how to ensure that they are effective, it would be easier for designers and developers of digital systems to design error messages and to evaluate them. It is foreseen that if the guidelines are successful in informing the development process and assisting early evaluation of interface prototypes, the amount of user testing that has to be done to evaluate the error scenarios in a system will be reduced. Also, for smaller projects where the cost of user testing is comparably high (Nielsen and Molich 1990), some assurance can be given that the error messages in a system are effective and valid without conducting large scale user testing.

In order for a set guideline to be valuable it has to be tested to confirm their validity and effectiveness. This was done through two usability studies. The first usability study tested whether domain experts can successfully use the guidelines to find potential usability problems with existing systems through a heuristic evaluation. This indicates whether the guidelines are easy to understand and can be used to find usability problems with an error message. In other words, the first usability study tested whether the proposed guides are effec-

tive.

The proposed guidelines were also tested by running a usability study on the error messages of an existing system with representative users. This second usability study tested whether the pain points and concerns of users are addressed by the created design guidelines. This gives a measure of the validity of the guidelines. Since both these tests were performed on the same system, a comparison of the potential usability problems, identified by the experts, and the concerns of the representative users could be done. It is a general concern whether the problems identified through a heuristic evaluation constitutes problems for real users, and whether all problems can be identified (Nielsen and Molich 1990). With the comparison between the problems identified by the experts and the concerns of real users, a further indication of the validity of the guidelines themselves and the validity of expected results of a heuristic evaluation could be retrieved. This also increases the credibility of the overall results.

1.3 Problem Statement

The research into the design of effective warning messages is extensive and includes a wide variety of scenarios and applications, which includes warnings displayed in digital systems. Unfortunately, the same cannot be said for the design of error messages. Although some guidelines and research exists, it is usually not the primary focus of the research and the guidelines are only for a specific application. Thus, the problem that this research addresses is the lack of general guidelines for error message design in digital systems.

Additionally, in heuristic evaluations (using existing heuristics for successful user interface design), it is seen that the heuristics regarding error messages are the most difficult to apply and usability problems related to error messages are likely to prove especially costly (Nielsen and Molich 1990). With a set of guidelines that focuses specifically on error messages, a separate heuristic evaluation can be completed which focuses only on error messages.

1.4 Research Questions

The overarching question that the research aims to answer is: what are the general guidelines to consider when designing error messages in digital systems? Specific questions that are also addressed are:

- What are the primary concerns found in the literature regarding the design of warning and error messages?

- How do these concerns translate into guidelines for error messages design in digital systems?
- Are the guidelines effective, that is to say, how easy is it to understand the guidelines and use them to analyse error messages?
- Are the guidelines valid, in other words, do the design guidelines reflect the pain points and concerns users have regarding error messages in digital systems?
- What are the relative importance of the different guidelines?

1.5 Assumptions and Limitations

As indicated in the sections above, the research aims to create and test design guidelines for error messages in digital systems. The starting point of developing design guidelines for error messages is to use existing literature to inform the development of the guidelines. The literature used includes research on error message design and warning design theory. The assumption is that warning design theory is applicable to error message design. The reasoning behind this assumption is briefly discussed above in section 1.1 and also in the literature review in Chapter 2. This assumption is not directly tested, but since the guidelines are tested, it is indirectly validated.

The evaluation of the created design guidelines included a laboratory usability study in the form of individual interviews. This method has small sample sizes and ecological invalidity (Rainer and Stefan 2010). The ecological invalidity is the concept that people act differently in a laboratory study in comparison to a real life situation.

Coupled with this is the problematic nature of performing usability tests on error messages. In an ideal case, with the least amount of ecological invalidity, a user should be observed while an error message is generated as part of normal interactions with a system. However, because the research approach of this study includes usability studies on existing systems it would be difficult to get users to generate errors as they would in a real life scenario. This could be either because some errors are time dependent or the generation requires very specific interactions.

One example is an error message tested during the research where a user receives a message when they have been inactive for too long, in other words, their session has expired. Refer to section 5.2.2 for details on the error message. This is problematic because either the user will have to be told to not interact with the website or be distracted from their task for long enough. Another example is an error that is generated when a user has to specify a

password, but an error is received when the password is too weak. More details on this error is shown in section 5.2.10. For this error to be tested the user should either be given free rein to enter any password that they want and perform enough user tests so that a reasonable amount of the users encounters the error scenario, or the users should be instructed on exactly what password to enter.

To get around these types of obstacles in the testing of error messages, the usability tests were performed using static images of error messages produced in an existing digital system. This is expected to introduce more ecological invalidities in comparison to the testing scenarios given above, but it will greatly reduce the complexity and time to evaluate error messages. This means that more error messages can be tested with more users in the same amount of time when using static images in comparison to more complex testing scenarios. Furthermore, because the research focuses on the design guidelines and not on the error messages themselves, it is assumed that the data gathered using static images rather than a functioning system would be sufficient to draw conclusions about the design guidelines.

A further limitation of the research is the size of the usability study which includes only errors generated on one digital system, in this case an Internet banking website. As discussed in more detail in Chapter 2, the environment within which error messages are interacted is expected to have an effect on the effectiveness of the error messages. For example, in an environment with a high perceived risk (like an Internet banking website), people are expected to act differently towards error messages in comparison to an environment with lower perceived risk. Therefore, the results of the usability study, and the conclusions drawn regarding the design guidelines, cannot be said to be valid for all environments.

1.6 Brief Chapter Overview

The rest of this document includes more extensive details about the research. This includes the literature review in Chapter 2 which discusses the literature regarding error and warning design as well as the theoretical grounding of the research and a discussion surrounding usability testing.

The research follows a design science research methodology. This is discussed in Chapter 3 which includes a discussion on relevant research philosophies and paradigms and a discussion on design science research. Chapter 3 also includes details on the research design, data collection and analysis approach, and ethical considerations.

Chapters 4 through 6 include the initial design guidelines created from the literature review,

the analysis of the results of the usability tests, and the updated guidelines based on the results of the usability tests respectively.

2 Literature Review

This chapter discusses and reviews the literature of interest regarding the design of error and warning messages. The aim is to explore the body of knowledge relating to error message design. This review was used to inform a set of design guidelines for error message design, which is the object of study. The methodology of the study, and how the literature study forms part of it, will be discussed in Chapter 4. This chapter also includes a short review of popular usability testing techniques. This shows how a set of guidelines for error message design might be used in practice.

In the sections below, literature on both warning and error design is described, where the literature on warnings presents the bulk of research done. It is important to understand that the purpose of warnings and errors, and the context in which they are used, differ. Warnings are usually shown to a user before an action to give information regarding possible risk. However, errors are shown to a user after an unsuccessful action and it gives information regarding the reason that the action failed. Although errors and warnings are presented to a user for different reasons and at separate times, the theory and design considerations of warnings are close enough to errors that warning messages are applicable to the study. This is reflected in the design guidelines that were created from the literature and is discussed in more detail in Chapter 4. The assumption that warning design theory is applicable on error messages is not directly validated in the research, but the proposed guidelines based on the theory is validated.

The literature review is divided into two main sections. The first is a section regarding user behaviour, discusses the theoretical backing for error and warning design and why users exhibit specific behaviours. The second section entails warning and error design, which covers the more concrete design considerations and suggestions found in literature.

2.1 Introduction

Overall, warnings should be regarded as safety communications, and their purpose can be analysed on four levels, namely to: create a safer world, provide information, influence behaviour, and also to function as a reminder (Laughery 2006, Laughery and Wogalter 2006). In short, the goal and purpose of a warning is to firstly prevent injury, illness, or damage by delivering accurate and appropriate information when and where it is needed, and secondly to affect behaviour to achieve this intent (Laughery and Wogalter 2006). For the warning to be successful it must be purposefully designed for a specific situation, and if the warning is unsuccessful it can be responsible for undesired consequences (Duarte et al. 2012).

Thus, warnings should be considered as a last resort to protect people from risk. This is illustrated by the hazard control hierarchy, sometimes referred to as the safety hierarchy. Before a warning is deemed necessary and appropriate, it should first be determined whether it is not feasible to firstly design out the risk, or secondly to guard against it (Laughery and Wogalter 2006, Laughery 2006, Bravo-Lillo, Cranor, Downs and Komanduri 2011). Only when it is not possible to design out or guard against risk is it appropriate to use a warning. This is also true for error messages where error prevention is considered more important than effective error messages (Molich and Nielsen 1990).

To explain the concept of designing out and guarding against warnings, Laughery and Wogalter (2006) give examples of both. Examples of designing out can be to use non-flammable propellant in hairspray or removing sharp edges in products or pinch points in industrial equipment. Examples of guarding against risk can include guardrails on a motorway, or forcing an operator of a punch press to use both hands therefore ensuring that his fingers are not in the way.

2.1.1 Error Messages in Digital Systems

To make sense of certain parts of the literature review, it is necessary to understand a few common ways error and warning messages are displayed in digital systems.

One common method to display a warning or error message is to use a modal message, sometimes referred to as a pop-up message. With a modal message the error or warning is displayed within a dialog or window, over the application or website. While the dialog is visible, the user is not allowed to interact with anything other than the dialog. This prevents the user from continuing with the current task without first dismissing (closing) the error or warning dialog. This mode of display interrupts the current task and forces the user to interact with the error or warning before being allowed to continue. The opposite of this method would be to show the error or warning message, but not disallowing interaction with the rest of the application.

An important area where error and warnings are displayed and which should be carefully considered, is web forms. Web forms are found on websites and usually consists of a set of fields that a user has to fill in. An example of a web form is a registration page which requires the user to specify a user name, password, email address and physical address. Web forms are an important case because each field of a web form has the potential to have erroneous content. Therefore, an error message can be shown for each field separately at the same time. Additionally, interaction with the web form can also produce a warning or error message even if the individual form fields are all without errors. For example, a

communication error when trying to save the completed form, or a security warning related to supplying sensitive information. It should be noted that this is not only applicable to web forms, but also to any type of digital system where multiple elements exist simultaneously that have the potential to require the display of an error or warning message.

2.2 User Behaviour

There are a lot of factors found in the literature that describe the effectiveness of warnings and why they succeed or fail. On a more theoretical level, for a warning to be effective it must be noticed by a user, the information must be encoded and understood, and it must provide the right information to result in the required behavioural changes (Laughery 2006). The sections below include a discussion on the theoretical framework used to inform the literature review, followed by discussions about how and why warnings and errors grab attention and why users choose to comply, and a section on habituation.

2.2.1 Theoretical Frameworks

Popular theories that are used to inform warning and error message design is communication theory and human information processing theory. Communication theory is described by Laughery and Wogalter (2006) as a basic model that consists of a source, medium, message, and receiver. In terms of warning and errors the source is the originator of the error or warning, the medium is the method of communication (for example a sign or a message on a computer screen), the message is the content of the warning or error and the receiver is the target audience of the message.

The human information-processing model is a staged model where information flows through different processing steps. If the information is successfully processed in one step, it flows to the next step. If the processing fails at any step the flow is blocked and therefore the information processing fails (Laughery and Wogalter 2006).

Wogalter et al. (1999) proposed another theoretical framework called the communication human information processing (C-HIP) framework, which is a combination of the communication and human information processing frameworks. A diagram of the C-HIP framework is shown in figure 1.

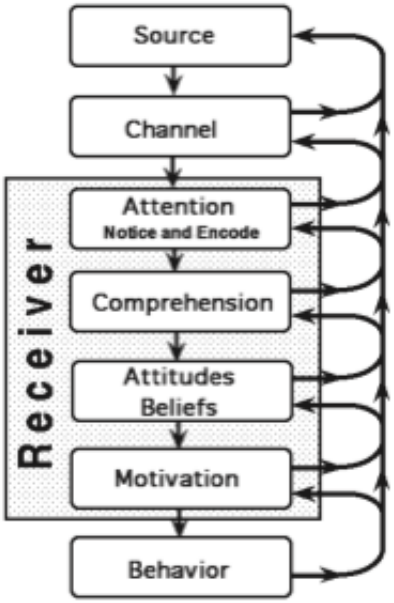


Figure 1: *The communication, human information processing (C-HIP) model (Laughery and Wogalter 2006)*

Laughery and Wogalter (2006) explains the C-HIP framework noting that it contains the elements of the communication theory as steps in processing like the human information processing framework. The receiver step contains four additional steps (attention, comprehension, attitudes and beliefs, and motivation) that describe the important aspects that affects the success of a warning or error message. The model also includes additional feedback loops showing that the different steps influence each other. An example of this cross-step influence would be when a warning is noticed and read, but comprehension fails, the user can always go back one step and read the message again.

From the C-HIP framework it is clear that the processing steps that form part of the receiver are key elements that will determine whether the warning or error can affect some user behaviour. Although not arranged specifically like the model, the following sections discuss these elements in more detail and evaluate how they affect the success or failure of a warning.

2.2.2 Attention (Notice and Encoding)

Before a warning can be effective and lead to behavioural changes, it must first attract attention (Laughery 2006, Laughery and Wogalter 2006, Laughery et al. 1993), and keep attention long enough for the warning to be comprehended (Egelman, Cranor and Hong 2008). This is the first receiver step in the C-HIP model. In the context of warnings, attention does

not only refer to whether it is seen or heard, but whether the information was encoded (read or listened to), and stored in memory (Laughery and Wogalter 2006). If the warning does not capture attention and is not seen, it cannot be expected to lead to behavioural changes (Laughery et al. 1993). In other words, if people fail to notice and understand the content of the warning, they will not have sufficient motivation to comply (Laughery 2006). This is in line with the C-HIP model where attention precedes comprehension. It is difficult to attract attention in real-world scenarios (Bartsch et al. 2013), and people do not typically seek out warnings by themselves (Laughery and Wogalter 2006, Laughery et al. 1993). Therefore, attracting attention is an important factor to consider during warning or error design.

2.2.3 Comprehension

After a warning is noticed and encoded, the user (receiver) tries to comprehend the encoded information (the second receiver step in the C-HIP model). Understanding is an important factor in warning effectiveness. Without understanding a warning, the comprehension step will fail, which will lead to an overall failure of the warning (Laughery et al. 1993). Therefore, design should not only focus on attention and motivation, but also on understanding and wrongful risk assumptions based on misunderstandings (Bravo-Lillo et al. 2011). This is specifically important in on-line security because users struggle to understand the digital risks (Egelman et al. 2008, Bravo-Lillo et al. 2011). This misunderstanding also affects the third receiver step of the C-HIP model where the wrong attitude or belief regarding the risk in question leads to warning failure.

2.2.4 Compliance Decisions

The effectiveness of a warning is dependent on how noticeable and understandable it is. But even if a warning is noticed and understood, it still fails when the user does not comply (Laughery et al. 1993). It is, therefore, important for a warning to provide all the required information for the recipient to make an informed compliance decision (Laughery and Wogalter 2006). When considering that warnings are intended to prevent damage or loss, compliance can be seen as the most important factor to judge warning effectiveness (Duarte et al. 2012). Following the C-HIP model, the compliance decision follows after the attention and comprehension steps and is affected by both the user's attitudes and beliefs, and motivation. The compliance decision of a user is affected by different factors which include properties of the warning message as well as characteristics and the environment of the user (receiver) (Laughery 2006).

One way to view the compliance decision process of users is to see it as a cost-benefit anal-

ysis, where the benefits of complying must outweigh the cost (Laughery and Wogalter 2006). Research has shown consistently that a reduced cost of compliance leads to increased compliance (Laughery 2006, Adams et al. 1998), and that warnings that seem satisfactory (is noticeable and understandable) might not be effective because of the amount of effort that is needed to comply (Wogalter et al. 1989).

The cost of compliance includes any negative economic factors that the user associates with compliance and can include financial cost of protective clothing, the time and effort it costs to carry out the activities required to comply, or even not doing something pleasing that is possibly unsafe (Adams et al. 1998, Wogalter et al. 1989). Apart from this, normative influences and the possibility of others' approving or disapproving of actions also affect compliance (Adams et al. 1998).

The cost-benefit analysis performed by users also include the possible consequences of compliance or non-compliance. This includes a mental analysis of the likelihood of something serious happening to them and whether the actions prescribed by the warning will prevent or reduce the likelihood of injury (Adams et al. 1998). It is important to note, however, that people tend to perceive themselves as less vulnerable than others and assign a lower risk to themselves compared to others (DeTurck et al. 1999, Adams et al. 1998).

The compliance decision is also affected by the social context and the people around the receiver. A well-documented and robust effect is that people tend to follow the behaviours of others they observe, even if that behaviour might not be in their best interest (Laughery and Wogalter 2006, Edworthy and Dale 2000, DeTurck et al. 1999, Wogalter et al. 1989). This is also seen in the compliance or non-compliance of safety warnings. For example, DeTurck et al. (1999) shows that people tend to imitate the safety behaviour of others. This effect is shown to have limits such as if the role model experience has a significantly negative effect when not complying with warnings, the users tend to follow the safety warnings instead of the compliance behaviour of the role model. This effect is more pronounced in scenarios where the user's perception is that the product use will not have a direct effect on them or if the perceived cost of compliance (nor non-compliance) is low (DeTurck et al. 1999, Edworthy and Dale 2000). This interaction is explained by stating that users use the behaviours of those around them as a heuristic cue for their own choices. This is because safety behaviour is mediated through the perceived hazardousness or risk, which is significantly affected by the safety behaviour of role models (DeTurck et al. 1999).

User trust also has a significant effect on warning compliance, where if users do not trust a specific warning, they will be more likely to ignore it (Egelman et al. 2008). In the context of web browsing, if people trust a specific website (for example a banking site) they tend to ignore security warnings (Kauer, Pfeiffer, Volkamer, Theuerling and Bruder 2012). It can be

said that the trust placed in the specific content provider is higher than the trust placed in the warning and therefore the warning is distrusted. This is especially dangerous because the manner in which non-technical users assess the level of trust they place in a website does not reflect real security concerns (Kauer et al. 2012).

The literature also shows that the familiarity of a product or environment affects the compliance decision. Familiarity in this sense refers to previous experience with the particular (or similar) product or environment (Laughery and Wogalter 2006). Research mostly shows increased familiarity leading to decreased compliance, although there are studies that show the opposite (Laughery and Wogalter 2006, Laughery 2006).

A possible explanation for this contradiction, discussed in Laughery and Wogalter (2006), is that the effect that familiarity has on compliance is mediated by perceived risk or hazardousness. When a specific interaction is perceived to have low risk, then familiarity with the product leads to people not noticing or reading warnings, because they are not looking for them. In the case where there is a high perceived risk, it is more likely that people will look for warnings and additionally the cost of non-compliance is perceived to be higher. Both these effects will lead to an increased likelihood of compliance.

2.2.5 Habituation

An important factor that affects the attention, comprehension and compliance of warnings, and therefore the effectiveness, is habituation. Habituation is defined as a reduction in attentional response due to repeated exposure to a stimulus (Kim and Wogalter 2009). In this case the stimulus is the warning or error message. In terms of the C-HIP model habituation will have an effect on all four receiver steps.

The habituation can have detrimental effects on the effectiveness of a warning or error message. According to Kim and Wogalter (2009) the detrimental effects can lead to a failure of a message in three ways: firstly, the message loses its capability to alert, which reduces the user's ability to notice it. Secondly, habituated users may stop attending to a message prematurely before all the information can be encoded. And lastly, when a user is habituated to a specific message, other messages with a similar appearance may also suffer from the same habituation effects. The first two failure modes negatively affect the attention step in the C-HIP model.

In software systems, habituation is especially evident in the non-compliance of security warnings in web browsing. This is discussed in more detail in Krol et al. (2012). Since users are exposed to frequent messages that include false alarms, they quickly become desensitized. If a user is given a security warning about a website that they know to be legit-

imate, the user distrusts the warnings due to false positive, which leads to the disregarding of subsequent messages. In other words, the attitudes regarding the warning affects the motivation to comply. Additionally, a misunderstanding of the risk associated with a warning can lead to desensitization, which is important because users do not understand the risk and repercussions around digital security (Krol et al. 2012, Bravo-Lillo et al. 2011).

Apart from habituation in warning and error messages, the literature also shows dishabituation and recovery effects of warnings. Dishabituation is an increase in alertness when the format of a warning changes, and the recovery effect is the return to the low alertness level when the warning is changed to the habituated format (Kim and Wogalter 2009).

In order to combat the effects of habituation it is firstly necessary to reduce the frequency of warning messages and to prevent the occurrence of warnings in uncritical situations (Bartsch et al. 2013, Maurer et al. 2011). This is in line with the safety control hierarchy (Laughery and Wogalter 2006, Laughery 2006, Bravo-Lillo et al. 2011). Additionally, critical or very serious warnings should be styled differently to decrease the habituation effects that non-critical warnings will have on critical ones (Egelman et al. 2008).

2.3 Warning and Error Design

A lot of research has been done on the attributes of warning and error messages and how they are affected by and are related to attention, compliance and habituation. These attributes include (but are not limited to): size, colour, shape, location, linguistics, and use of pictorials (Laughery and Wogalter 2006, Laughery 2006). The sections below highlight the important factors discussed in the literature regarding these warning and error message attributes.

2.3.1 Content

The content of a warning or error message refers only to the words used in the message, which excludes layout related factors such as colour, font and size. The content of a warning or error is an important factor that contributes to its effectiveness (Harbach et al. 2013). On a very basic level it is important that the warning or error message clearly communicates all the information required to make an informed compliance decision.

When designing warning and error message content, the explicitness of the content is an important factor that impacts the success of a message (Laughery and Wogalter 2006, Wogalter et al. 1997). The research shows that being explicit in the information communicated to

a user will enable the user to make better compliance decisions because the cost and benefits are clear (Laughery 2006). For a warning, the content should include details about the hazard, consequences of non-compliance as well as instructions on how to mitigate the risk (Laughery and Wogalter 2006, Adams et al. 1998). The literature also shows that providing clear choices (if applicable) of how to mitigate the risk increases the effectiveness of a warning (Egelman et al. 2008). An error, however, should inform the user of what has happened, what the cause of the error is and how to recover from the error (Kelkar et al. 2013).

In warnings it is specifically important to be explicit when describing the consequences of non-compliance (Egelman et al. 2008), and this becomes more important when the consequences become more severe (Laughery and Wogalter 2006). It is also more effective to communicate personal risk rather than technical risk, although both are important ((Kauer et al. 2012). This is especially true for digital risks, which are typically misunderstood, and when analogies to the physical world are incomplete (Bravo-Lillo et al. 2011). Care should be taken when describing consequences because lack thereof can lead to distrust from the user. Distrust can stem from a user's perception that the consequences are exaggerated, or from a misunderstanding of the more explicit content, both of which lead to doubts and reduces motivation to comply (Bartsch et al. 2013).

In order to communicate successfully, care should be taken when deciding on the length of the message. The trade-off is that longer messages can be more explicit about the warning and give the user all the required information. Unfortunately, longer messages can have negative effects on compliance due to over warning and can cause a reduction in the effectiveness of communication because users might be selective in the parts they read (Laughery and Wogalter 2006). Although it seems counter intuitive, some research suggests that a longer message can lead to increased willingness to read it due to an increase in perceived risk (Silver, Leonard, Ponsi and Wogalter 1991). However, it is expected that the initial level of perceived risk will have an effect on the user's willingness to read the message. The research also shows that users do not have to read an entire message to know what action to take (Egelman et al. 2008). This said, it can be expected that users will then not read any warning or error messages thoroughly and, thus, to increase the length of the message could worsen the problem (Bravo-Lillo et al. 2011) and lead to ineffective warnings.

In the end it is expected that more effective communication can be achieved by using shorter content that is easier to read (Harbach et al. 2013). A factor related to the length of the warning message is the linguistic complexity (or language difficulty). Sometimes in an effort to reduce the length of a warning, the complexity of a message is increased. This can lead to confusion and can lead to the reader failing to comprehend the message (Harbach et al. 2013, Barlow and Wogalter 1991). Although it is an important factor, brevity should not be improved at the cost of simplicity and readability. In the end the goal is to provide

understandable information that the reader needs in order to make an informed compliance decision (Laughery and Wogalter 2006).

In order to facilitate better compliance decisions, the user needs to have all the relevant information to make the correct decision. This might be a difficult task when considering the negative effects of long warning messages. To increase the capacity of a warning to be able to warn without becoming too verbose, the content can be updated dynamically as a way to convey only the required information (Laughery and Wogalter 2006). This dynamic content can serve to personalise a warning to fit the specific characteristics and capabilities of the user, or dynamic content can be used to fit content to the user's intention by naming concrete consequences related to the user's intention (Laughery and Wogalter 2006, Bartsch et al. 2013). Research shows that by dynamically changing the content of a warning based on the specific context, users are able to assess the risk more correctly (Bartsch et al. 2013). For example, the user's intention should be taken into account in on-line security warnings such as mentioning credit fraud when viewing on-line shopping websites.

A design element that is frequently used in warnings is signal words. These are specific words used to give a general indication of the level of hazard (Laughery and Wogalter 2006). Typically, warnings tend to use three signal words namely *Danger*, *Warning*, and *Caution*. According to warning design standards, *Danger* denotes the highest risk scenario where the message involves an immediate hazard that will result in severe injury or death; *Warning* is intended for hazards that may cause injury; and *Caution* is used for hazards that can cause minor injuries or damage (Wogalter and Silver 1995). Research shows that Danger attracts the most attention (Laughery and Wogalter 2006, Laughery 2006), but people do not assign different levels of risk to *Caution* and *Warning* (Wogalter and Silver 1995). Consequently, Wogalter and Silver (1995) suggest that these signal words should be used in line with the guidelines set out in warning design standards in order to reduce incorrect risk assigned to signal words by users.

The tone of the error and warning messages also has an effect on their effectiveness. Generally, it is considered good practice for a warning or error message to use a kind and helpful tone instead of an accusatory tone (Kelkar et al. 2013). In other words, an error message should rather blame the problem on system deficiency instead of the user (Molich and Nielsen 1990). This change in tone should deliver a better user experience and reduce stress levels, which can lead to better comprehension. In contrast to this, Rainer and Stefan (2010) reported that polite request and warning text, which points to a voluntary decision, reduces the probability of consent. It is presumed that this is due to users being habituated to end user license agreement (EULA) dialogs, and users tend to agree to everything that looks superficially like EULA. As long as this factor is kept in mind when designing messages, it should not present a problem.

The technical expertise of the target audience should be kept in mind when designing error and warning messages. This includes steering away from technical terms in the warning content (Bravo-Lillo et al. 2011, Harbach et al. 2013), as well as keeping possible misconceptions users might have in mind, like misconceptions about computer security and anti virus software (Bravo-Lillo et al. 2011, Krol et al. 2012).

Another factor to consider in the encoding of a warning is the ambiguity of the message. If a warning message is ambiguous, it will lead to incorrect encoding of the message. Also, as a result, the ambiguous message can lead to incorrect behaviour and uncertainty (Duarte et al. 2012). Although ambiguity is usually due to confusing content, the content may become ambiguous only in a specific context. The context of an error can also be affected by the placement and layout of the error, therefore, the ambiguity is also a factor in layout.

2.3.2 Pictorials

Pictorials (or icons) can be used in error and warning messages to communicate information about the hazard, consequences and severity, as well as to convey instructional information in a space-saving, non-verbal way (Laughery 2006, Laughery et al. 1993, Davies et al. 1998). The use of pictorials has a positive effect on attention because it makes the warning or error message more conspicuous (Laughery and Wogalter 2006, Davies et al. 1998). The conspicuousness of a warning is improved when a pictorial is used in conjunction with text as opposed to a warning with just an icon or text (Davies et al. 1998). Pictorials also improve compliance by providing extra information that helps with better informed compliance decisions (Laughery 2006).

A pictorial has different characteristics that have an impact on its effectiveness. In a study by Isherwood, McDougall and Curry (2007) a list of characteristics is given along with their relative importance. These characteristics include correctness (the extent to which the icon represents real objects), complexity (the intricacy or amount of detail), semantic distance (how close the icon is to the function it represents), and familiarity (both from frequent use and familiarity of depicted objects).

The study shows that semantic distance and familiarity have the most effect on successful icon identification. Semantic distance is most important for situations where immediate understanding of an icon is important. For situations where icons are used on a regular basis, familiarity with what is depicted in the pictorial is most important. Other literature suggests that icons that are abstract in nature or too complex result in lower comprehension (Davies et al. 1998).

In terms of warnings and errors, research suggests that people are already familiar with a

red ‘x’, a yellow exclamation mark and a blue ‘i’ depicting an error, warning and informational message respectively (Kelkar et al. 2013). Therefore, it should be considered to be included in error or warning displays. If new or untested icons are used, proper testing should be performed to make sure that they are understandable and noticeable (Davies et al. 1998).

2.3.3 Layout

In general, users do not actively look for warnings. Therefore, making sure that the warnings are visually conspicuous and stand out from the background will increase the effectiveness of a warning (Laughery et al. 1993). The elements of the visual style of a warning that has been shown to have an effect on attention are: size, placement, borders, contrast, and colour, where the specific colour is not as important as the amount of colours being used (Laughery et al. 1993). Larger sized warnings are generally more conspicuous and is, therefore, easier to notice. More specifically, it is the size relative to other items in the surrounding area that is important (Laughery and Wogalter 2006, Barlow and Wogalter 1991).

Apart from the visual styling, placement of an error or warning should also be considered. It is important that information is presented when and where it is needed. Therefore, warnings should be located close to the hazard in both time and space (Laughery and Wogalter 2006).

In software systems the mode of display of an error or warning can affect the success of the message. It is generally seen as a usability problem to present a user with pop-up dialog with an error or warning, because it interrupts the flow of interactions (Kelkar et al. 2013). Since users dislike disruptions, it is questioned whether pop-up messages are effective (Krol et al. 2012). This said, passive (non-blocking) warnings are not nearly as effective and usually go unnoticed (Egelman et al. 2008).

Bravo-Lillo et al. (2011) tested inhibitive warnings where users were presented with a blocking dialog that prevented the user from proceeding past the warning by either forcing them to wait for a certain time, or by only allowing them past once they have performed some specific interaction with the message directly. The result of the study shows that these dialogs increased the effectiveness of the warning. Although this study concludes that the inhibitive warnings can increase effectiveness, the cost in terms of usability and frustration incurred might be too high. It might, therefore, be better to focus on other methods of designing effective warnings and errors that do not go against design and usability guidelines.

Another design concern relating to layout and mode of display found in the literature is the timing of validation messages in web forms. A regular occurrence on websites are forms that need to be filled in, such as a registration form for example. These are referred to as web forms. A web form usually consists of a number of form fields each of which can potentially

have different rules on what constitutes valid data. For instance, what constitutes valid data in an email address field is different from a telephone number field. It is, therefore, expected that error or warning messages can exist for the separate form fields at the same time.

The design of errors and warnings surrounding web forms is important for this study because it is a common occurrence on websites and where error messages are frequently displayed. A contentions point found in the literature regarding validation in web forms is whether validation messages should be displayed immediately (during form completion) (Al-Saleh et al. 2012) or after from submission (Bargas-Avila et al. 2007).

Although showing validation messages after form submission is not recommended by usability guidelines and ISO design standards (Al-Saleh et al. 2012), Bargas-Avila et al. (2007) show empirical evidence that immediate feedback does not enhance form performance and users tend to ignore the feedback while completing the form. In contrast to this, Al-Saleh et al. (2012) continue from the research presented by Bargas-Avila et al. (2007), and show that immediate in-line feedback does have a positive effect on form performance. The study also shows that non-critical warnings or information shown in-line is ignored by users similar to the effects noted in Bargas-Avila et al. (2007). It is, therefore, suggested that in-line immediate error feedback should be shown, especially for complex form fields, but further study is required to validate this.

This concludes the discussion of theory found in the literature regarding the design of warning and error messages. The remainder of the chapter is devoted to a discussion of the role and methods of usability testing in interface design.

2.4 Usability Testing

The research reported in this dissertation aims to produce a set of guidelines for error message design. The purpose of these guidelines is to assist in the evaluation of a user interface during the development process. It is important to know how user interfaces are typically tested during the development process, because this will show how the guidelines for error message design will be used. There are two major classes of usability testing. The first class is empirical user testing where an interface is tested with representative users. The second class is usability inspection methods which typically include inspection with developers and usability experts instead of real users. These methods are considered cheaper and easier and can be used earlier in the development process (Hollingsed and Novick 2007).

The sections below will give a short summary of different usability testing techniques, starting with empirical user testing and followed by the four most common usability inspection meth-

ods namely: heuristic evaluations, cognitive walkthroughs, pluralistic usability walkthroughs, and formal usability inspections (Hollingsed and Novick 2007). Of these four, the heuristic evaluation will be covered in the most detail because it is the inspection method where the guidelines are used the most directly.

2.4.1 Empirical User Testing

Empirical user testing involves testing a user interface or product with real users. This is sometimes referred to as *usability testing*, but in this discussion to distinguish it from other usability inspection methods, it will be called *empirical user testing*. Rogers, Lamson and Rousseau (2000) describes empirical user testing as the traditional usability test of a product in a controlled laboratory setting. Performing the tests in a controlled environment enables the evaluators to control the environmental and social influences on a user as well as what the user does. The goal of this type of test is to determine whether a product or interface is usable by the intended user base and can be used to complete the tasks it was designed for. The data that is gathered during an empirical user test can include video recordings of the user, logged keystrokes, mouse movements etc. Sometimes, users are also asked to think out loud while they are completing a task to get an insight into what the users are thinking. In order to gather additional information from the user, a satisfaction questionnaire can be used, or structured or semi-structured interviews can be conducted with the users.

Jeffries and Desurvire (1992) compare empirical user testing to other usability inspection methods and give some advantages that empirical testing have. The major advantage is that because the testing is done with real users, the problems identified are per definition problems that will bother actual users. There is also no need to try to predict the impact of a usability problem, the impact should be obvious from the test itself. These results are also more convincing to the developers and designers of the system in comparison to other expert based tests where there might be some doubt about whether the problems identified by usability experts are really problems for users. Some classes of problems are also more likely to be discovered when real users use the application because they are able to interact with the system in unexpected ways, which is less likely to happen in inspection methods. Empirical user testing is considered expensive in comparison to usability inspection methods, but it is not advisable to substitute empirical testing entirely with usability inspections (Jeffries and Desurvire 1992, Hollingsed and Novick 2007).

Empirical user testing is considered not only to be expensive, it is also considered to be time-consuming and is typically only applied later in the development cycle (Jeffries and Desurvire 1992). Testing earlier in the development cycle is possible, however, when using low-tech prototypes of interfaces (Hollingsed and Novick 2007). The perceived high cost in

terms of both resources and time is one of the main reasons why empirical testing is not conducted in practice (Nielsen 1994).

2.4.2 Heuristic Evaluations

Heuristic evaluations were first proposed in Nielsen and Molich (1990). This is described as an informal method of usability analysis where evaluators are asked to analyse a user interface and comment on what is good and what is bad. A defining factor of a heuristic evaluation is that the analysis is guided by a set of established usability principles, sometimes referred to as heuristics (Nielsen 1992).

The use of heuristics is advantageous due to the expectation that problems identified using heuristics will be easier to resolve because the heuristic itself should give guidance on how to resolve the problem (Nielsen 1995b). An additional advantage of using guidelines or guiding principles is that it forces the evaluator to carefully examine the interface (Jeffries, Miller, Wharton and Uyeda 1991). It is important to keep the number of heuristics used in an evaluation low to ensure that they are intellectually manageable (Molich and Nielsen 1990). A number of between 5 and 10 is suggested which ensures that the heuristics are sufficiently discriminating while remaining easy to remember (Rogers, Sharp and Preece 2011). The research also shows that when a large amount of guidelines is used, the evaluators perform poorer (Jeffries and Desurvire 1992).

Heuristic evaluation is difficult and a single evaluator will never find all the usability problems in a system (Nielsen 1995b). In general, individual evaluators are not very good at finding usability problems. An experiment shown in Nielsen and Molich (1990) showed that an individual only found between 20% and 51% of the usability problems in a system. The interesting thing is that in the evaluation, different people found different problems and in some cases an evaluator who was able to find many problems in an interface missed easy problems, and evaluators who struggle to identify problems sometimes identified hard problems. This presents the possibility of aggregating the results of multiple heuristic evaluations of the same system by collecting the usability problems found by the different evaluators into a single set of problems.

It is recommended by Nielsen and Molich (1990) that the number of usability problems found with an aggregate of more than one evaluation increases rapidly from one to five evaluations, but reached a point of diminishing returns from ten. From these results it is recommended to use between three and five evaluators in a heuristic evaluation where the aggregate of five evaluations is expected to find about two thirds of the usability problems in a system.

An enhancement to the heuristic evaluation process that is sometimes added is having the

evaluators rate the severity of usability problems identified. As described in Nielsen (1995c) the severity rating process is completed after the initial evaluation. All the results from the evaluators are first aggregated to form a unified list of identified usability problems. Each evaluator is then given this list and asked to rate the severity of each problem. It is recommended that more than one evaluator (ideally at least three) complete the severity rating process to ensure the quality and reliability of the severity rating. The purpose of the severity rating process is to enable the allocation of resources to the most severe usability problems first, and to assist in the estimation of effort still required. It also gives an indication on the current quality of the user interface which can assist in the decisions on whether to release the product.

Nielsen (1995c) states that the combination of the following four factors define the severity of a usability problem:

- The frequency of the problem (is the problem common or rare).
- The impact of the problem (how easy is it to overcome the problem).
- The persistence of the problem (is it a problem that will continually bother the user or is it something that is encountered once, and that the user can overcome it once they know about it).
- The market impact of the problem. For example, a problem that might be easy to overcome and which is infrequent might have a devastating impact on the popularity of a product.

These factors are typically combined into a single severity rating to facilitate rating and prioritisation. The following scale is also suggested (Nielsen 1995c):

0. This is not usability problem.
1. This is only a cosmetic problem that should be fixed if extra time is available.
2. This is a minor problem with a low priority to fix.
3. This is a major problem with a high priority to fix.
4. This is a critical usability problem that has to be fixed before the product can be released.

In the initial proposal of heuristic evaluation in Nielsen and Molich (1990), evaluations were performed by people without usability expertise. In subsequent research, however, it was

shown that usability experts were much better at finding usability problems (Nielsen 1992, Jeffries and Desurvire 1992). Nielsen (1992) included an analysis of how good novice evaluators, usability experts and double experts are at identifying usability issues. Double experts in this case refers to people who are usability specialists with specific expertise in the type of interface being tested. The research shows that double experts perform significantly better than usability experts, and that they are especially good at identifying problems that are unique to the type of interface being tested.

Although it is possible to complete a heuristic evaluation with minimal usability training, more individual evaluations (10 to 20) have to be completed and aggregated to result in the identification of most of the usability issues (Jeffries and Desurvire 1992). Nielsen (1992) found that two to three double experts were able to find most problems (81% to 90%). With three to five expert evaluations 74% to 87% of the problems could be identified. With novice evaluators, fourteen evaluations are needed to find more than 75% of the problems and with the recommended 5 evaluators, only 51% of issues were found. The sourcing of usability experts can be difficult and expensive, especially if they are expected to have expertise in the specific type of interface being tested (Nielsen 1992). If it is not possible to get these resources, using engineers or developers can be a suitable alternative (Jeffries et al. 1991).

One of the advantages of heuristic evaluations is that it is less expensive compared to other methods (Nielsen and Molich 1990). Specifically, it is faster and more resource efficient in relation the amount of usability problems it can be expected to identify (Georgsson, Weir and Staggers 2014).

Research shows that heuristic evaluations are able to find many usability issues, including serious issues, but the problem is that the experts required to perform the evaluation can become expensive (Jeffries et al. 1991). Additionally, it is observed that the problems identified will be dominated by minor problems (Jeffries and Desurvire 1992). More problematic even, is the possibility of wasting time on a non-issue that might even reduce the usability of a product (Jeffries and Desurvire 1992).

The problem with the large amount of minor problems that heuristic evaluations usually identify is that extra effort needs to be put in to prioritize the issues and schedule resources to address them. This increases overhead cost and together with the need to use usability experts. It is questioned whether the overall cost and effort associated with heuristic evaluations are really more cost-effective than empirical user testing (Jeffries and Desurvire 1992).

Another advantage of heuristic evaluations is that it can be used early in the development process (Nielsen and Molich 1990, Hollingsed and Novick 2007) when it is easier and less expensive to fix usability problems. Heuristic evaluations also do not require the same

amount of advanced planning as empirical user testing (Nielsen and Molich 1990). Another advantage of heuristic evaluations is that it is an intuitive evaluation method, and it is easy to motivate people to do it (Nielsen and Molich 1990). The results of a heuristic evaluation do not require as much interpretation as empirical user testing, which adds to the ease of evaluation (Nielsen 1995b).

Although heuristic evaluations are able to find usability problems, it is not fool proof and some issues are missed. Research consistently shows that different methods have various strengths (Jeffries and Desurvire 1992). Because of this, it is recommended that the design and development process should not rely on heuristic evaluations alone, but should use other methods as well to find problems missed by heuristic evaluations (Nielsen and Molich 1990). More specifically, it is recommended in Nielsen (1995a) that when using more than one evaluation method, an iterative design process should be followed between evaluations, typically a heuristic evaluation will be completed first to find any obvious usability issues. These usability issues should first be addressed and thereafter the updated user interface should be subjected to empirical user testing that will both validate the result of the updated design as well as find the problems that was missed during the heuristic evaluation.

2.4.3 Cognitive Walkthroughs

The cognitive walkthrough process, as proposed in Polson, Lewis, Rieman and Wharton (1992), is described as an evaluation methodology that focuses on ease of learning of the interface. With a cognitive walkthrough it is possible to identify some usability problems, specifically learnability problems that might be experienced by users that have not received formal instruction.

A cognitive walkthrough involves the author of a specific design presenting the proposed design to a group of peers. The group of peers then evaluates the solution using a set of criteria that is appropriate for the specific class of design issues. The evaluation involves the reviewers stepping through user actions and for each action determining:

1. What the goal of the user is leading up to the action.
2. Whether the prompts and labels in the interface will guide the user to take the correct action given that the current goal, and
3. How the goal of the user will change based on the feedback from the interface after the action.

By following this process, the reviewers attempt to identify actions that will be difficult for the

average user to choose or to execute.

This evaluation can be performed at any time of the design and development process, which implies that issues can be identified without empirical user testing. The cognitive walkthrough process is not expected to identify all the problems with an interface, but it does provide a systematic process to evaluate an interface in depth with less cost and effort than empirical user testing. The results of the walkthrough can also assist in focusing subsequent empirical user testing on identified problem areas.

Although a cognitive walkthrough can be successfully used to find usability issues when performed by non-human factor engineers (Jeffries et al. 1991), it has a more intuitive appeal to people with an understanding of cognitive theory (Kahn and Prail 1994). Kahn and Prail (1994) also state that the concepts and terminology used in cognitive walkthroughs can be difficult to understand for people without experience in human factor engineering.

2.4.4 Pluralistic Usability Walkthroughs

The pluralistic usability walkthrough is presented in Bias (1994). The details of this inspection method, as given in Bias (1994), is discussed briefly in this section. The pluralistic usability walkthrough can be described as a usability inspection method where three types of participants (an actual user, product developers, and usability experts) are included in a panel that analyses a user interface, but all the participants are asked to assume the role of a user.

In a typical pluralistic usability walkthrough, the evaluators are presented with a typical task that would form part of the usage of the user interface being tested. With the task in mind, the participants are shown a hardcopy of a screen in the user interface and then they are asked to write down what they would do in pursuit of completing the given task. Participants are expected to write down in as much detail as possible what they would do, ideally to a keystroke or mouse movement level. Once all the participants have finished writing down the action they will perform, each participant will share their action and discuss it with the group, starting with the representative user.

One of the defining factors of a pluralistic usability walkthrough is the inclusion of representative users, developers, and usability experts, each of which has an important purpose or function in the walkthrough. By including a representative user and the developers, the developers are given first-hand knowledge into what real users feel and how they interact with the design. This is also expected to increase the empathy of the developers (their sensitivity to real user concerns). The purpose of the expert is to firstly provide their superior usability intuition and secondly, to assist the user to express their concerns in a meaningful way that

can be translated into improvements to the interface. Additionally, the usability experts have to advocate for the concerns of the user and prevent the developers from explaining away their concerns. It would be possible for a pluralistic usability inspection to be conducted without usability experts, but the results are expected to be sub-optimal and care should be taken that the user's opinion is not overridden by the developers.

Another benefit of pluralistic usability walkthroughs is the fact that it can be conducted early in the design process when a prototype is not yet available. This means that performance and satisfaction data from users can be gathered earlier than with empirical user testing. Because this inspection method includes a group discussion, the results can include solutions to identified problems. The cost and effort involved in pluralistic usability walkthroughs are also considered less than with empirical user testing.

The primary limitations of pluralistic usability walkthroughs are firstly that the evaluation can only progress as fast as the slowest participant on the team. This implies that the participants will not always have a good grasp on the flow of the interface. Secondly, the evaluation can only continue down the interaction path selected beforehand. This means that some insights that could have been gathered when users perform unexpected or unplanned interaction, would be missed. Additionally, participants that selected an incorrect action, or the action that was not planned for, will have to reset their cognitive process in order to continue down the selected path. These limitations do not discount the findings of a pluralistic usability walkthrough, but simply implies that the results should be analysed while keeping the limitations in mind.

2.4.5 Formal Usability Inspections

Formal usability inspections were first proposed in Kahn and Prail (1994). Formal usability inspections can be defined as a formal process that, when followed, enables the developers and designers to review a product and the potential task performance of users. This will enable them to find valid usability problems and describe them in an effective manner, which means they can be easily solved.

A formal usability inspection is performed by a team of four to eight engineers. Each inspector in the team is given an inspection package that includes:

- A product description in the form of screen drawings and explanatory text, and can include storyboards and prototypes.
- User profiles, which are descriptions of typical users that will use the system. The minimal requirements of a user profile are: a label, a description of the user education,

and a description of the user's experience that is relevant to the current product begin tested.

- Task scenarios, which describe specific scenarios that will be tested. Each scenario should include the user goal expressed in user terms, the starting point of the task which should include the user's situation and the product state, and any other intermediary situations that the user will encounter. The task scenario should not include specific steps that the user should complete. It is part of the inspection process to determine what the steps are that the user will follow.

The user profile is important because it assists the inspectors to analyse the product from the user's perspective and also impresses on them that all users are not the same and will interact with differently with the system. The task scenarios are important because it forces the inspection to deal with real work, and not inspect the product at an abstract level.

The user profile is important because it assists the inspectors to analyse the product from the user's perspective and also impresses on them that all users are not the same and will interact differently with the system. The task scenarios are important because it forces the inspection to deal with real work and not inspect the product at an abstract level.

The formal usability inspection process involves the inspection team completing each of the task scenarios for each user profile. For each pair of task scenario and user, the inspectors will attempt to determine what problems the user will experience (identify usability problems). This inspection does not necessarily have to include multiple user profiles. The number of profiles and scenarios covered in an inspection will be dependent on the requirements of the inspection and the time available.

Apart from the inspection package, the evaluators are also provided with a task performance model flowchart that shows an abstract depiction of the cognitive steps that users go through when performing tasks. During the inspection of each pair of task scenario and user profile, the inspectors will apply the model to determine what the cognitive phases are that a user has to complete. By using this tool, the inspectors are given an insight into the parts of the scenario completion that are carried out in the mind of the user, and helps the inspectors to consider what could go wrong in these invisible mental phases.

A formal usability inspection will usually also use heuristics to guide the inspection. Where the task performance model helps the inspectors to know where to look, heuristics will help the inspectors to know what to look for. The heuristics that are used in the inspection can include general usability heuristics or heuristics that are specific to the project being tested.

The benefits of completing a formal usability inspection is primarily that the final product

will be more usable, effective and efficient because users will encounter fewer problems. Additional benefits include the exposure of the engineering team to user-centered design which is expected to lead them to create better designs in the future. Peer engineers in the inspection team are exposed to projects other than their own which will enable them to support the project in the future, and the created scenarios and user profiles can be reused in future inspections and user testing.

The primary focus of this inspection method is to formalise the inspection process that will enable the inspection team to use their time efficiently. The formal usability inspection provides a clear agenda for the inspection which contributes to the efficiency of the inspection because the inspectors do not waste time figuring out what to do. Although it takes times to set off and prepare for a formal usability inspection, in the long run it increases the inspections effectiveness and efficiency.

2.5 Conclusion

The review above described the literature of interest regarding error and warning message design. This included a discussion around the theoretical considerations or error and warning design and why users behave in a certain manner, and the more technical considerations when designing error and warning messages, where the technical considerations are informed by the theory. This literature was used to inform a set of design guidelines that is the object of the research (the guidelines are discussed in detail in Chapter 4).

The sections above also included a summary of different usability testing methods that are used in practice. This showed that the guidelines developed in the research can assist in the evaluation process. When using the developed guidelines as heuristics the most relevant use of the developed guidelines would be in heuristic evaluations, which is found to be a useful method with which to identify usability problems early in the development process (Nielsen 1992). Additionally, heuristics are also used in pluralistic usability walkthroughs (Bias 1994). Although heuristics do not form an integral part in all usability evaluation methods, the literature shows that a successful development process should include the use of different usability testing methods (Nielsen and Molich 1990, Jeffries and Desurvire 1992, Nielsen 1995a, Hollingsed and Novick 2007).

The following chapter will place the literature review in the larger context of the study and describe the research methodology.

3 Methodology

In this chapter a discussion regarding the research methodology followed in this study is given. The research follows a design science research approach which focuses on the creation of an artefact, in this case guidelines for error design in digital systems. The sections below include an overview of the research philosophy and the paradigms that are prevalent in information system (IS) research and an in-depth discussion about design science research. Following the discussion regarding research paradigms and design science research, the research will be described and framed within the design science research approach to show its validity. Thereafter, a discussion around the specific research tasks that were completed and ethical considerations is included.

3.1 Introduction

The research aims to create and validate a set of guidelines for error message design in digital systems. The overarching question that the research answered is: what are the general guidelines to consider when designing error messages in digital systems? In order to answer this, the following sub questions are answered:

- What are the primary concerns found in the literature regarding the design of warning and error messages?
- How do these concerns translate into guidelines for error messages design in digital systems?
- Are the guidelines effective, that is to say, how easy is it to understand the guidelines and use them to analyse error messages?
- Are the guidelines valid, in other words, do the design guidelines reflect the pain points and concerns users have regarding error messages in digital systems?
- What are the relative importance of the different guidelines?

By answering these questions, the research is able to present a set of guidelines for error message design in digital systems. The research includes an analysis of the guidelines which shows that they are effective and valid.

3.2 Research Philosophy

In IS research there are four paradigms of note namely: positivism, interpretivism, critical research (Myers 2013), and design science (Vaishnavi and Kuechler 2004). These paradigms inform the research assumptions and process, and can be defined in terms of its ontological, epistemological and axiological perspective. Vaishnavi and Kuechler (2004) define these perspectives as follows:

- *Ontology* is the study of the nature of reality, that is, how the researcher views reality.
- *Epistemology* is the study of knowledge, which includes how it can be defined, how it is created, and what constituted valid knowledge.
- *Axiology* is the study of values. Research produced is measured and motivated by that which is valued by individuals and the larger research community.

3.2.1 Positivism

In positivist research the goal is to provide explanations of the world, where the ontological view is that the world consists of one reality that is observable and measurable (Hirschheim and Klein 1989). In this world view causal relationships between phenomena exist and these phenomena can be explained and predicted through general laws (Orlikowski and Baroudi 1991). An additional assumption is that observation is value free, that is, that the researcher plays a passive role where the phenomena being observed is unaffected by, and independent of the observation (Orlikowski and Baroudi 1991, Myers 2013).

The epistemological view of positivism is that knowledge is created in the form of generalised theories that can both explain as well as predict phenomena and also requires that theories be empirically testable and falsifiable (Orlikowski and Baroudi 1991). In positivist research value is placed on these theories and their ability to explain, predict, and be generalised (Vaishnavi and Kuechler 2004).

3.2.2 Interpretivism

Similar to positivism, interpretivism seeks to explain phenomena, but this happens within the realm of individual consciousness and subjectivity (Hirschheim and Klein 1989), where people each create their own subjective meaning of the world with which they interact (Orlikowski and Baroudi 1991). Since meaning is assigned subjectively, the ontological perspective is

that there does not exist a single reality, but a myriad of different perspectives (Hirschheim and Klein 1989).

The epistemological standpoint of interpretivism, unlike positivism, does not search for causal and empirical explanations but rather has a focus on sense-making (Hirschheim and Klein 1989). The purpose of interpretative study is to understand phenomena and this is achieved by studying the subjective meaning people assign to them (Orlikowski and Baroudi 1991). Such a study is heavily dependent on the context and therefore generalisation to a larger population is not possible nor is it sought (Orlikowski and Baroudi 1991). Thus, the axiology of interpretative research is *understanding* (Vaishnavi and Kuechler 2004).

3.2.3 Critical

The ontological perspective of critical research is that everything, and everyone, has unfulfilled potential, but also that the capacity to enact change is constrained by alienation from potential fulfilling environments by the prevailing authorities (Orlikowski and Baroudi 1991). The critical researcher, therefore, tries to impart understanding and awareness of these constraints, thereby empowering people to act towards eliminating that which constrains them (Orlikowski and Baroudi 1991).

The epistemology of critical research is similar to interpretivism, especially its stance on objective observation (Myers 2013). In a pursuit of emancipation, the researcher must identify the shared understanding of the obstacles to communication, which can only be done by obtaining insider knowledge through genuine participation. This would not be possible through objective observations alone (Hirschheim and Klein 1989).

Where critical research differs from interpretivism is that instead of only seeking explanation, critical research seeks to emancipate people from constraints and compulsions towards a state of justice and freedom (Hirschheim and Klein 1989). This emancipation is obtained by taking a critical stance towards the status quo and exposing the alienating and restrictive social conditions (Orlikowski and Baroudi 1991).

3.2.4 Design Science

A final research paradigm discussed here is design science. The research paradigms mentioned above are mostly concerned with explaining and predicting human behaviour (Hevner, March, Park and Ram 2004). Due to the explanatory nature of other research paradigms the resultant output is less applicable to problems seen in research and practice (Peffers et al. 2007). In contrast to this, design science seeks to extend the capabilities of humans

and organisations by creating artefacts and addressing research through the creation and evaluation of artefacts that aim to address certain business needs (Hevner et al. 2004).

The ontological perspective of design science is that of multiple world-states (Vaishnavi and Kuechler 2004). These world states are described by Gregor and Hevner (2013) as follows: the first world is an objective world that contains material things, which includes the instantiations of artefacts. The second world is a subjective world which includes the ideas and experiences of users and designers. Finally, the third world is an abstract world of man-made entities such as language, theories, models and constructs. This view is seemingly similar to the multiple reality view of interpretivism, but it differs in the fact that most design science researchers still assume a stable underlying physical reality that is necessary to be able to create an artefact (Vaishnavi and Kuechler 2004).

In design science, knowledge is tightly coupled with the creation of an artefact, where the methods, techniques and constructs used when creating, describing, and evaluating the artefact contributes to the body of knowledge (Vaishnavi and Kuechler 2004, Gregor 2006). Truth is then proven through construction and circumscription, where truth is measured by the predictability of interaction of an artefact with information (Vaishnavi and Kuechler 2004).

What is valued in design science research most is functionality in the form of manipulation and control over an environment. Additionally, design science research also values the more traditional research values like truth and understanding (Vaishnavi and Kuechler 2004).

An important part of design science is the information technology (IT) artefact. The definition of artefacts is very broad and includes constructs, models, methods, and instantiations (Hevner et al. 2004). Examples of artefacts include, but are not limited to, algorithms, human computer interfaces, and system design methodologies (Vaishnavi and Kuechler 2004). Design science revolves around the artefact where the output of research is a description of the artefact that addressed a specific business need, and the research methodology is concerned with designing and evaluating the artefact (Hevner et al. 2004, Gregor and Hevner 2013).

3.3 Design Science Research

In the above section, the research paradigms of note are discussed, including design science. In the literature, design science is handled as both a research paradigm and a research methodology. In the case where it refers to a methodology it is called design science research. This section describes the design science research methodology in more detail.

3.3.1 Design Science Research Methodology

Design science research follows an iterative process of expert activities that produce an artefact but also includes evaluation, the result of which informs future iterations by proving better understanding of the problem and improving quality of the artefact (Hevner et al. 2004). In an attempt to formalise the process, Peffers et al. (2007) present a design science research methodology that provides a model for performing research, as well as a mental model for evaluating and presenting research. The use of this process will not only guide the research process but also legitimise it by giving researchers an understanding of the elements on which good research is built.

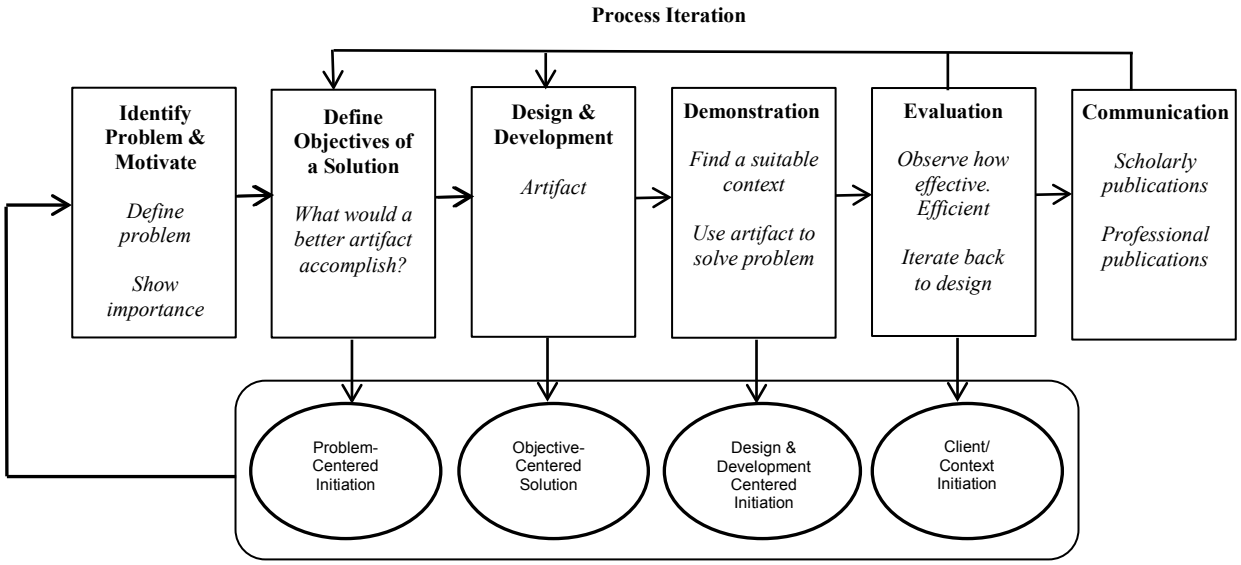


Figure 2: *Design Science Research Methodology Process Model (Peffers et al. 2007)*

The design science research method proposed by Peffers et al. (2007) is depicted in figure 2. The model consists of six steps, the description of which, as given by Peffers et al. (2007), is summarized below.

1. *Problem Identification and Motivation* - Define the research problem and also present a justification of the value of a solution. This justification provides motivation for the pursuit of the solution and gives audiences reason to accept the results of the research. It also gives an understanding of researcher's reasoning with regards to the problem.
2. *Definition of the Objectives for a Solution* - Give objectives of the solutions that will be designed. This is inferred from the problem definition. These objectives can be stated in quantitative or qualitative terms.

3. *Design and Develop* - Create the actual IT artefact by first determining the desired functionality and architecture of the artefact, and then creating an instantiation of it.

4. *Demonstration* - Demonstrate that the artefact can solve one or more instances of the problem. Demonstration can come in the form of an experiment, simulation, case study, proof, or another appropriate activity.

5. *Evaluation* - Observe and measure if, and how well, the designed artefact solves the given problem. This involves the comparison of the objectives of a solution with the observed results recorded during the demonstration of the artefact. The evaluation can take many forms and can include any appropriate empirical evidence or logical proof. After the evaluation has been completed the researcher can decide to go through another iteration of the design process in order to improve the artefact, or proceed to the communication step.

6. *Communication* - Communicate the research outcomes to the appropriate audiences. This includes communication of the problem addressed and its importance, and the designed artefact including the utility, novelty, effectiveness, and the rigour of its design.

Apart from the steps described above, the model also includes provisions for different starting points based on the focus of the research. A normal problem centered approach will start at the first step and move sequentially through the steps, but if the focus is different the starting point shifts, as seen in figure 2.

3.3.2 Design Science Research Guidelines

Another way to inform and evaluate design science research is to make use of guidelines for research presented by Hevner et al. (2004), which state that the guidelines help to distinguish between design science research and system building efforts by helping researchers to successfully represent the problems, solutions, and solution processes which clarifies the knowledge produced. It is up to researchers, reviewers and editors to determine how and when to apply these guidelines, as the use is very dependent on the specific study in question. As long as each element is addressed in some manner, the research should present effective and complete results (Hevner et al. 2004). An explanation of each guideline, as described in Hevner et al. (2004), is given below:

1. *Design as an Artefact* - Any design science research should deliver a viable artefact that addresses an important and relevant problem. The instantiation of the artefact demonstrates feasibility of both the designed product and the design process followed.

2. *Problem Relevance* - The relevance of the design science research and the artefact is dependent of the relevance of the problem addressed. The relevance of the research and problem is always measured relative to a specific community, in other words, the research and artefact do not necessarily have to have relevance to all communities, just the communities that are relevant to the specific research. The relevant community includes the people who plan, manage, design, implement, operate and evaluate the information system and technologies.

3. *Design Evaluation* - A crucial part of design science research is the evaluation of the designed artefact. The evaluation must rigorously demonstrate the utility, quality, and efficacy of the designed artefact through evaluation methods. The requirements of design inform the evaluation criteria and methods used. Since the design science research process is iterative, the evaluation of a design can be used to inform the next construction phase of the next iteration.

4. *Research Contributions* - Any effective and relevant design science research must provide contributions to the body of knowledge. The research can contribute in different areas, namely: the designed artefact, the foundations of the existing design science knowledge, and evaluation methods and metrics. Proper design science research must provide clear and verifiable contributions in at least one of these areas.

5. *Research Rigour* - Design science research requires the use of rigorous methods in both the design and evaluation processes. By using appropriate methods from the existing knowledge base, rigour can be guaranteed.

6. *Design as a Search Process* - The design of an artefact is basically a search process with the aim to find an effective solution to a given problem. This is seen in the iterative nature of the design science process.

7. *Communication of Research* - Without the effective communication of research, it will not fulfil its purpose, which is to contribute to the body of knowledge. The research must be communicated in such a manner that it will be valuable to both technology and management orientated audiences. For the technology orientated audiences enough detail should be included that the artefact can be constructed and used within an appropriate context. This enables the use of the research both by practitioners, who will make use of the artefact, as well as researchers, who will build on the research with further extensions and evaluations. For managerial orientated audiences, emphasis should be placed on the importance of the problem addressed and the novelty and effectiveness of the proposed solution.

3.4 Research Design

When considering the underlying research goal, namely to create a set of guidelines for design of error messages in digital systems, the design science paradigm is most suitable. The underlying philosophical assumptions, which are in line with that of design science, discussed in section 3.2.4, are as follows:

- *Ontological assumption* - The research assumes the existence of a stable physical reality within which a set of design guidelines for error message design can be tested and verified with some degree of generality. In addition to this, other world-states (or views) that are assumed. This includes firstly an objective reality where the guidelines exist, secondly a subjective reality which constitutes the experiences of users surrounding error messages, and lastly an abstract world that include language, theories and constructs which entails not only the theories, models and frameworks that informed the design guidelines, but also the language used in error messages.
- *Epistemological assumption* - The research assumes that the proposed set of guidelines constitutes knowledge. In addition to this, the design and evaluation process followed in the research are informed by, and contributes to the body of knowledge. The validity of knowledge, in this case the proposed guidelines, are proven through a heuristic evaluation process where the guidelines are demonstrated as well as an evaluation process where the results from the heuristic evaluation are compared to the results from a usability study on error messages.
- *Axiological assumption* - The research assumes that the proposed guidelines are valuable in their ability to inform design. Apart from this, the process of creation and validation of the guidelines are also assumed to be valuable.

3.4.1 Research Approach

The research followed the design science research methodology process proposed by Peffers et al. (2007), which is discussed in section 3.3.1. Although the design science research method is iterative, the research only involved one iteration of the process. Subsequent iterations are left for future research. Each research action which corresponds to the step in the process is shown below:

1. *Problem Identification and Motivation* - The problem identified and the motivation thereof was the lack of general design guidelines for error message design in digital systems

that are not bound to specific use cases or environments. The existence of such guidelines will enable designers and developers to design effective error messages without having to conduct large usability studies. This can reduce the effort and cost associated with design and development of digital systems. This is discussed in more detail in section 1.2.

2. *Definition of the Objectives for a Solution* - The objective of the proposed guidelines was to create a single resource that can be used to inform the design of error messages. Secondary objectives that assisted in the completion of the primary objective was:

   - To determine the concerns found in the literature regarding error and warning messages.
   - To translate the literature into a set of guidelines for error message design in digital systems.
   - To determine whether using the guidelines are an effective way to find usability issues with error messages.
   - To determine whether the guidelines are valid.
   - To determine what the relative importance of the guidelines are.

3. *Design and Develop* - The artefact in question is a set of design guidelines for error design in digital systems. This was created by looking to the literature discussed in Chapter 2 and distilling it into a set of well-defined guidelines presented in Chapter 4.

4. *Demonstration* - In order to demonstrate the use of the guidelines on an instance of the problem, the first part of a two-part case study was performed. The second part of the two-part case study was performed in step 5.

   The case that was studied is an existing Internet banking website. The first step in the case study is a heuristic evaluation of the system using the design guidelines (the artefact) as the heuristics. More detail on the process will be provided in section 3.5. The heuristic review demonstrates how well experts can understand and apply the guidelines to determine potential usability problems with the errors and how to address them.

5. *Evaluation* - To evaluate whether the artefact (guidelines) addresses real user concerns, a usability study (in the form of individual interviews) was performed on the same Internet banking website as in the previous step. This is seen as the second part of the case study. The purpose of this usability study is to determine what the pain points and concerns of users are with regards to error messages. The results of this usability study was compared firstly to the design guidelines themselves as a measure

of validity of the guidelines, and secondly to the results of the heuristic evaluation to show whether the guidelines are effective in revealing usability problems that real users experience. More detail on the specifics of this process is discussed in section 3.5.

6. *Communication* - The details of this process, along with the results of the study is communicated in this dissertation.

3.4.2 Evaluation of Research Approach

Apart from the research process steps above it is also important to evaluate the suggested research based on the design science research guidelines proposed by Hevner et al. (2004). A short discussion of each of the guidelines together with an analysis on whether the performed research adheres to it appears below:

1. *Design as an Artefact* - The proposed research delivers a designed artefact in the form of guidelines for error message design in digital systems.

2. *Problem Relevance* - The research includes an analysis of the error messages in an existing Internet banking website. Tan and Teo (2000) shows that, among others, trialability and self-efficacy are important factors that contribute positively to increased adoption of Internet banking. Trialability refers to the ability of users to experiment with an innovation before adoption. If error feedback is optimised and it is clear that errors can be rectified, users will perceive a more predictable environment, and their fears of the unknown can be minimized (Tan and Teo 2000). Self-efficacy refers to user's perceived capability to complete a task (Bandura 1982), which will be improved when the error messages are improved.

   Therefore, the developed guidelines for error message design will assist in the creation of better error messages which can increase the adoption of Internet banking websites.

   Heuristics that inform the design of error message are included in general guidelines of user interface design (Molich and Nielsen 1990). However, existing heuristics for error messages are difficult to apply, and the usability problems of error messages are likely to be especially costly (Nielsen 1992). With a set of guidelines for error message design, a separate, more focused, heuristic evaluation of the error messages can be conducted to find usability problems that are easily missed in general heuristic evaluations.

3. *Design Evaluation* - The evaluation of the design guidelines consists of a case study of an existing Internet banking website, which includes both a heuristic evaluation and individual interviews with users.

4. *Research Contributions* - The contribution to the body of knowledge consists of a synthesis of literature in the form of design guidelines (the artefact) as well as an evaluation proving the relevance and effectiveness of the guidelines.

5. *Research Rigour* - The research was rigorous in terms of both the research approach, which is built on existing design science research framework, as well as the evaluation methods which will also be conducted based on best practices in the literature.

6. *Design as a Search Process* - Traditionally design science research constructs an artefact by searching for a solution through an iterative process. The research presented in this dissertation only involved a single iteration of the design science research method. However, the research did contain a search process firstly in the sense that the literature was used (searched) and synthesized to create the initial design guidelines, and thereafter a more optimal solution was searched for by conducting and analysing the results of two usability studies.

7. *Communication of Research* - As previously mentioned, the research results is communicated in this dissertation.

3.5 Data Collection and Analysis

The research included the completion of three major steps in the design science research methodology, as described in section 3.3.1. The design and development step (step 3 in the design science research method) consisted of creating design guidelines for error messages in digital systems from the literature. More details on this is shown in section 3.5.1.

The demonstration and evaluation (steps 4 and 5) are both performed in the form of a case study, where an existing Internet banking website is analysed. As previously stated, the demonstration was done by performing a heuristic evaluation of the case, and the evaluation was done by performing a usability study in the form of individual interviews of the same case and comparing the outcomes with that of the heuristic evaluation.

The data collection, sampling strategy, and data analysis for each of the above mentioned steps are discussed in the sections below.

3.5.1 Literature Analysis

The first step that was completed in the research was the designing and developing of the guidelines for error design. The data gathered for this are relevant literature (documents)

regarding the error and warning messages. The sampling or data gathering approach was that of a search process to find the most important literary sources. The process started with a keyword search in Google Scholar. The literary sources that were uncovered in this search were added to a list of possibly relevant documents. Documents were then selected based on their relevance judged firstly on the title and then the abstract. Preference was given to well-cited books and peer reviewed journal articles or peer reviewed conference papers published in reputable conference proceedings. Literature in the list of possibly relevant documents was then studied, and if the content of the document was deemed relevant and important, it was included in the literature review.

As each document was studied, the references used were taken and any documents that were deemed possibly relevant were added to the list of possibly relevant documents. A search was also conducted on Google Scholar for any literature that referenced the document being studied. Any possibly relevant documents found in this way were also added to the list of possibly relevant documents. The relevance here was also judged in the same manner as in the keyword search.

This process was iterated where new documents were studied and referenced as well as documents that referenced them were considered. Sometimes as a specific document was studied, a new term or seemingly important concept was identified. These terms or concepts were then also used in a keyword search to identify any other possibly relevant documents. This whole search process was reiterated until no new possibly relevant documents were found.

The analysis of the documents gathered followed a qualitative analysis process of immersion, coding and interpretation. The result of this is a set of design guidelines for error messages in digital systems presented in Chapter 4.

3.5.2 Development of Guidelines

After the literature had been analysed, it was possible to create a set of guidelines for error message design. In the development process the application of the guidelines has to be considered. This informs the structuring of the guidelines.

The proposed guidelines are intended to be used by experts to evaluate error messages by using the guidelines as heuristics, in other words, performing a heuristic evaluation of error messages. A heuristic evaluation is a type of expert-based test, which is described by Lazar, Feng and Hochheiser (2011) as experts using structured methods to find interface flaws. The experts are people who are familiar with interaction design but not with the specific interface being tested. The purpose of this type of evaluation is for the experts to uncover the more

obvious flaws in a user interface. Heuristics are short sets of no more than 10 interface design rules. By keeping the number of heuristics low, ensures that they are intellectually manageable (Molich and Nielsen 1990).

While the proposed guidelines were created, it was found that the literature contained many concerns that could form part of the guidelines, which could potentially deliver a set of more than 10 guidelines. In order to reduce the number of guidelines, guidelines with concepts that are similar could have been merged together into one guideline. The drawback of this, however, is possible loss of specificity in the guidelines which could lead to either concerns in the literature not being properly addressed or to the guidelines becoming too convoluted and confusing, and ultimately hard to apply. As an alternative to this approach, the guidelines were created from the literature and then divided into three major, mutually exclusive, concerns. The result is that each grouping could contain fewer than 10 guidelines each. Due to the mutual exclusivity of the groupings, each group can be used separately from the others, meaning that in total only 10 guidelines (heuristics) need to be used at a time.

3.5.3 Selecting a Case

Since the aim of the study is to demonstrate and evaluate the proposed design guidelines, it is deemed sufficient to use only one case. The selection of the case was done using a purposive and convenient sampling method, in other words, purposefully selecting a case that is a good candidate to be studied and which was also convenient.

The case selected was that of an Internet banking website. An Internet banking website is a good candidate on which to test the proposed guidelines, because the use of Internet banking is not uncommon, and an Internet banking website stands to benefit from more effective error messages as highlighted in section 3.4.2.

In this dissertation the anonymity of the bank whose Internet banking website was selected as a case is kept by cropping out any obvious branding marks from all images of the website. The specific identity of the bank is not an important factor that could affect the results of the study. In the usability test, however, the full screen was given to experts and was presented to users during the individual interviews. This was done in order to keep the ecological invalidity as low as possible.

In the heuristic evaluation and individual interviews (in other words, demonstration and evaluation steps in the design science research methodology), the same 10 error messages were used. The data gathering process for the error screen shots and the details surrounding them were gathered with the assistance of the Internet banking website developers and representatives. This started with a meeting with some of the stakeholders and develop-

ers of the system. In this meeting, the purpose and aim of the research was explained to the stakeholders and high level information of the system was gathered via an informal discussion around the display and origin of error messages in the Internet banking system. Following this, a day was spent at the developers' office where a systematic analysis was conducted of a recent version of the Internet banking website. Also at the developers' office informal interviews were conducted with developers to gather insights into the causes, system design considerations, and mitigation of error messages.

It is expected that the more error messages that are used in the demonstration and evaluation of the proposed guidelines, the better the results will be and the more individual guidelines could be verified. However, if the number of errors that has to be tested is not kept to a reasonable number, the demonstration and evaluation will become unnecessarily difficult to complete, mostly due to the expected unwillingness of users to spend excessive amount of time as research informants. The selection of the error messages used in the research was, therefore, selected to span multiple distinctly different scenarios while keeping the total number as low as possible.

In the initial analysis of the Internet banking website, more than 40 error messages resulting from unique interactions were gathered. This was too many to use in the research. The error messages were compared to one another and 10 error messages were selected that are in a certain way different from the rest of the selected messages. This ensures that the breadth of research is kept and that most of the individual guidelines could be evaluated, but at the same time the total number of error messages to be tested is kept to a manageable number.

3.5.4 Heuristic Evaluation

In the research three heuristic evaluations were performed by three different experts. Two of the experts are academics, both with PhDs in human computer interaction (HCI) and usability related topics. The third expert is a user experience (UX) practitioner with more than 10 years' experience.

To assist the experts in the evaluation process two instruments were created. The first is an evaluation guide which includes all the details necessary to complete the evaluation process. This includes the design guidelines or heuristics that the experts need to use and the details about the 10 error messages that the experts need to evaluate. For each error message a static screen shot is supplied together with a description of the scenario surrounding the error message. This evaluation guide is shown in Appendix B.

The second instrument used was a questionnaire that the experts had to complete. The

questionnaire included a list of the guidelines in the form of questions, space to answer or to indicate the level of compliance with a yes-no selector or a Likert scale respectively, and space to also add comments. This tool assists the experts to follow a structured analysis process. The use of the questionnaires also makes the analysis of the results easier because the different experts each provided their results and suggestions in the same format. The heuristic evaluation form used is provided in Appendix C.

The completed questionnaires constitute the data that was gathered and analysed. The sampling process to select the experts to perform the evaluation was purposive and convenient. Three experts were used in the research, which is believed to be enough to demonstrate the value of the design guidelines.

The purpose of the heuristic evaluation is to show that the proposed guidelines are relevant and useful. In order to do this, a systematic qualitative evaluation of the data was done.

3.5.5 Individual Interviews

After the heuristic review that aims to demonstrate how the proposed guidelines can be applied, the next step of the design science research methodology is to evaluate the guidelines or artefact. Evaluation of the proposed guidelines is necessary because without it, it would be unclear whether the individual guidelines are valid. This is because during the creation process the results of different studies were combined, where most of the research used is concerned with warnings and not with error messages. In other words, using results from studies on warnings and applying them to error messages is not necessary valid, and the combination of the results of different studies will not necessarily result in a valid guideline. An evaluation of the guidelines as a whole is also necessary to be able to confirm the validity of the artefact and not just its parts.

The evaluation process was done by performing a usability study in the form of individual interviews with representative users of the Internet banking website. With the results from this usability study it is possible to validate the guidelines in two ways. Firstly, the results give an indication of whether individual guidelines are valid (in other words, that real users have problems with the error messages that match the guidelines). Secondly, the guidelines as a whole can be validated by comparing the results of the interviews with the heuristic evaluation. This shows whether the use of the guidelines would identify the same usability problems and result in similar suggestions to resolve them.

The instruments used for the usability study includes the same screen shots of the Internet banking website under review that was used in the heuristic review, as well as an interview guide. The interviews were semi-structured and the interview guide contains broad ques-

tions that need to be discussed. Firstly, (and mostly) the contents of the interview guide were informed by the design guidelines. Secondly, the interview guide was informed by the C-HIP model discussed in section 2.2.1, where the effectiveness of the error should be measured on whether interviewees noticed and encoded the error, whether the content was understood, and whether the error will affect the correct behaviour. The interview guide is given in Appendix D.

The advantages of using interviews, according to Lazar et al. (2011) are that it can deliver very detailed information regarding problems in comparison to other methods such as surveys. The trade-off for the depth of information is the increased effort that is required in order to perform interviews and analyse the data. This means that far fewer interviews can be conducted in comparison to surveys.

Selection of participants, similar to the expert selection, was done using purposive and convenient sampling. All the participants were over the age of 18, and did not bank with the bank whose Internet banking website was selected as the case. In total, 14 individual interviews were conducted. This delivers enough insights to be able understand what the user pain points are surrounding the error messages, and is enough to provide a meaningful analysis of the design guidelines.

As part of the individual interviews demographics of the participants were recorded. The demographics include:

- Whether English is the participants' home language. This is important because all the error messages are in English.
- The profession of the participant. This is included to get an idea of whether the participant works in a technical field associated with the development of digital systems.
- Usage patterns of computers, the Internet, and Internet banking websites. This is included to estimate how familiar the participant is with digital systems, specifically Internet banking websites.

The specific questions asked is available in the interview guide in Appendix D.

In total 14 individual interviews were completed. This includes two participants whose home or first language is English. The low number is not unexpected since, according to South Africa's 2011 census, less than 10% of participants listed English as their first language (Anon 2012). Seven of the 14 interviewees have listed a profession that is technical. All the participants have indicated that they use computers and browse the Internet on a daily basis. Seven of the 14 participants have indicated that they use Internet banking on a

weekly basis and the rest have indicated that they use Internet banking on a daily basis. The demographics gathered does not include gender or age because of ethical considerations, but the selection process is done to ensure that the participants includes people of different age groups and genders.

The collected data of the evaluation step consists of the content of the individual interviews. The data was analysed with a systematic qualitative process. The results were first analysed on their own to determine what the pain points and concerns of normal users are surrounding the error messages. Thereafter, the results were compared to that of the heuristic evaluation in order to identify the similarities and differences.

3.6 Ethics

The proposed research has minimal ethical concerns since the participants were not gathered from vulnerable groups, and all participation was done on a voluntary basis. Additionally, the anonymity of all participants is to be kept. To ensure that the study follows the ethical guidelines of the University of Pretoria, ethical clearance was obtained from the ethics board of the University of Pretoria. This letter of clearance is shown in Appendix A.

3.7 Conclusion

The sections above gave an outline of the specifics of the research and what all the research tasks completed are. This was all framed within the design science research methodology. The suitability of the design science research methodology for this specific study was shown and research conducted was shown to follow the requirements for good design science research. The following chapter will show the results of the heuristic evaluation and individual interviews.

4 Guidelines for Error Message Design In Digital Systems

In the previous chapter the methodology of the research is given. The research follows the design science research methodology, and according to it, after the research problem has been identified, motivated, and objectives defined, the next step is to design and develop the artefact. In this case, the artefact is a set of guidelines for error message design in digital systems. In this chapter the proposed guidelines are listed and described.

The guidelines were developed from the literature regarding error and warning design theory as discussed in Chapter 2. The development of the guidelines follows a qualitative analysis process of immersion, coding and interpretation of the literature.

The guidelines given in this chapter was demonstrated and evaluated through two usability studies, the results of which are given in Chapter 5. By using the results of the demonstration and evaluation process the guidelines was updated and enhanced. The updated guidelines are shown in chapter 6.

4.1 Introduction

From the literature review shown in Chapter 2, specifically the section on design (section 2.3), it was possible to generate a set of guidelines for error message design. The process followed to generate the guidelines started with an analysis of the literature review, where all existing concerns surrounding warning and error design were identified. This includes suggestions, theory, and guidelines. The identified concerns were then grouped and, when possible, arranged based on importance. This was then further simplified by grouping similar concerns that could be expressed as a single guideline. This ensures that the minimal amount of guidelines is created. This process (listing, grouping, ordering) was performed iteratively until a final set of guidelines was generated.

This process was performed while constantly considering that the application of the guidelines is error messages, and not warnings, which constitutes the focus of most of the literature used. This resulted in the identification of concerns in the literature that are obviously not applicable to error messages. Along with that, certain concerns, although applicable to error messages, were also excluded to keep the guidelines as concise and general as possible. These items were purposefully excluded from the guidelines and for reference are mentioned in section 4.3.

4.2 Guidelines

According to Lazar et al. (2011) heuristics are a short list of interface design rules, that should not exceed 10 rules. Since the proposed guidelines are expected to be used as heuristics, it is important to keep their number below 10. The ease of use of the guidelines are expected to increase with a smaller number of guidelines in a set, because the number of rules that needs to be remembered and considered at the same time is more manageable. Since the number of unique concerns regarding error message design identified in the literature was more than 10 in total, it was necessary to group the concerns into mutually exclusive groups, each of which has fewer than 10 guidelines associated with them. The purpose of this is that each grouping can be considered separately, meaning that the use of the guidelines will be easier.

This grouping in the proposed guidelines is given as three supreme rules. The supreme rules can be seen as a set of heuristics themselves where an error message should adhere to all three rules. In the case of rules 2 and 3, a set of sub guidelines are given. These sub guidelines are less strict than the supreme rules and are more subjective. The proposed guidelines are discussed in detail in the section below.

4.2.1 Supreme Rules

There are three supreme rules to adhere to when designing error messages. They are listed, in order of importance, as follows:

1. Never show an error message if the cause of the error can be designed out or guarded against.
2. Error messages must be noticeable.
3. The content of the error message must be easy to understand and should communicate the information required.

More information on the specifics and the source of the rules are shown in the sections below.

4.2.2 Supreme Rule 1: Hazard Control Hierarchy

The first rule: “Never show an error message if the cause can be designed out or guarded against”, comes directly from the hazard control hierarchy. This rule states that a warning

should only be shown when it is not feasible to design out the risk or guard against it (Laughery and Wogalter 2006, Laughery 2006, Bravo-Lillo et al. 2011). The same can be said for errors, where the error should only be shown when the action that produces the error cannot be designed out or guarded against. The result of this strategy will be a reduction in the number of error messages in a system. This will have a positive effect on habituation, which will result in more effective errors across the board. Therefore, this first rule is seen as the most important of the supreme rules.

An example of designing out the cause of an error would be removing a required field from a form that a system does not need. Guarding against an error would be not allowing invalid characters in a numeric field or using date pickers when filling in date fields.

4.2.3 Supreme Rule 2: Noticeability

The second rule has to do with how noticeable an error message is. This rule is more important than the third rule, because without actually noticing the error message, the user cannot be expected to understand the content and, therefore, it cannot affect the user's behaviour. This is in line with the C-HIP model discussed in section 2.2.1.

In order to make an error message noticeable, the guidelines below are suggested:

1. Make sure error messages are visually conspicuous. An error message can be made more noticeable by:
   (a) increasing the relative size,
   (b) using more contrasting colours, and
   (c) increasing the font size and weight (Laughery and Wogalter 2006).
2. Error messages should be placed near the cause of an error in terms of location and time (Laughery and Wogalter 2006).
3. Where applicable, use immediate in-line validation (Al-Saleh et al. 2012). Immediate in-line validation is a design pattern where error messages for form fields are shown immediately after the field has been entered and showing the error in line with the form field. The opposite of this would be showing error messages at the top of the page only after a form submission has been attempted.
4. Use pop-overs or modal views only when it is necessary to stop the user from proceeding with the current task or where the error message is critical enough to interrupt the current task flow.

4.2.4 Supreme Rule 3: Understandability

The third supreme rule has to do with the understandability of the error content. Without the user being given the required information in an easy-to-understand format, the error message will not be successful in its communication or in its ability to affect the user's behaviour.

The following guidelines, in order of importance, can be used to create understandable error messages:

1. Use a kind and helpful tone, never blaming the user but rather assigning blame to system deficiency (Kelkar et al. 2013, Molich and Nielsen 1990).
2. Use uncomplicated language. Examples of what to steer away from includes long sentences, complicated sentence structures, complicated or compound words, and the unnecessary use of conjunctions (Harbach et al. 2013).
3. The error message content should include details about:
   (a) what the effects of the error is,
   (b) what the cause of the error is, and
   (c) how to recover from the error (Kelkar et al. 2013).
4. Steer away from technical terms that the target audience might not fully understand (Bravo-Lillo et al. 2011, Harbach et al. 2013). If it is not possible to remove technical terms from the error, make sure to keep possible misconceptions about the technical terms in mind when creating the content (Bravo-Lillo et al. 2011, Krol et al. 2012). An example of this is misconceptions users have about phishing and computer viruses which lead to the misunderstanding of security messages (Krol et al. 2012).
5. Ensure that the error message is unambiguous (Duarte et al. 2012).
6. Make the error message as short as possible without compromising on the aforementioned guidelines.
7. Add icons where applicable, focusing on semantic distance (how close the icon is to the function it represents) or using well known icons (Kelkar et al. 2013).

4.3 Warning Design Topics Excluded

Some topics discussed in the literature review in Chapter 2 on warning design theory are not specifically mentioned in the design guidelines for error message design as mentioned

above. They were purposefully excluded from the guidelines for different reasons. The concerns excluded from the guidelines, and the reasons why they were excluded include:

- *Social Context* - This is the effect seen where people tend to follow the warning compliance behaviour of those around them (Laughery and Wogalter 2006, Edworthy and Dale 2000, DeTurck et al. 1999, Wogalter et al. 1989). This was excluded because interaction with error messages on digital systems, unlike warnings, are typically experienced individually.
- *Trust* - The trust that is placed on a warning can affect users' compliance decision (Egelman et al. 2008). Although this effect that is seen for warnings can have an effect on error messages, it is excluded to keep the guidelines more general, and more focused on the error message itself instead of the design of the greater digital system.
- *Habituation* - Although habituation is not specifically mentioned as a separate guideline, by following the above mentioned guidelines, the quality of error messages should be increased which will reduce habituation.
- *Signal Words* - Signal words in warning messages are an important factor to consider. Error messages, however, do not have a similar set of commonly understood words associated with them to indicate severity. It was, therefore, excluded from the guidelines.

4.4 Conclusion

An important step in the design science research methodology is the development of the artefact. In this case the artefact is a set of guidelines for error message design in digital systems. These guidelines were developed by considering the literature on warning and error design theory. The literature that informed the development of the guidelines is discussed at length in Chapter 2.

To minimize the number of guidelines that have to be considered at the same time, the guidelines were divided into three mutually exclusive groups, referred to as supreme rules. This reduces the mental burden required to use the guidelines, thereby making them easier to use. To ensure that the design of an error message is optimal, the error message should adhere to all three supreme rules. The details of each of the supreme rules are presented in the sections above, and in the case of rules 2 and 3, secondary guidelines are presented that can be used to analyse an error message's compliance to the supreme rules separately.

In the design science research methodology, after the artefact is created, the next step is to demonstrate, and then evaluate the artefact. In the research, both these steps involved a usability study of the same case. Firstly, the demonstration step involved a heuristic evaluation by three experts of 10 error messages that can be generated in an existing Internet banking website. The evaluation step involved the completion of individual interviews with representative users where the same error messages used in the demonstration step was discussed. The results of both these usability studies is then used to determine the effectiveness and the validity of the guidelines.

The results of the demonstration and evaluation steps are shown in the following chapter. The updated guidelines that follow from the analysis of the results of the demonstration and evaluation steps is shown in Chapter 6.

5 Results

In the previous chapter the proposed guidelines for error message design in digital systems are presented. The development of the guidelines, or artefact, constitutes the third step in the design science research process. This follows after the research problem definition and motivation. The next step after the development of the artefact is the demonstration step, followed by the evaluation step. Both of these steps are covered in this chapter.

5.1 Introduction

Both the demonstration and evaluation steps in the research involved a usability study of the error messages of an existing Internet banking website. To demonstrate the use of the guidelines, three experts were asked to evaluate 10 error messages that can be encountered in an existing Internet banking website by using the proposed guidelines as heuristics. The purpose of the demonstration step (heuristic evaluation), was to determine whether the proposed guidelines are understandable and can be used to find potential usability problems with error messages.

To facilitate the evaluation process, the experts were supplied with an evaluation guide and a questionnaire to complete for each error message. The evaluation guide included all the details necessary to complete the evaluation process. This included the design guidelines or heuristics that the experts had to use and the details about the 10 error messages that the experts had to evaluate. For each error message a static screen shot was supplied together with a description of the scenario surrounding the error message. This evaluation guide is shown in Appendix B. The questionnaire included a list of the guidelines in the form of questions. This tool assisted the experts by allowing them to follow a structured analysis process. The use of the questionnaires also made the analysis of the results easier because the different experts each provided their results and suggestions in the same format. The heuristic evaluation form used is provided in Appendix C.

The evaluation step involved completing 14 individual interviews with representative users, where the same 10 error messages used in the demonstration step were discussed. The purpose of the individual interviews is to determine what the pain points and concerns of real users are regarding the error messages in the Internet banking website being tested. This, on its own, is expected to give an indication of whether the proposed guidelines reflect the concerns of real users. This is an important question to answer, because many of the guidelines originated from the literature on warning design theory that was not originally intended for error messages. Therefore, the validity of the guidelines has to be established.

Apart from a separate analysis of the results of the heuristic evaluation and individual interviews, the results of both studies are also compared to one another. This is possible because the same set of error messages was used in both studies. The purpose of this analysis is to determine how effective the guidelines are at finding valid usability problems. In other words, in the cases where both studies identified the same issues, the guidelines can be said to be effective. In the cases where the two studies identified different issues it points to areas where the guidelines are either lacking or are invalid.

The sections below firstly present the results and analysis of these usability studies described above. Thereafter, an analysis of the proposed guidelines based on the results and analysis of the usability studies is given.

5.2 Error Evaluations

In this section the results of the two usability tests performed are shown. The results, and analysis thereof, are shown separately for each error message that was tested. For each error message, the results of the individual interviews, the heuristic evaluation, and a comparison of individual interviews and heuristic evaluations is given. After the results of all 10 error messages, a discussion around more general observations is presented.

5.2.1 Login Failure

When a user tries to log in to the Internet banking website and enters an incorrect password or CSP (customer selected pin), the user is presented with the error page, seen in figure 3. The error is shown in red above the login form with the text: “The details you have entered are incorrect. Please re-enter and submit it again”.

Individual Interviews

The error message was easily noticed by twelve of the informants, only two informants struggled to notice it right away and one informant did not notice the error at all. The case where the message was not noticed was mainly due to the informant being distracted by the other elements on the screen.

The informants who noticed the error understood what caused it and what was expected from them (that they should re-enter their credentials). Two informants, however, indicated that they would have liked more information on how to reset their password and CSP if they forgot them.

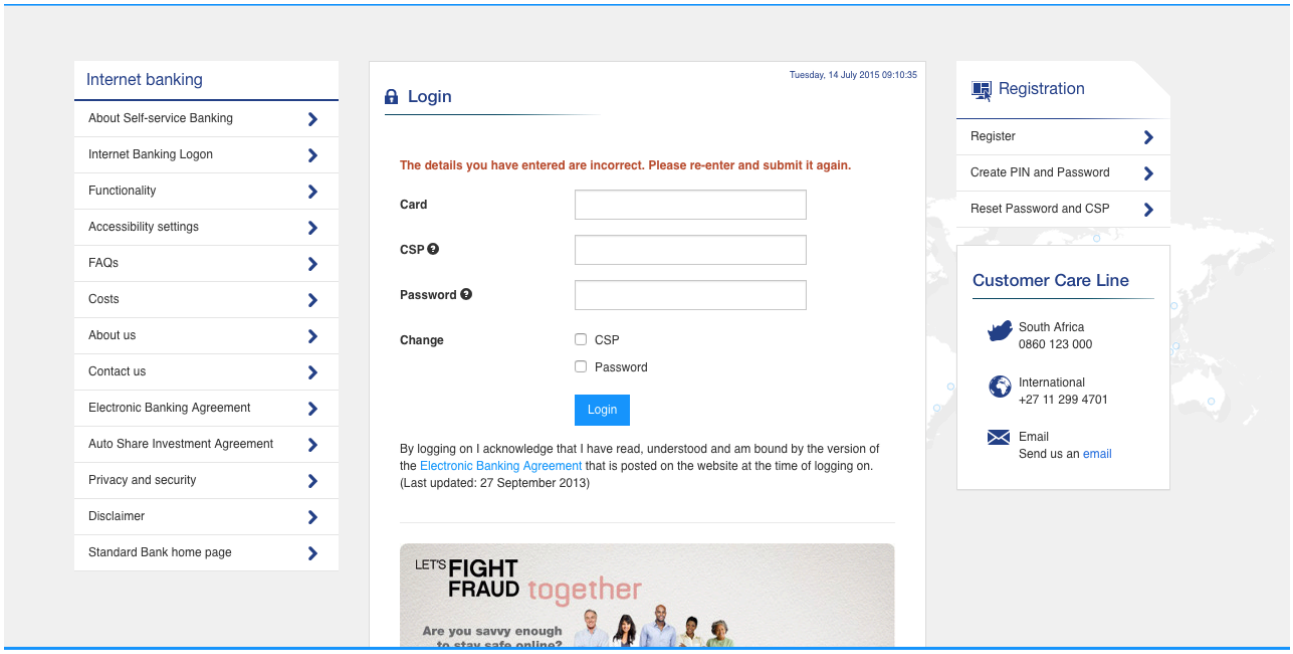


Figure 3: *Login error page with text: "The details you have entered are incorrect. Please re-enter and submit it again."*

Nine informants also complained about the explicitness of the cause of the error message indicating that they would have liked to know which field was incorrect. Five of the more technical informants understood that is it bad practice in terms of security to give information on which field was entered incorrectly, but they still mentioned that they would have liked the information. For example:

> "I understand that it can't quite say which one of them [is] incorrect, but I would really like it to." - Informant 7

In terms of general form layout and usability, four users were upset that the form cleared all entered data and that they have to enter it again. Two informants also thought the form layout was a bit confusing and five informants mentioned that they did not know what a CSP was.

From the interviews it is suggested that the error should be made more prominent to make it stand out from the rest of the elements on the page. The messages should be more explicit on what can be done to recover from the error including how to reset or recover your credentials. It is also worth mentioning that although the cause of the error cannot be made explicit due to security concerns, most users would have preferred it to say which fields were incorrect.

Heuristic Evaluation

The experts agreed that the error message is necessary and that it gives the right amount of details, keeping security requirements in mind. Two of the experts suggested that the error message is not noticeable enough and should be made more prominent by increasing the font size as well as by using a standard error icon. One expert also indicated that the language might be confusing and that the text should mention the login button rather than the term “submit”.

Comparison

The individual interviews and heuristic evaluation both indicate concerns around the visibility of the error relative to its surroundings, but the heuristic evaluation fails to reveal shortcomings in the explicitness of ways to recover from the error message, such as, adding details regarding the password and CSP reset procedure.

5.2.2 Session Expiry

The error given in figure 4 is presented to a user after their session expires due to inactivity. It is shown on a new page with the text: “Your session has expired. If you are not taken to the logoff page automatically click the logoff button”. An important aspect of this error is that it is not shown as soon as the session expires. This message only appears when a request to the website is made after the session has already expired. For example, when the session times-out while taking too long to fill in the new beneficiary form, the error page will be shown only after the form submission has been attempted.

Individual Interviews

The informants understood the necessity of their sessions expiring and security implications thereof, but they were unhappy that they were not given a warning before the session expired. Ten of the informants indicated that they would have preferred it if a message was shown before the session expired giving them the option to prevent the session from expiring. It was suggested by six of the informants that the message should be shown with a countdown timer showing the remaining time before session expiry.

Nine of the informants were upset that the data they entered would be lost. For example, one informant said:

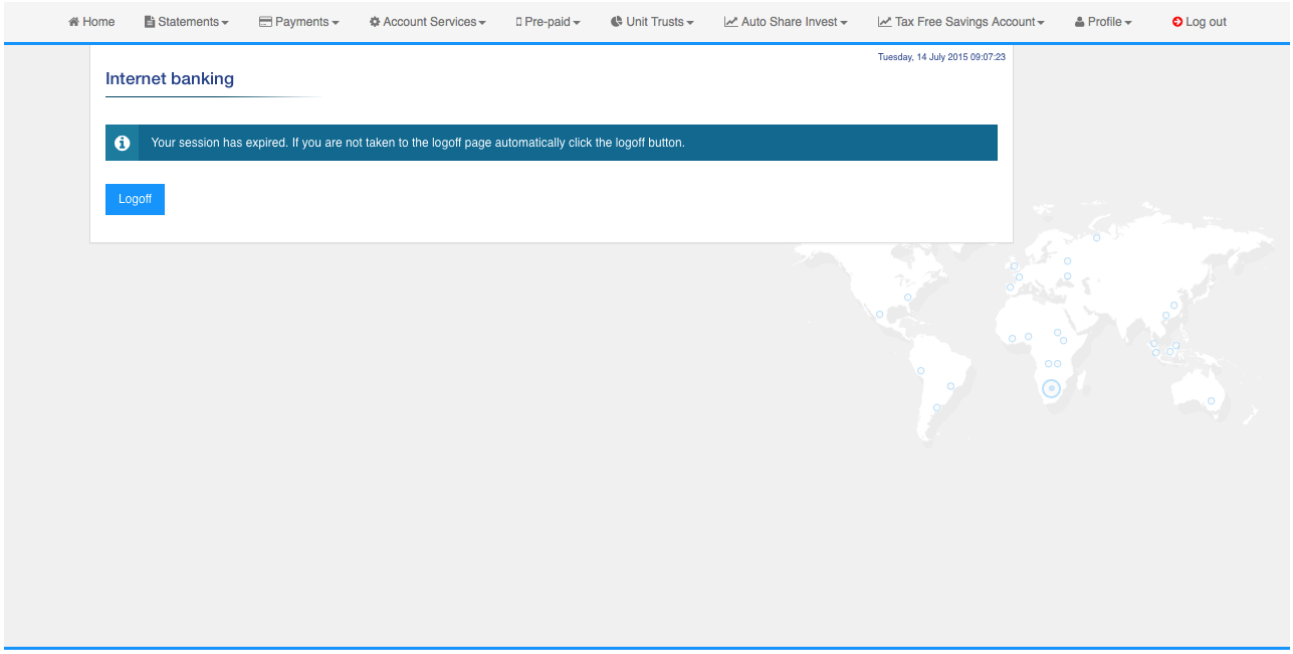


Figure 4: *Session expiry page with text: "Your session has expired. If you are not taken to the logoff page automatically click the logoff button."*

> That's very bad. That is not something you would want. Especially when you entered a bunch of data. *Translated from: "Dis baie bad, nee dit wil 'n ou nie hê nie. Veral nie as jy 'n klomp data ingevul het nie."* - Informant 8.

To address this issue, two informants suggested that the system should remember the entered data and give an option to continue the next time you log in.

The informants were familiar with the concept of session expiry, comprehended what happened and understood what was expected from them. Three informants, however, were unsure whether they were already logged out or not. This is mainly because session expiry was equated to being logged off, which is technically not true, although the effect is the same.

Four of the informants also indicated that they would prefer to be taken to the login page or have an option on the error page to log in again. For example, one informant said:

> [The system] should not just kick me out, [it should] also give me the option to come back, you know. *Translated from: "Hy moet my nie net weg jaag nie, maar die opsie gee om terug te kom, (jy) weet."* - Informant 4.

From the interviews it is suggested that the error should be guarded against by giving the user a warning before the session expires with a countdown timer, and by providing the

opportunity to delay the session expiry. Due to the confusion between session expiry and being logged out, the users should either be taken to the login page or a login form should be added to the error page. Additionally, to reduce the severity of the error, and implicitly design out a part of it, the system could be extended to prevent data loss when the session expires.

Heuristic Evaluation

The experts agreed that the error message is necessary and noticeable. The one expert suggested the use of a pop-up warning with a timer to warn the user of session expiry before it times out.

In the analysis of the error content, one expert indicated that the error might cause confusion between the terms “log off” and “log out” at the top of the page. A second expert indicated that “session expired” might be too technical. A third expert indicated that the content of the error message might be improved by including details on why session expiry is necessary and updating the error to give instructions on how to log in again.

In the analysis of the error content, one expert indicated that the error might cause confusion between the terms “log off” and “log out” at the top of the page. Another expert indicated that “session expired” might be too technical. Yet another expert indicated that the content of the error message might be improved by including details why session expiry is necessary and updating the error to give instructions on how to log in again.

Other enhancements suggested by the experts included extra detail in the error message on how long a user can be inactive before session expiry, and that the option to log in again should be added to the page.

Comparison

The individual interviews and heuristic evaluation both reveal that a session expiry warning should be added to prevent the error scenario. Both also reveal some inconsistencies in the wording used in the error message and that session expiry can be too technical for some users. The addition of a login form on the error page was suggested.

The only difference between the results of the two usability studies is the suggestion that the system should prevent data loss when a session expires. This is mentioned by two informants in the individual interviews, but is not mentioned by any of the experts.

5.2.3 Payments Above Monthly Limit

The Internet banking system used in the evaluation included a monthly payment limit assigned to a banking profile. When an attempt is made to pay beneficiaries with a total above the monthly limit, an error is received. An example of a page with this error message is shown in figure 5. The error is placed inside a red box above the area where the payment amounts are specified. The error text of this example is: "You have exceeded your EAP limit by R1 000.00. Please visit your branch to increase your limit or reduce the amount you want to pay", where the amount in the message is variable based on the payment amounts.

Individual Interviews

Nine of the informants were very upset that the only way to increase their monthly limit was to visit the branch. This frustrated them and as one pointed out:

> I don't want to go to a branch at all. I specifically use Internet banking so that I don't have to go to a branch. *Translated from: "Ek wil glad nie na 'n branch toe gaan nie. Ek doen juis internet banking sodat ek nie na 'n branch hoef te gaan nie."* - Informant 14.

The most prevalent change suggested by the informants was to add the ability to change limits on the Internet banking website. Eleven of the 14 informants suggested this. Four informants also indicated that they would welcome options like sending a message or phoning the bank to increase their limits. Nine of the informants mentioned that they would have liked more information on different ways to change their limit.

Another problem that was frequently pointed out was that the informants did not know what EAP meant. Six informants mentioned this. This was also apparent in the fact that three informant assumed the error referred to the daily limit instead of the monthly limit. Although most of the informants were otherwise fine with the error message, three informants did indicate that the error should be shown immediately if you are over the limit, not after the form was submitted.

From the interviews the first suggestion is to be more explicit in the methods given to recover from the error, in other words, how to increase the limit. Although options are available to send a message to the bank or make a phone call this is not mentioned in the error message. Secondly, the use of the misunderstood and technical terms in the error message should be changed. Lastly, immediate validation should be used as amounts are entered to inform users if payments are too large.

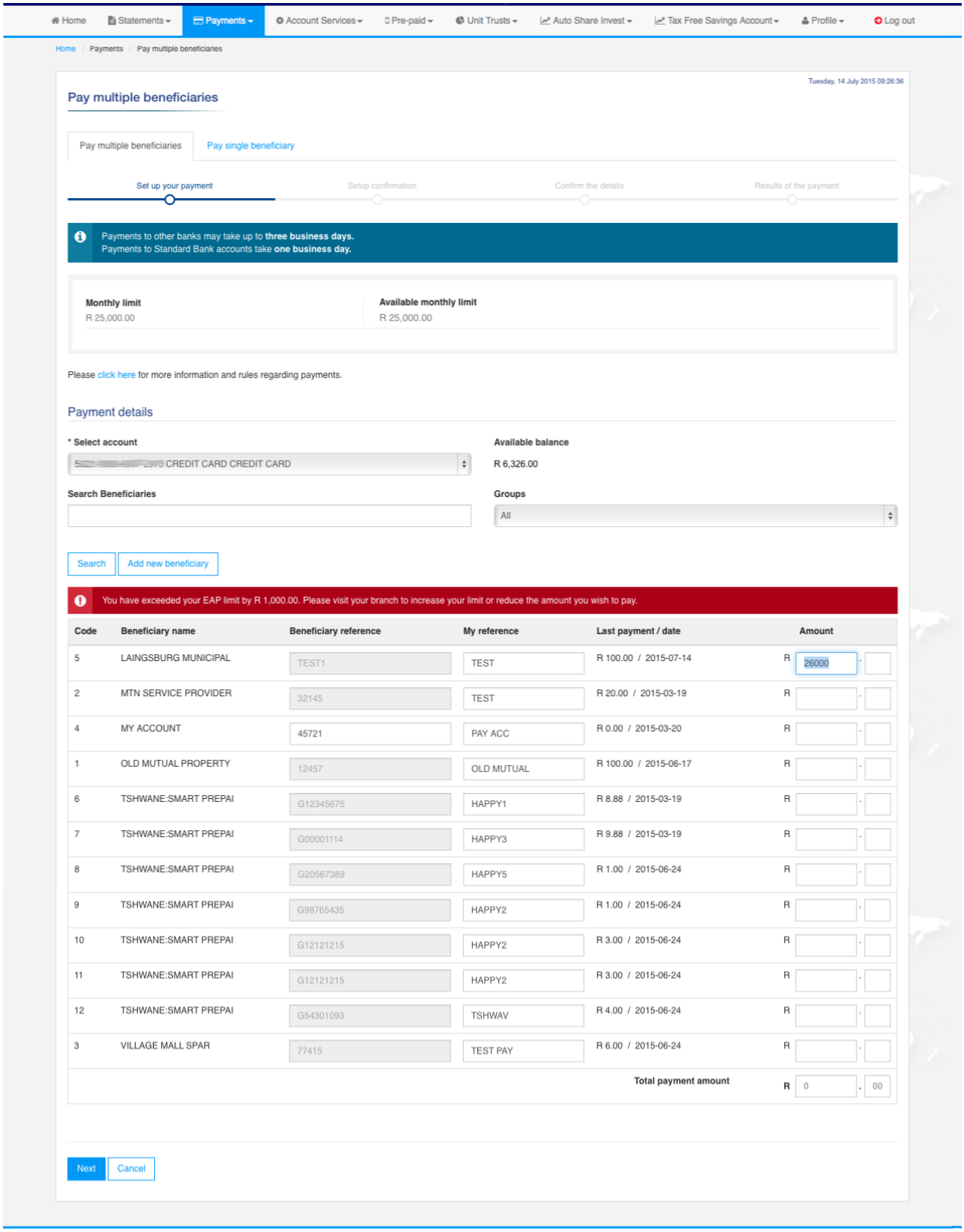


Figure 5: *Payment error page with text: "You have exceeded your EAP limit by R1 000.00. Please visit your branch to increase your limit or reduce the amount you want to pay."*

Heuristic Evaluation

The experts agreed that the error message is necessary and understandable, except for EAP which might not be understood by users. One expert indicated that visibility of the error might be an issue depending on the scroll position, the length of the page and which beneficiaries are used. One expert also indicated that the error message should give more details on how to recover from the error. Two of the experts suggested that a pop-up message should be used, one of which indicated that alternatively, the error should appear next to the input fields.

Comparison

The biggest problem that the individual interviews highlights is the lack of options to recover from the error. This was only mentioned by one of the experts. The individual interviews do reveal that immediate feedback should be given but the interviews do not specifically mention the placement of the error, and the heuristic evaluation indicate that the error should be positioned next to the input field but they do not mention anything about the timing of the error.

5.2.4 Incorrect One Time Password

The Internet banking system used in the usability studies includes a security precaution where a user is sometimes asked to confirm their identity via a one-time password (OTP). The validation is performed by sending an OTP to the user's mobile phone who is then asked to fill in the received OTP on the Internet banking website. When the password is entered incorrectly, the error page in figure 6 is shown. The error message is placed inside a red box above the OTP entry box with the text: "The one-time password you have entered is incorrect. Please re-enter and click Continue".

Individual Interviews

The error message was easily noticed, and because the use of an OTP in an Internet banking system is a common feature, it was easily understood and the informants knew what caused the error and how to resolve it. Even so, one informant commented that the font could be larger and another two informants wanted the error message to indicate what to do when the default OTP delivery method was not available. Five informants also indicated that a competing bank enabled them to accept or reject an authentication request on their cellphones instead of manually entering the one-time password on the website.

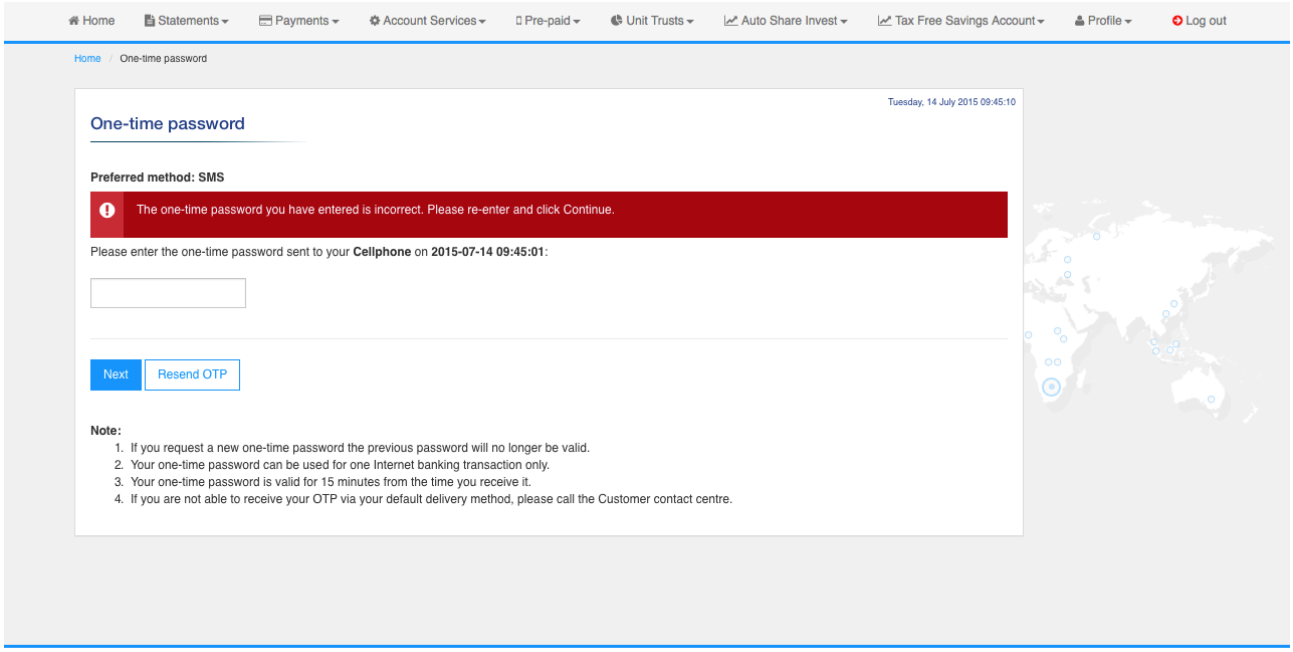


Figure 6: *One-time password error page with text: "The one-time password you have entered is incorrect. Please re-enter and click Continue."*

From the interviews is clear that users would prefer a less error prone solution where they do not have manually enter a password, in other words, that the error should be designed out. Additionally, it is suggested that the font size should be increased.

Heuristic Evaluation

The experts agree that the error message is necessary, understandable and, apart from increasing the font size, visibility should not be a problem.

Comparison

The heuristic evaluation and individual interviews deliver the same comments, except that the individual interviews result in the added suggestion to design out the error scenario.

5.2.5 Insufficient Funds

The error message seen in figure 7 is received when a user tries to transfer more money out of an account than is available. The error is placed in a red box at the top of the page with the text: "Transfer unsuccessful. You have insufficient funds in this account to make the requested payments". It is important to note that this is the third step in the transfer process

where the transfer amount is entered in the first step and is confirmed in the second step. In all three steps, the available funds are visible somewhere on the page.

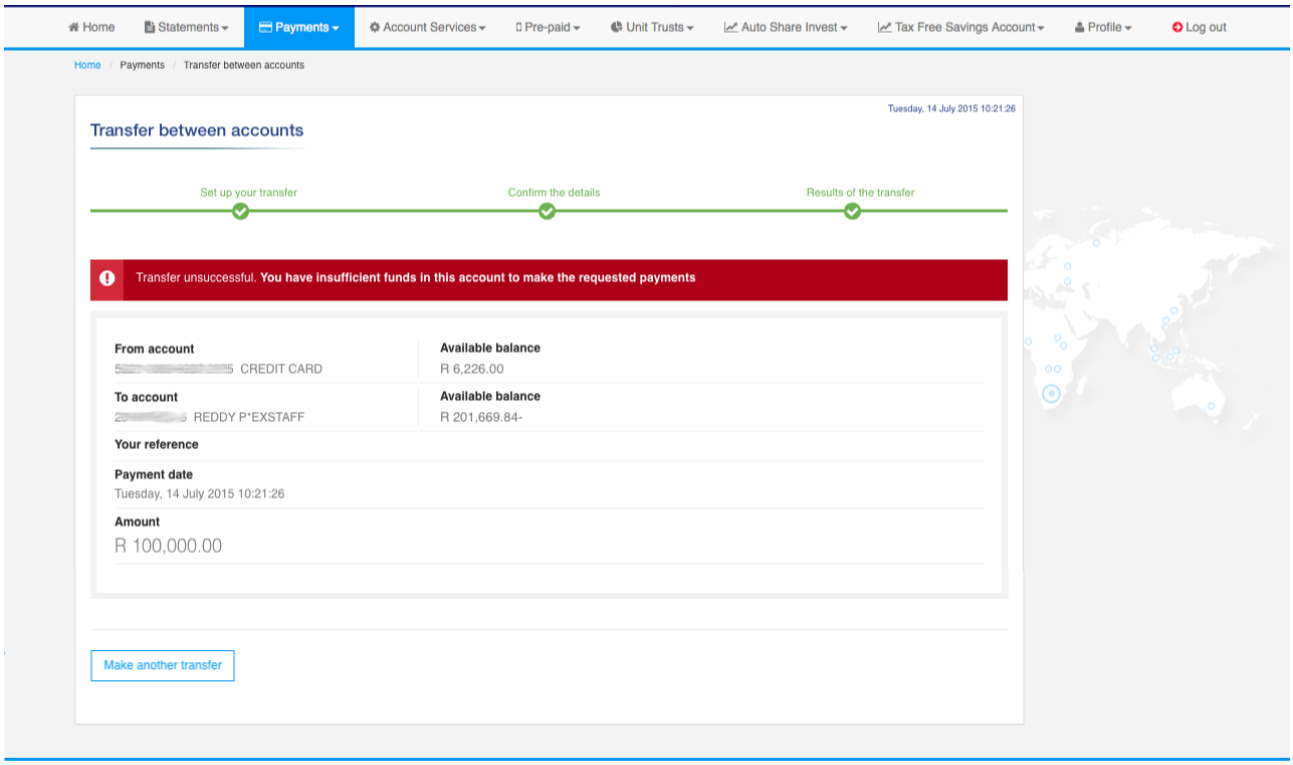


Figure 7: *Inter-account transfer error page with text: "Transfer unsuccessful. You have insufficient funds in this account to make the requested payments."*

Individual Interviews

From the interviews it was clear that the error message was easily noticed by the informants and they understood the content of the error including what caused the error and how to recover from it. Eleven of the informants, however, indicated that they would have liked the error to be shown sooner, five informants said that users should be informed as they are entering the amount to transfer. For example, one informant said:

> "In an ideal world, as I am typing in, something must say: hey don't even click next." - Informant 10.

Apart from this, seven informants also indicated that they would prefer an option to edit the transaction when the error occurred rather than having to start over. One informant also commented that considering the tick mark in the last step in the breadcrumb above the error and the button text "make another transfer", that apart from the actual error message, it appears as if the transfer was successful.

From the interviews it is suggested that the error message should be moved closer to the cause of the error and that immediate in-line validation should be used. This will mitigate all the other issues mentioned by the informants.

Heuristic Evaluation

The experts agreed that the error message should be given earlier in the transfer process and two experts suggested using immediate in-line validation. The experts also indicated that the error message does not give explicit details on how to recover from the error. The experts, however, did not agree on the mitigation of this issue. Two experts suggested the error message should be extended, where the one suggested a text to prompt the user to change the amount, and the other to prompt the user to change the amount or account. One expert commented that users would easily deduce what is expected

Comparison

The individual interviews and heuristic evaluations both result in the suggestion that immediate in-line validation should be used. The individual interviews do provide different examples on how to recover from the error message including, reducing the amount transferring from a different account, or first adding more funds to the account. None of these suggestions are wrong, but none of them are entirely correct either because the mitigation steps would be dependent on the reason why the transfer was made in the first place. Therefore, the comment by the one expert that the steps to recover are easily deduced is believed to be correct based on the results of the individual interviews.

5.2.6 Missing Beneficiary Reference

When adding a new beneficiary, users are required to fill in a beneficiary reference. When payments are made to a beneficiary, the beneficiary reference will appear as a description on the transaction statement of the beneficiary's account.

When the beneficiary reference is not entered while adding a new beneficiary, the page in figure 8 is seen. The error is shown above the beneficiary reference entry field with the text: "You have not entered a Beneficiary reference. Please enter it in the field and try again".

An important interaction to take note of is when users submit the new beneficiary form by clicking the next button on the bottom of the page, when the beneficiary reference is not filled in, the page reloads and is scrolled to the top. This means that the error message is not visible at first.

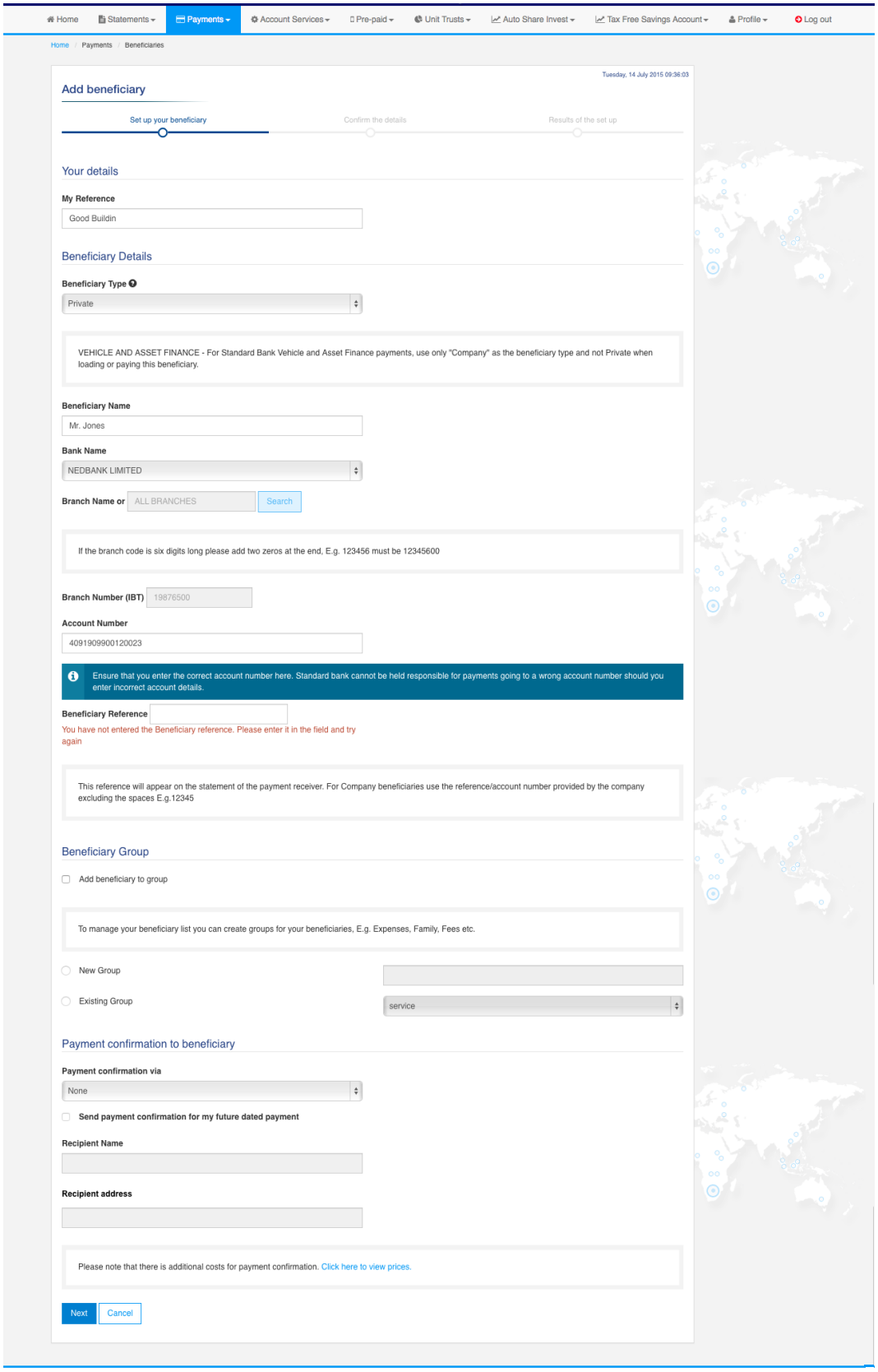


Figure 8: *New beneficiary page with error text: "You have not entered a Beneficiary reference. Please enter it in the field and try again"*

Individual Interviews

This error scenario presented the most problems in terms of visibility because the error message is displayed lower down on the page and is not immediately visible. This caused the informants to become confused and frustrated. Three informants failed to notice the error because they never scrolled down the page. Even when the informants did scroll down the page, they still struggled to notice the error message, where only three informants easily noticed the error message, and one informant did not notice the error at all. From the comments of the informants it is deduced that the error is hard to find due to the general clutter on the page.

The suggestions from the informants to remedy this issue included increasing the visibility of the error by using a bolder font and using a redder colour or adding an error message at the top of the page. Two informant suggested that the same red banner visual style used in the previous error messages should be used in this error. One informant also suggested that it is necessary to add some indication that the field is required rather than relying on only an error message. Apart from this, two informants mentioned that the beneficiary reference should not be a required field when adding a new beneficiary.

From the interviews it is suggested firstly that the beneficiary reference should not be included as a required field. This can be entered when paying the beneficiary. This would constitute the designing out of the error. Secondly, if the first is not a viable option, then visibility of the error should be increased relative to the rest of the page by either decreasing the clutter on the page or by using larger, bolder text and more contrasting colours.

The other significant problem was the fact that the user has to scroll to see the error. This can be remedied by adding a message at the top of the page to list all the errors on the page. When this is done the conspicuousness of the in-line error message would not be as important. Therefore, it is suggested that the error message at the top of the page should use the red banner visual style used in the previous errors, but the in-line error message should not use the red banner visual style.

Heuristic Evaluation

According to the experts, the error message is necessary but problems regarding understandability and visual conspicuousness are highlighted. Two of the three experts indicated that the error is not noticeable enough, where the one expert attributed it to the large amount of other distractions on the page. In order to remedy this, the experts suggested making use of the red banner visual style of previous errors. Although two of the experts did not think the error message was too technical in nature, the one expert indicated that the term benefi-

ciary reference could be better explained. The experts also indicated that the error message should include details on how to recover from the error and that the message could be shorter.

The one expert also indicated that it might be confusing that a beneficiary reference is required since the reference might change between payments to the same beneficiary. The expert also indicates that users might not know what a beneficiary reference is.

Comparison

Although the error message can be designed out by not requiring users to fill in the beneficiary reference when creating a new beneficiary (which is the best solution), it was not mentioned often in either the individual interviews or the heuristic evaluations. The most consistent comments between the two usability tests were that the error message was not noticeable.

The individual interviews and heuristic evaluation reveals that a bolder font and more contrasting colours can be used to increase the visibility of the error message. Although it is suggested by both the individual interviews and heuristic evaluation, the use of the red banner visual style for the error can be problematic considering that there might be multiple errors on the same page. Therefore, it would be more advisable to use the red banner style with a summarizing error at the top of the page. The heuristic evaluations do not reveal the possibility of using an additional error message at the top of the page while the individual interviews do.

5.2.7 Back-end System Communication Error

The page in figure 9 is seen when a user requests a provisional account statement, but the website fails to communicate with the back-end system responsible for generating the statement. The error is placed inside a red box at the top of the section where the requested statement would be shown. The error message has the following text: “Technical Error: Please try again later”.

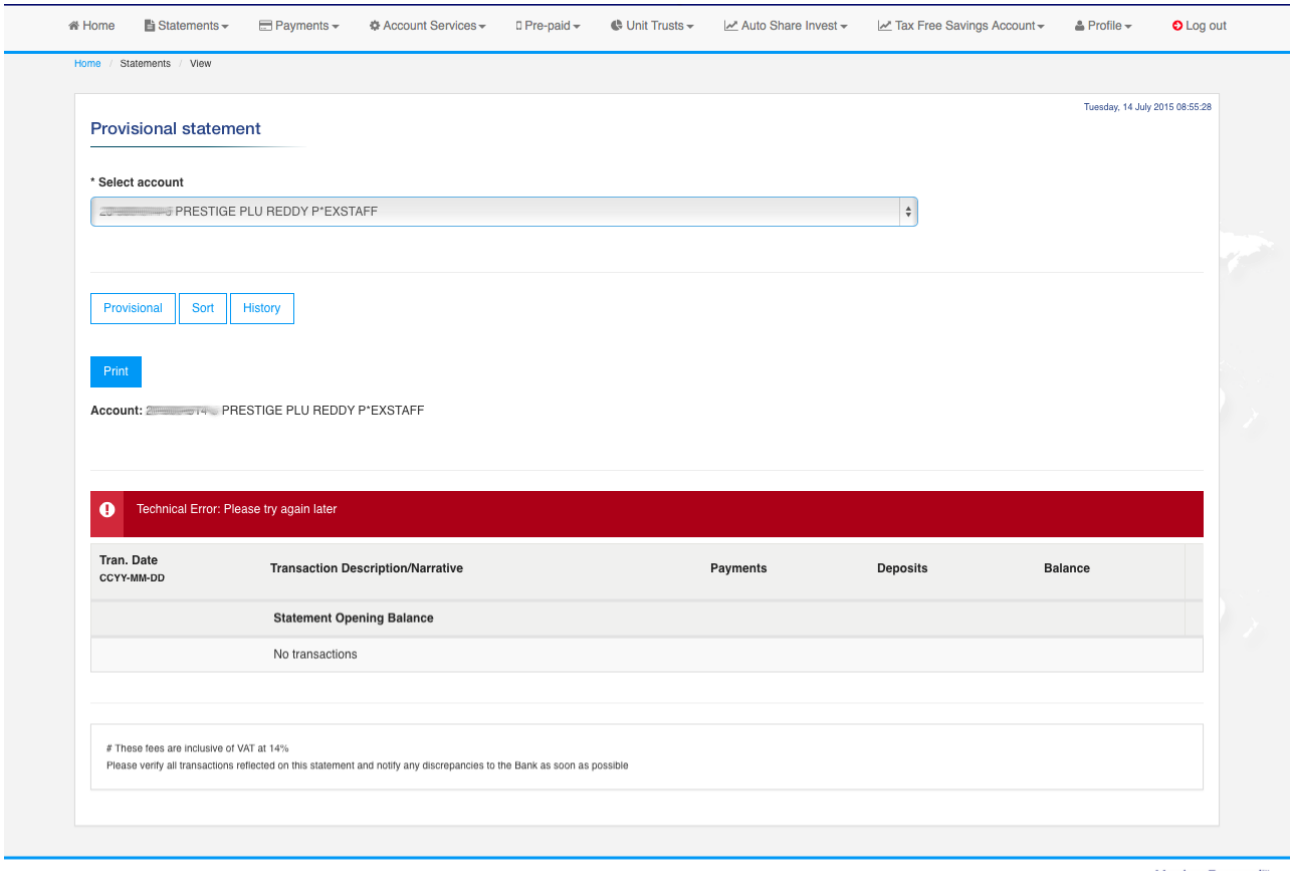


Figure 9: *Provisional statement page with error text: "Technical Error: Please try again later"*

Individual Interviews

All the informants did not like the error message. Seven of the informants specifically indicated that they would like more information on what the cause of the error is. For example, one informant said:

> "Tell me what it is. Is it something I've done, or is it you guys?" - Informant 10.

Five informants also suggested that information should be given on how to report the issue, such as a contact number or reference number.

To recover from the error two informants suggested an additional feature where the requested statement can be emailed to them once the system is back up again. Three of the informants also said that they would prefer it if they were not allowed to use a feature in the first place if it is not functional.

From the interviews it is suggested that the content of the error message should be changed

to be more specific about the cause of the error and give more concrete steps to either mitigate or report the issue. Another viable suggestion is to disable the feature when the system cannot handle the request at that moment. Although this does not technically remove the error but simply moves it in time, it might be better because it could present a less frustrating experience due to expectations being better managed. The additional option to email the statement when the system is back up is a good suggestion that designs out the negative aspect of the error where the user does not know when to retry the request.

Heuristic Evaluation

The experts agreed that the error is necessary and noticeable but that it does not communicate all the required information. The consensus was that the error message does not contain an explicit cause and gives unsatisfactory suggestions on how to recover from the error. Unsurprisingly, the wording was also flagged as being overly technical in nature. Additionally, the one expert suggested the use of a pop-up error message, while the other suggested that a notification that the system is down could be shown before the user even attempts to retrieve a statement.

Comparison

The individual interviews and heuristic evaluation both agree that the error message is lacking in terms of explicit information on the cause of the error as well as on how to mitigate it, and that the error message is too technical in nature. Both usability tests also suggest that a notification could be shown before the user attempts the request. The individual interviews results have the additional suggestion to allow the user to request that the statement should be emailed to the user when the system is operational again.

5.2.8 Repeating Payment Too Far In The Future

The error message presented in figure 10 is shown when a user tries to add a repeating payment for longer than 12 months. The error is shown in red in the middle of the page with the text: “Future dated payment end date is greater than 12 months”.

Individual Interviews

A large proportion of the informants commented that they did not understand why the system would not allow them to add repeating payments for longer than 12 months. Five of the informants mentioned this, stating for example:

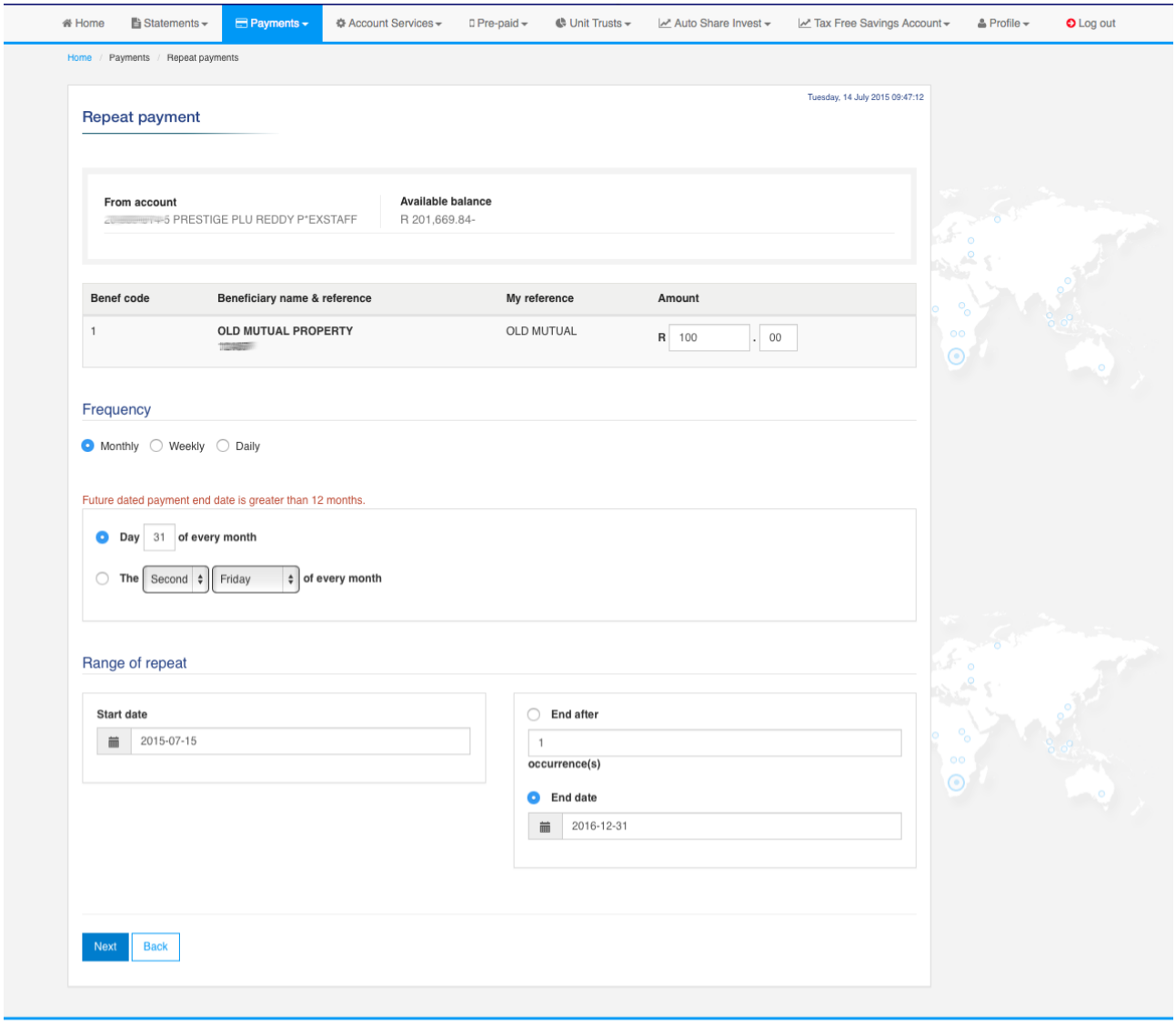


Figure 10: *Repeating payment page with error text: "Future dated payment end date is greater than 12 months."*

> "I would believe that I can schedule a payment for longer that 12 months, it's my money." - Informant 7.

> Why must the bank decide for me how long I can add a repeating payment for? [...] I don't like it when other people think for me. Translated from: "Hoekom moet die bank vir my besluit hoe lank ek 'n repeating payment kan opsit? [...] Ek hou nie daarvan as ander mense vir my dink nie." - Informant 14.

Related to this is the expectations of users which could be better managed. Six informants indicated that they would expect requirements for entry fields to be stated beforehand. Four

informants also commented that they would prefer it if the date picker did not allow them to select the invalid date in the first place.

This error was also not particularly noticeable, where four informants struggled to notice it at first or mentioned that the message is not noticeable. Two informants commented that they thought the error should conform to the red banner visual style and two informants also suggested that the error message should be placed closer to the erroneous entry field.

Other comments included that the error should be shown before the form is submitted, and that the maximum end date should be explicitly mentioned in the error message.

From the interviews it is suggested that the error message should be designed out, enabling users to add a repeating payment for any date range. If this is not possible, the date picker should not allow the users to select invalid dates and the date range and requirements should be clearly stated on the form. It is also suggested that immediate in-line validation should be used. To address the habituation of previous error message visuals it is suggested that the banner style should be used at the top of the page to indicate any errors on the page together with in-line errors for the specific erroneous input fields.

Heuristic Evaluation

The experts agreed that the error can be guarded against by not allowing the user to select invalid dates in the date picker. Two of the experts also indicated that the message is not noticeable and suggested that the error should be moved closer to the date entry field and that the red banner visual style should be used. It was also suggested that an icon should be included. The content of the error was considered to be too short and not clear enough regarding the cause of the error or ways to recover from it.

Comparison

The individual interviews and heuristic evaluation deliver very similar comments and suggestions regarding the error which include not allowing the selection of invalid dates in the date picker, placing the error message closer to the invalid field, and using the red banner visual style. From the individual interviews immediate error feedback is suggested, along with possibly designing out the error by allowing repeating payments for longer than 12 months. The experts suggested using an error icon and indicated that the error message content is lacking.

5.2.9 Transaction History Too Far In The Past

In the Internet banking website used, users have the option to view historical information on transactions that have been made. There are different options available to specify which transactions should be shown, one of which is to specify a date range that will result in transactions that were performed inside the range being shown. The system only supports the viewing of the transaction history for up to 12 months in the past. However, if a user enters dates further back in the past when specifying the date range, an error is received. An example of the page with the error message is available in figure 11. The error message is shown in red at the top of the filters section with the text: “Payment history can only be displayed up to 12 months back”. The error is also only presented after the view button at the bottom of the page is clicked.

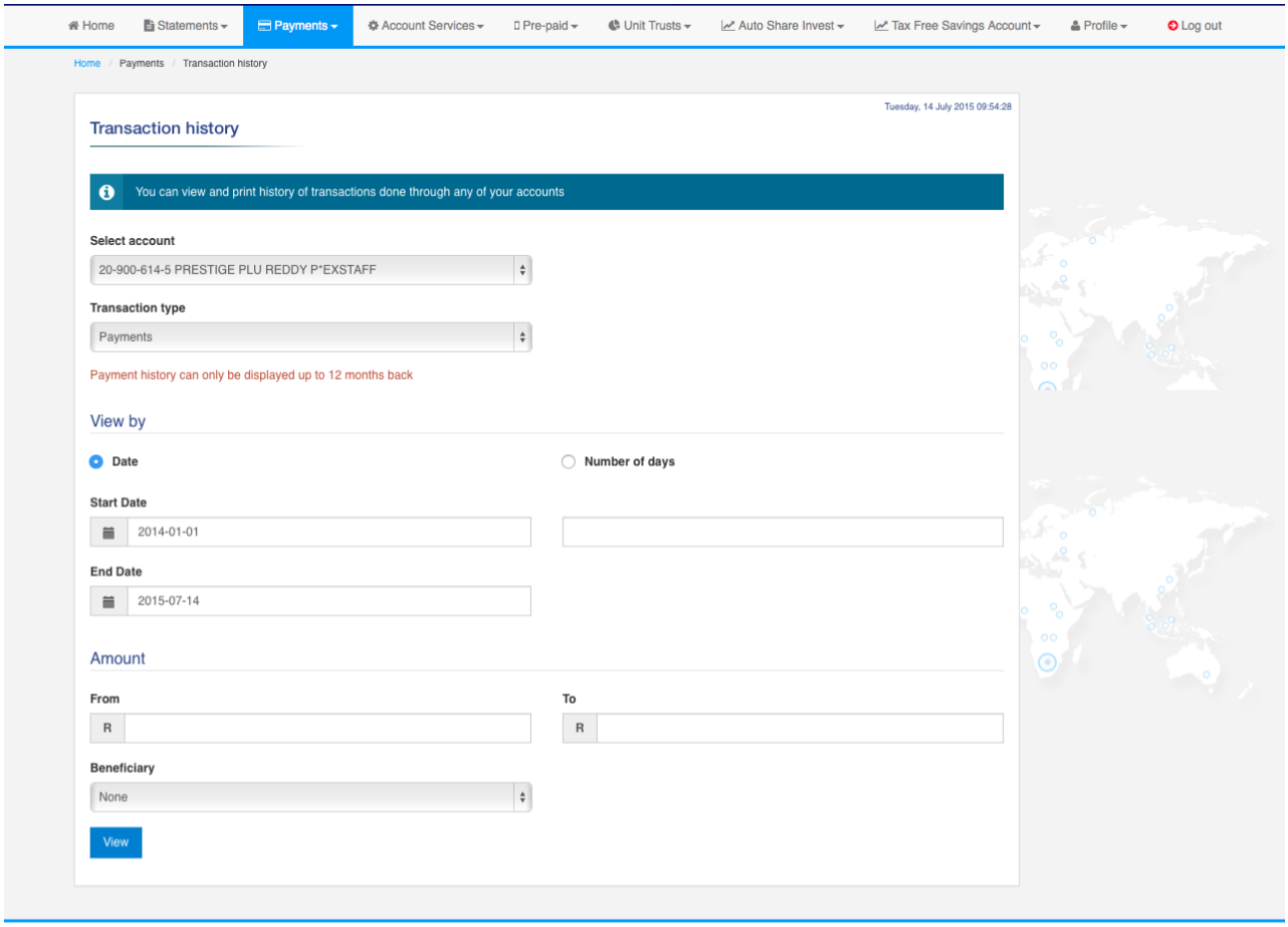


Figure 11: *Transaction history page with error text: “Payment history can only be displayed up to 12 months back”*

Individual Interviews

The most frequent comment made by informants was that they wanted or expected the system to allow them to view transaction history further back. Six of the informants mentioned this. Related to this, the informants also wanted to know why it is not possible and suggested that the requirements should be stated beforehand. Four of the informants also said that the date picker should not allow them to select invalid dates in the first place.

Two of the informants suggested that more information should be given on what to do next like adding information on how to contact someone for assistance. Some other suggestions for mitigation of the error include automatically filling in the largest end date after the error and adding the ability to request the transaction history to be made available via email or at a later time on the website.

Most informants did notice the error without too much effort, but one informant indicated that the error should be made visible before form submission. One informant also suggested that the error should be placed closer to the erroneous form field. For example, the informant said:

> You cannot really see for which text box that [the error message] is for. *Translated from: "Jy kan nie regtig sien watter text box daai voor is nie."* - Informant 5.

From the interviews it is suggested that the error should be designed out by enabling users to search transaction history for any date range. If this is not possible it is suggested that the date picker should be restricted to only allow valid dates. Furthermore, it is suggested that the requirements for the date range should be added as an informational entry in the form. The content of the error message should also be updated to include explicit details on alternative methods to get the requested transaction history and possibly adding features to request it on the website. It is also suggested that immediate in-line validation should be used.

Heuristic Evaluation

Although the experts did find problems with the error message, they each made a different suggestion on how to solve the problem. One expert indicated that the error can be designed out by enabling the system to show data further back. Another expert indicated that the error can be guarded against by not allowing invalid dates in the date picker. Two of the experts indicated that the error can be made more noticeable by using immediate in-line validation and by using the red banner visual style.

The experts agreed that the error content is understandable and, although the cause is not explicitly mentioned, two of the experts indicated that it can easily be deduced by users. One expert also indicated that the error message should include more detail on how to recover from the error.

Comparison

Both the individual interviews and heuristic evaluation reveals the same strategies for designing out and guarding against the error: that the error is not noticeable enough, and that in-line intimidate feedback should be used. The individual interviews do, however, reveal that users would prefer more details on the cause of the error, such as why they cannot see transaction history further back, which is in contrast to the comments from some of the experts. Both studies also reveal the lack of explicit details in the error message on how to recover from the error along with suggestions on alternative methods to mitigate the issue.

5.2.10 Weak Email Encryption Password

Another feature available on the Internet banking website being tested is to schedule additional statements to be emailed to the user. As part of the process users have to enter a password that will be used to encrypt the email. When a user attempts to submit the form with a password that is too weak, the error screens in figure 12 and 13 are displayed depending on the shortcomings of the password. The error is shown in red beneath the password entry field with the either the text: “Your password must be a minimum of six characters long. Please re-enter and continue,” or the text: “Your password must contain a minimum of three letters and numerics. Please re-enter and continue”.

The scenario presented to the users and the experts includes first entering a password with less than 6 characters. This will result in the screen in figure 12. Thereafter, the password is updated to be longer but it does not include any numbers. This results in the screen in figure 13.

Individual Interviews

In general, the informants did not struggle to notice the error message, comprehended the error message and understood what was expected from them. Two informants, however, did not understand why the password was required in the first place. For example, one informant said:

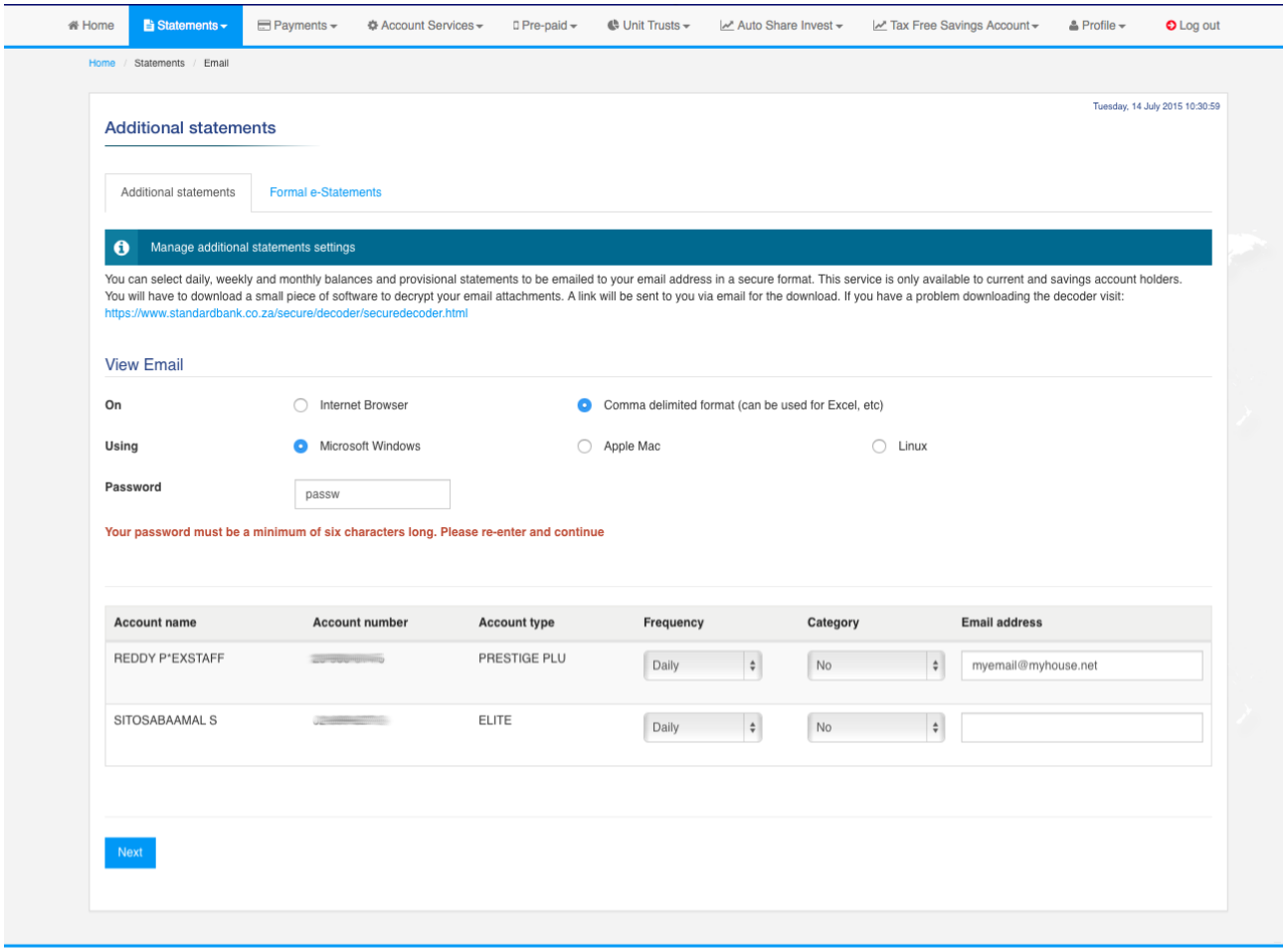


Figure 12: *Additional statement page with error text: "Your password must be a minimum of six characters long. Please re-enter and continue"*

> I would have been frustrated that I have to enter a password, if I must, because I am emailing it to myself. There should not be a reason to stress over it, which makes this message so much more frustrating. *Translated from: "Ek sou dan [ge]frustreerd gewees het dat ek 'n password moet intik, as ek moes, want ek email dit na myself, so daar behoort nie 'n rede te wees om te stres daaroor nie. Wat hierdie boodskap net so veel meer frustrerend maak."* - Informant 3.

The assumption made by this user is that the security of his email account is enough to safeguard his information and that extra encryption is not warranted. Another suggestion made by an informant was to auto generate a password and email it out or secure the email message by using the existing security systems available in the Internet banking system.

The informants were also upset by the fact that all the shortcomings of the password were not shown in the first error. Eight of the informants suggested that as the user types in the password, a message should be displayed next to the entry field that lets them know what

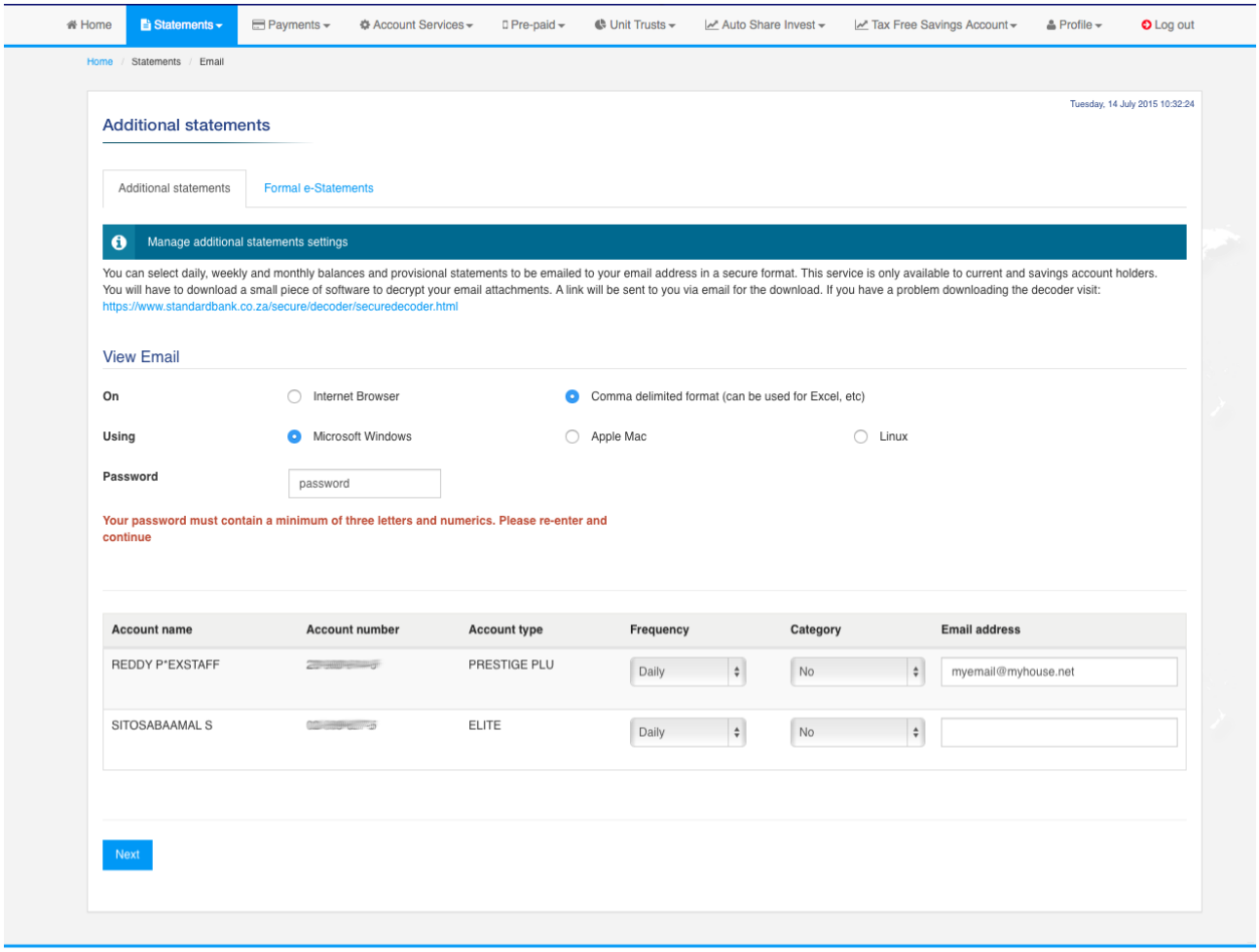


Figure 13: *Additional statement page with error text: "Your password must contain a minimum of three letters and numerics. Please re-enter and continue"*

requirements they have and have not met. Four of the informants also said that they would prefer it if the requirements of the password are shown beforehand.

From the interviews it is firstly suggested that the error should be designed out by not requiring a password. The best solutions from the informants are to firstly not encrypt the email since email security should be sufficient and, secondly, to use another identity verification method that is tied to the existing Internet banking security. If the password is required, it is suggested that immediate in-line validation should be used which shows all password requirements and shows the user what requirements they have and have not met as they type.

Heuristic Evaluation

The experts agreed that the error message content is not optimal and that it should give all the password requirements the first time that it is displayed. The one expert suggested that immediate in-line validation should be used to draw attention to the error. Two of the experts indicated that the error is too technical, with the term “characters” flagged as being problematic. Additionally, two of the experts indicated that the requirements should be given upfront before entry.

Comparison

From the studies it is clear all the requirements of the password should be shown at the same time. Both the individual interview and heuristic evaluation suggest using immediate in-line validation, but the heuristic evaluation also suggests showing requirements beforehand.

The individual interviews reveal concerns about the necessity of the password and ways to design out the error scenario. This is not revealed by the heuristic evaluation. The heuristic evaluation also indicates that the message might be too technical, but from the individual interviews this does not seem to be the case.

5.2.11 Individual Interviews - General Observations

During the individual interviews some observations were made regarding the error messages of the Internet banking website across the site. Firstly, generally the red banner visual styles seemed to be more noticeable in comparison to the form field visual style. This is firstly due to the red banner visual style being larger and using more contrasting colours and, secondly, because the form field visual style was mostly used in pages with more visual elements that cluttered the page and drew attention away from the error. From the two visual styles the informants preferred the red banner style.

In the interviews the informants usually understood that something went wrong, but in the cases where they struggled, it was mostly because they did not immediately notice the error message. In this case, by using consistent styling across the website, users will become habituated to the visual style and be more likely to notice error messages if they are looking for them. Some informants indicated that they would prefer consistent styling of the error messages and one informant, after seeing most of the error pages in the above sections, commented on the form field visual style:

> I am used to the site by now, so maybe it is a usage thing where if you use the site

> more you would know better what colour the error messages are and what you are looking for. *Translated from: "Ek is nou al gewoond aan die site so miskien is dit 'n usage ding: as jy die site meer gebruik weet jy meer wat is die kleur van die messages en waarvoor om te soek."* - Informant 6.

This, of course, would not be true for all error messages. In this case, users would be on the lookout for error messages because they understand that they have to resolve some problem in order to continue with the task at hand. In other systems or scenarios, error messages might be ignored or dismissed without reading due to the negative effects of habituation.

Another observation that was made is that, although the effects of the error message are usually not mentioned explicitly in the error message, users did not struggle to understand the implications of an error message. This is in contrast to a suggestion from warning design theory that states that the consequences of non-compliance should be included in the warning. This was included in the error message design guidelines stating that the effect of the error messages should be stated in the content. With warnings that are displayed before a relevant activity, the effects of non-compliance are not necessarily obvious. Error messages are shown after the relevant action, which means that the effects are usually obvious, such as: the transfer was not made, the email was not sent, or the report is not available.

The informants did not really comment on the tone of the error messages directly, but when the informants would discuss different methods of handling the error some would give suggestions of error messages that can be used. In these cases, the suggestions tended to use a much kinder tone and tended to be much more apologetic than the existing error messages.

5.2.12 Heuristic Evaluation - General Comments

In some cases, the experts consistently made the same comments on most of the error messages. This is excluded in the discussion above to keep it more concise, and is instead discussed here in a more general sense.

One expert indicated that the font size of the error messages could be bigger, specifically for the red banner visual style. Two of the experts also indicated that the red banner visual style should be used more often because it is more noticeable, and that error messages could be politer and use a more helpful tone. One expert raised concerns regarding how noticeable the error would be for the users who are colour blind. All the experts also noted that most of the errors do not explicitly state the effect of the error but two of the experts indicated that it should not be a problem because the effects are clear.

From the evaluation and comments made by the experts it was clear that some concepts used in the guidelines were not explained well enough and, therefore, the guidelines were ineffective. The first was the designing out and guarding against errors. In general, the experts rarely indicated that the error message could be guarded against or designed out compared to the possibilities highlighted by the individual interviews. In some cases, the experts would indicate that the error cannot be guarded against or designed out but would then, in comments on other questions, give examples of how they could. An example of this, was one expert who indicated that the error described in section 5.2.8 could not be designed out or guarded against, but in comments suggested that the date picker should be restricted to only allow the selection of valid dates. The suggestion would technically guard against the error being received because the user would have to manually enter an invalid date to get the error.

In other cases, the experts indicated that an error message could be designed out or guarded against but the comments made gave suggestions of improvements, that is not technically designing out or guarding against the error. For example, in the evaluation of the error regarding insufficient funds discussed in section 5.2.5, one expert indicated that the error message could be designed out or guarded against, but the details in the comments suggested that the system should alert the user earlier in the process that they are attempting to transfer more money than they have. With this suggestion, the error message would only be moved closer to its source, which is a valid suggestion, but it does not technically design out or guard against the error being received.

The distinction made here between designing out and guarding against error messages, and normal usability improvements is very technical and somewhat pedantic, but it points to a problem with the proposed guidelines. This is considered important because the designing out and guarding against error messages is the most important guideline.

Another guideline that seems to present some problems is the guideline around immediate in-line validation. For example, two of the experts both indicate that immediate in-line validation is required for the error received when setting up repeating payments too far into the future. This error is described in section 5.2.8. Although this by itself it not incorrect, both experts' comments on immediate in-line validation is that the date picker should not allow the selection of invalid dates.

Similarly, one expert indicated that a modal error message should be used for the password strength errors discussed in section 5.2.10, but the comments suggested that a tool-tip should appear with the requirements of the password. In this case, the wording in the questionnaire might have been confusing where the term pop-up errors is used instead of modal errors because it was considered to be easier to understand.

5.3 Design Guideline Evaluation

From the interviews and heuristic evaluation, it is possible to analyse the design guidelines themselves and thereby determine how relevant and how effective they are as well as how to improve them. The sections below will comment on the error message design guidelines as given in Chapter 4 based on the analysis above.

5.3.1 Supreme Rule 1: Hazard Control Hierarchy

From the results of the individual interviews it is clear that users prefer interactions that do not have the possibility of causing an error or that the system has something in place to prevent the error. The results of the heuristic evaluation show that it is less likely to find scenarios to design out or guard against error messages when using the guidelines and evaluation questionnaire in comparison to completing individual interviews. Supreme rule 1 can, therefore, be said to be valid but ineffective.

To remedy the ineffectiveness of supreme rule 1 it is suggested that the guidelines should be extended to include a more detailed explanation of designing out and guarding against errors with more concrete examples. It is also suggested that a set of guiding questions should be included that can facilitate the analysis, and creative problem solving process.

5.3.2 Supreme Rule 2: Noticeable

From the individual interviews, it is seen that in the instances where users failed to notice error messages, they were much more confused and frustrated in comparison to the cases where the users gave negative comments on regarding the content of the error. Therefore, supreme rule 2 is seen as more important than supreme rule 3.

The guidelines regarding visual conspicuousness and timing of error messages are considered to be relevant and valid based on the results of the individual interviews. The experts also successfully applied them revealing similar concerns to that of the individual interviews, which means that the guidelines are understandable and easy to apply.

The individual interviews also reveal that users would prefer that error messages should be immediately displayed when filling in form fields. Suggestions to use in-line error messages (the positioning the immediate feedback next to the entry field) are not as frequent or consistent, but a preferred alternative display method is not evident. From this it can be said that the guidelines regarding immediate in-line validation is relevant. As discussed above in

section 5.2.12, this guideline was sometimes not used correctly. In order to solve this, it is suggested that the guideline should be updated to include a better description of immediate in-line validation with more concrete examples on when it is applicable and when not.

The guideline regarding modal error messages could not be directly tested with the specific error messages which have been used in the usability studies because there are no modal error messages or cases where they should have been used. The expert evaluation, however, does reveal problems with the guideline as described in section 5.2.12. It is, therefore, suggested that the guidelines should be updated to include more details and concrete examples regarding modal error messages, as well as the term pop-up error should rather not be used.

The usability studies also revealed two additional guidelines that could potentially be added to the guidelines under supreme rule 2. The first guideline is that consistent styling should be used for error messages throughout a system. This does not mean that all errors should look the same, but that the different classes of error messages should have consistent visual styles. The second guideline is that when a web form includes validation on form fields, a summarizing error should be displayed at the top of the page after form submission listing all the errors for the form fields.

5.3.3 Supreme Rule 3: Understandable

From the individual interviews it can be said that most of the guidelines for supreme rule 3 are valid. The exceptions include the ambiguousness and length of the error message. The specific error messages used in the usability tests do not include messages that could successfully test these guidelines and, therefore, they cannot be confirmed to be valid. Another exception is the guideline stating that the effect of an error should be explicitly stated in the error message. This guideline is deemed invalid based on the results from the individual interviews which reveal that users understand the effect of the error message without it being explicitly mentioned in the error message.

From the heuristic evaluation the guidelines are also mostly understood and applied, revealing similar results to the individual interviews. In the same manner as the individual interviews, the ambiguousness and length is not directly testable. However, from the comments made by the experts it would seem like the guidelines are easy to understand. Regarding the inclusion of the effect of an error, the experts correctly identify the lack of the required details in the error message, but they also comment that it would seem obvious to users.

It is, therefore, suggested that the guideline regarding the effect of the error message should be removed. This is in line with existing guidelines for error messages in the literature (Molich

and Nielsen 1990). Further testing should be conducted to test the validity and to confirm the understandability of the guidelines regarding ambiguousness and length of the error messages.

5.4 Conclusion

This chapter presents results of the usability studies performed on an existing Internet banking website. The primary purpose of these studies is to demonstrate and evaluate the proposed guidelines. This follows from the design science research methodology. Since the design science research methodology follows an iterative process, the results of the evaluation of the guidelines can be used in another design and development step. The purpose of this is to propose an updated set of guidelines that is expected to be easier to use effectively and to be more valid (be closer to the real concerns of users). The next chapter presents the updated guidelines for error message design based on the results discussed in this chapter.

6 Contribution

In the previous chapter the results of the demonstration and evaluation of the artefact being studied (guidelines for error message design in digital systems) is presented. The demonstration of the artefact involved the completion of a heuristic evaluation of 10 error messages that can be produced in an existing Internet banking website. This was completed by three experts. The evaluation step involved completing 14 individual interviews with representative users where the same 10 error messages that were used in the heuristic evaluation were tested.

The results of the heuristic evaluation were used to determine if the originally proposed guidelines are easy to use and effective at finding usability problems with a given error message. The results of the individual interviews gave an indication of how valid the guidelines are (if they reflect the concerns and pain points that users have). The results of both these usability studies were also compared to one another to further contribute to the analysis of the effectiveness and validity of the proposed guidelines.

Based on the results of the previous chapter, an updated set of guidelines for error message design in digital systems is proposed. This follows the design science research methodology process which makes provisions for multiple iterations of the process. The updating of the guidelines can be seen as another design and development step that would constitute the start of a second iteration of the design science research process. The research however does not complete this second iteration.

The updated guidelines and other contributions are described in more detail in the sections below.

6.1 Introduction

This chapter presents the contribution of the research. This comes in the form of guidelines for error message design in digital systems. The guidelines, in its original form as shown in Chapter 4, were evaluated and the results of the evaluation were used to update and enhance the guidelines shown in this chapter.

The sections below include, the updated design guidelines that follows from the results of the evaluation of the originally proposed guidelines presented in Chapter 5. Some parts of the guidelines will match the original proposed guidelines that appear in Chapter 4 in the cases where they were found to be effective and valid. In the sections below, indications are given where the guidelines have been updated.

Another contribution is an updated heuristic questionnaire. This questionnaire is similar to the one supplied to the experts during the heuristic evaluation process. The purpose of the questionnaire is to assist in the evaluation of error messages by providing an easy to follow, repeatable process.

6.2 Updated Design Guidelines

There are three supreme rules to adhere to when designing error messages. They are listed, in order of importance, as follows:

1. Supreme Rule 1: Never show an error message if the cause of the error can be designed out or guarded against.
2. Supreme Rule 2: Error messages must be noticeable.
3. Supreme Rule 3: The content of the error message must be easy to understand and must communicate the information required.

More details on each of these are shown in the sections below.

6.2.1 Supreme Rule 1: Hazard Control Hierarchy

The first rule: “Never show an error message if the cause can be designed out or guarded against”, comes directly from the hazard control hierarchy. This states that a warning should only be shown when it is not feasible to design out the risk or guard against it (Laughery and Wogalter 2006, Laughery 2006, Bravo-Lillo et al. 2011). The same can be said for errors, where the error should only be shown when the action that produces the error cannot be designed out or guarded against. The result of this strategy will be a reduction in the number of error messages in a system. This will have a positive effect on habituation, which will result in more effective errors across the board. Therefore, this is seen as the most important rule.

Designing Out an Error Message

This extra explanation of designing out error messages has been added to the guidelines. This includes a more detailed explanation of designing out error messages with more concrete examples. The purpose of this is to make the guidelines more effective

by giving experts more information to help in identifying scenarios where errors can be designed out.

Designing out an error entails updating the system to remove the error state or scenario. After this it would be impossible to produce the error. A basic example of designing out an error would be an error that is shown when a required field in a form is left empty. By removing the requirement for users to fill in the field, the error cannot be produced any more. This is only valid, though, if the system can function without the specific data.

Another example of designing out an error would be a limit that is placed on a system that is not necessarily required. For instance, a system where a user schedules a task sometime in the future. The system can have a limitation where a task cannot be scheduled further than one year into the future. When a user tries to schedule a task further than a year into the future, an error is produced. To design out this error, the system can be changed to cater for schedules any time in the future.

Guarding Against an Error Message

This extra explanation of guarding against error messages has been added to the guidelines. This includes a more detailed explanation of guarding against an error message, how it differs from designing out, as well as supplying more concrete examples. The purpose behind this addition is to help in the identification of scenarios where an error message can be guarded against, thereby increasing the effectiveness of the guidelines.

Guarding against an error is also very important but it is less preferable than to completely designing out an error. When guarding against an error the goal is also to prevent the user from producing the error. The difference is that instead of updating the system to remove the constraint that produces the error, the user is stopped from entering invalid data or performing an action that produces the error.

Taking the example above of scheduling tasks, consider when a user tries to schedule a task in the past. If the system cannot be changed to allow this to happen, then it is impossible to design out the error. In that case, the error can still be guarded against by forcing the user to select the scheduled date from a calendar which does not have any entries for the past.

Another example would be when a form requires a user to fill in a number value such as when entering the amount when making payments. In this case, adding alphabet characters to the amount will produce an error and it is conceivable that the system cannot be updated to cater for alphabet characters in a numeric field. In this case, the error can be guarded

against by only allowing numbers in the field. On a computer, this means not responding to any invalid character keys pressed on the keyboard, and on a mobile device this will mean presenting the user with a software keyboard that only contains valid characters.

Guiding Questions

These guiding questions have been added to the guidelines. The purpose of this is to facilitate a brainstorming process or a creative problem solving process when trying to identify whether a given error message can be designed out or guarded against. This is also added to the questionnaire described in section 6.3 below.

To facilitate the analysis of an error message the following guiding questions can be used to determine whether it is possible to design out or guard against an error message:

1. What is the basic or underlying cause of the error?
2. Is this due to a limitation imposed by the system?
3. Can the system conceivably be updated or changed to remove the limitation?
4. In a perfect world, how should the system function to prevent the error from occurring?
5. Is the error due to an invalid user interaction?
6. Can the user be prevented from performing the invalid interaction?

6.2.2 Supreme Rule 2: Noticeability

The validity and importance relative to the other supreme rules was confirmed during the evaluation of the guidelines. However, some changes are made to the sub guidelines which follows from the suggestions made in section 5.3.2.

The changes include expansions of guidelines 2, 4 and 6. The purpose of this is to give a better explanation of the guidelines which is expected to increase the effectiveness of the guidelines (make it easier to identify possible usage problems). By giving more verbose explanations with more concrete examples, even people with less experience will be able to use the guidelines effectively.

The guidelines also include the addition of guidelines 3 and 5.

The second rule has to do with how noticeable an error message is. This rule is more important than the third rule, because without actually noticing the error message, the user cannot be expected to understand the content and, therefore, it cannot affect the user's behaviour. This is in line with the C-HIP model discussed in section 2.2.1.

In order to make sure that an error message is noticeable, the following guidelines should be met:

1. Make sure error messages are visually conspicuous. An error message can be made more noticeable by:
   (a) increasing the relative size,
   (b) using more contrasting colours, and
   (c) increasing the font size and weight (Laughery and Wogalter 2006).
2. Error messages should be placed near the cause of an error in terms of location and time (Laughery and Wogalter 2006). The time, in this instance, refers to the moment the action that causes the error is performed, and the location refers to the on-screen location of the item or action that caused the error.
3. Use consistent visual styling of the different classes of errors throughout the system. The classes of errors typically refer to the level of the error for example: page level errors, form errors, or form field errors.
4. 4. Where applicable, use immediate in-line validation (Al-Saleh et al. 2012). Immediate in-line validation is a design pattern used to display error messages for form fields. Firstly, the error message should be shown immediately. That means that as soon as the form field has been entered, or even as the user is filling it in, the error should be displayed. (The opposite would be showing the error message after the form has been submitted). Secondly, the error should be shown in-line with the form field, which means it should be placed right next to (usually on the right hand side) the erroneous form field. (The opposite of this would be to show the error message on the top of the page or above the form field).

   Immediate in-line validation is applicable mostly in web forms where a user fills in multiple entry fields of a form. Each field in the form can have validation or errors associated with them, therefore, each field should show error messages, separately, immediately and in-line with the field.

   An example of where immediate in-line validation should be used is in a registration form with an email address entry. An email address has specific rules that govern its validity. As a user fills in the field, and while the email address is invalid, an error

message should be displayed on the right hand side of the email field indicating that it is invalid.

5. In forms, a combination of a summarizing error at the top of the form and individual errors for each erroneous form field should be used. The summarizing error does not have to be displayed immediately, and may only be added after a form submission.

6. Use modal errors appropriately. A modal, sometimes referred to as a pop-over error, is an error message that is displayed over the current screen or page and prevents the user from interacting with the system until the error message is dismissed. This forces the user to acknowledge the error message and makes the message highly visible. The disadvantage of this type of display is that it interrupts the user and breaks the flow of the current task. Therefore, a modal error should only be used where it is necessary or important enough to stop the user from proceeding with the current task.

6.2.3 Supreme Rule 3: Understandability

Supreme rule 3 and the sub guidelines remained the same except for the removal of the requirement to describe the effect of an error message in rule 3.

The third supreme rule has to do with the understandability of the error content. Without the user being given the required information in an easy-to-understand format, the error message will not be successful in its communication or in its ability to affect the user's behaviour.

The following guidelines, in order of importance, can be used to create understandable error messages:

1. Use a kind and helpful tone, never blaming the user but rather assigning blame to system deficiency (Kelkar et al. 2013, Molich and Nielsen 1990).

2. Use uncomplicated language. Examples of what to steer away from include long sentences, complicated sentence structures, complicated or compound words, and the unnecessary use of conjunctions (Harbach et al. 2013).

3. The error message content should include details about:

   (a) What the cause of the error was, and

   (b) How to recover from the error (Kelkar et al. 2013).

4. Steer away from technical terms that the target audience might not fully understand (Bravo-Lillo et al. 2011, Harbach et al. 2013). If it is not possible to remove technical terms from the error, make sure to keep possible misconceptions about the technical terms in mind when creating the content (Bravo-Lillo et al. 2011, Krol et al. 2012). An example of this is misconceptions users have about phishing and computer viruses which lead to the misunderstanding of security messages (Krol et al. 2012).
5. Ensure that the error message is unambiguous (Duarte et al. 2012).
6. Make the error message as short as possible without compromising on the above guidelines.
7. Add icons where applicable, focusing on semantic distance (how close the icon is to the function it represents) or using well known icons (Kelkar et al. 2013).

6.3 Guidelines As an Evaluation Tool

To assist in the heuristic evaluation process, a questionnaire was created. The questionnaire contains all the guidelines in the form of questions with either a Likert scale or a yes-no selector to indicate whether or to what extent the guideline is met. This can facilitate the analysis process as well as the aggregation of results from multiple experts.

With the updates made to the guidelines as shown in section 6.2, an updated questionnaire can also be created. This involves rewording the guidelines in the form of questions as well as determining whether it is more suitable to use a Likert scale or a yes-no selector for the answer.

Apart from updating the questionnaire based on the updates to the guidelines, the questionnaire is also restructured to move the answer to the supreme rules after all the leading questions or sub guidelines.

The questionnaire is not seen as a substitute for the guidelines shown in section 6.2 above, it is rather meant to be used in conjunction with them because the questionnaire does not contain all the information available in the guidelines. The updated questionnaire is shown in Appendix E.

6.4 Conclusion

Following the results of the demonstration and evaluation of the guidelines first proposed in Chapter 4, an updated set of guidelines for error message design in digital systems could be

created. This is in line with the iterative nature of the design science research methodology. The updating of the guidelines (or artefact) can be seen as the design and development step of the next iteration of the design science research process. The research does not include a full second iteration.

The sections above describe the contribution of the research which includes the updated guidelines for error message design in digital systems as well as a questionnaire that can assist in the evaluation process.

7 Conclusion

This dissertation showed all the details surrounding research performed to fulfil the requirements for a M IT (Information Systems) degree. The aim of the research was to produce a set of guidelines for error message design in digital systems.

From a search of relevant literature on error messages it was found that although some research exists, it is usually relevant to a specific application or specific concern. In contrast to this, the related field of warning design has seen much more research. The starting point of the research was that although different, warnings and error messages have many similarities, which means that some of the theory that informs the design of warnings might also be used to inform the design of error messages.

From the results of this study it was seen that many of the concerns found in warning design theory is indeed applicable to error messages but, as expected, some concerns do not apply to error messages. From the results of the research some additional guidelines were also suggested that apply to error messages, but not warnings.

The rest of this chapter will summarize the key aspects of the research and also give suggestions for future research.

7.1 Research Problem Identification and Motivation

The problem that was identified at the start of the research was that although some research on error message design does exist, and it does present guidance, it does not have the same depth as the research on warning design. In order to fill this gap, the existing research on error message design was combined with that of warning design and summarized into a set of guidelines for error message design in digital systems. Additionally, although existing heuristics on the design of user interfaces include guidelines for error messages, it is seen that the heuristics for the design of error messages are generally difficult to apply (Nielsen and Molich 1990).

It is expected that the use of the guidelines will make it easier for designers and developers of digital systems to design effective error messages. The guidelines could also be used in targeted heuristic evaluations to get a measure of the effectiveness and correctness of existing error messages. Both these expectations will reduce the amount of user testing that has to be done to validate the design of the error messages. This is advantageous, firstly, because user testing can be expensive and time consuming (Nielsen 1992) and, secondly, because the validation of error messages and the scenarios that produce them is problematic.

These benefits will, however, only be gained if the guidelines are effective and valid. The guidelines will be considered effective if it is easy to understand and used to find usability problems with an error message. The guidelines will be valid if the content is complete and reflect what users really care about and contribute positively to the effectiveness of an error message.

7.2 Research Methodology

The research paradigm adopted for the research was design science. Design science is primarily concerned with the description of an IT artefact that addresses a specific business need (Gregor and Hevner 2013). An IT artefact in design science can be many things including constructs, models, methods, and instantiations (Hevner et al. 2004). Examples of this include algorithms, human computer interfaces, and system design methodologies (Vaishnavi and Kuechler 2004). The research was primarily concerned with the creation of guidelines for error message design in digital systems.

The methodology followed to developed the artefact, in other words the guidelines for error message design, was the design science research methodology proposed by Peffers et al. (2007). The design science research methodology describes six steps in an iterative process that needs to be completed. Below is a list of the six steps as completed in this research, with a brief description of its content in terms of the research:

1. *Problem Identification and Motivation* - As discussed above, the problem identified was the lack of general guidelines for error message design in digital systems. The motivation of the research is that the use of such guidelines will enable designers and developers to design effective error messages as well as enable the evaluation of the error messages in a system. This is expected to decrease effort and cost of system development.

2. *Definition of the Objectives for a Solution* - The solution proposed in the research aimed to provide a set of rules or guidelines that can inform error messages design, is shown to be valid and effective, and includes a hierarchy of importance.

3. *Design and Develop* - The development of the guidelines started with a literature review of the body of knowledge regarding warning design and error message design. The result is presented in Chapter 2. The literature then informed the initially proposed guidelines. This is presented in Chapter 4.

4. *Demonstration* - After the guidelines were developed, it had to be demonstrated. This was done by asking three experts to complete a heuristic evaluation of 10 error mes-

sages found in an existing Internet banking website. The purpose of this is to show whether the guidelines are easy to understand and can be used to find potential usability issues with error messages.

5. *Evaluation* - After the demonstration of the guidelines with experts, 14 individual interviews were conducted. In the interviews the same 10 error messages that were used in the demonstration step were presented to the interviewees and discussed in a semi structured interview. The results of the interviews gave an indication of the validity of the guidelines (whether they reflect what users really feel). Additionally, the results from the individual interviews were compared to the results of the heuristic evaluation, which gave a further indication of the validity and effectiveness of the guidelines.

   The results of both the demonstration and evaluation steps, as well as the comparison of the results is shown in Chapter 5.

6. *Communication* - The final step in the design science research methodology is to communicate the details and findings of the research. This is presented in this dissertation.

The design science research methodology is an iterative process. In the research, a second design and development step was completed where the results of the demonstration and evaluation were used to update the initially proposed guidelines. The updated guidelines are given in Chapter 6. The updated guidelines are expected to be more effective and valid than the originally proposed guidelines.

The research paradigms and the methodology of the research was discussed in more detail in Chapter 3.

7.3 Research Questions

At the onset of the research a set of research questions was proposed. Answering these questions was expected to solve the research problem. The overall question that the research answered was: what are the general guidelines to consider when designing error messages in digital systems? In order to answer this, secondary questions were also asked. The questions, and the answers revealed in the research, are as follows:

- What are the primary concerns found in the literature regarding the design of warning and error messages? - The research included a literature review process that searched the existing literature and identified the design considerations and theories regarding warning and error message design. This is discussed in Chapter 2.

- Are the guidelines effective, in other words, how easy is it to understand the guidelines and use them to analyse error messages? - A heuristic evaluation of existing error messages was completed that showed how easy it is to use and understand the proposed guidelines. An analysis of the originally proposed guidelines from the results of the heuristic evaluation is shown in section 5.3.
- Are the guidelines effective, in other words, how easy is it to understand the guidelines and use them to analyse error messages? - A heuristic evaluation of existing error messages was completed that showed how easy it is to use and understand the proposed guidelines. An analysis of the originally proposed guidelines from the results of the heuristic evaluation is shown in section 5.3.
- Are the guidelines valid, in other words, do the design guidelines reflect the pain points and concerns users have regarding error messages in digital systems? - Individual interviews with representative users were conducted on existing error messages and the results showed which guidelines are valid and where the guidelines were lacking. An analysis of the originally proposed guidelines from the results of the individual interviews is shown in section 5.3.
- What is the relative importance of the different guidelines? - The results of the individual interviews also gave an indication of what users struggle with and how it affects the effectiveness of an error message. This was used to conclude on the relative importance of proposed guidelines. This is discussed as part of the analysis of the originally proposed guidelines in section 5.3.

With these answers to the secondary research questions, the main research question was answered in the form of an updated set of guidelines that was changed based on the results of the usability tests performed to answer the secondary questions. The updated guidelines are presented in Chapter 6.

7.4 Summary of Contribution

By following the design science research methodology described above in section 7.2, the research was able to produce a set of guidelines for error message design in digital systems. The guidelines are expected to be effective and valid, based on the results of the demonstration and evaluation steps in the design science research methodology.

A part of the evaluation process was to determine the relative importance of the guidelines. The initially proposed hierarchy of the guidelines followed directly from the C-HIP theoretical model. This model includes the steps that are followed in the processing of information,

specifically, it includes four steps that a receiver of a warning follows. A receiver in this case is the person seeing or hearing the warning. The steps that are followed according to the model are: attention, comprehension, attitude and beliefs, and motivation. The order of the steps is important because the preceding step has to succeed before the following step can be attempted.

The implication of this is that for a warning to be effective it must first be noticed by someone (the warning should grab and maintain the receiver's attention). After this, the receiver has to take in the information supplied by the warning and comprehend it. After successful comprehension of the information, it is analysed through internal thought processes that weigh the validity of the information through the attitude and beliefs of the receiver.

After this, the motivation of the receiver to comply will affect the final compliance decision. More details on this theoretic model and its implication is discussed in section 2.2.1.

The impact of the theoretical model on the guidelines was that guidelines regarding the visual conspicuousness of the error message are seen as more important than guidelines regarding understanding. This was corroborated by the results of the evaluation process of the guidelines. The implication is that, in at least a limited sense, the theoretical model used (which was originally proposed for warnings), is shown to also apply to error messages.

The contributions of the research, as described in Chapter 6, includes the final guidelines that were updated based on the demonstration and evaluation of the initially proposed guidelines. In addition to this a questionnaire was also developed that can assist designers and developers to use the guidelines. The questionnaire includes all the guidelines in the form of questions that have to be answered. The questionnaire gives users of the guidelines the ability to follow a structured and repeatable evaluation process. The results of the guidelines also make it easier to compare the results of more than one person analysing the same error message.

7.5 Concluding Remarks

This dissertation presents the process that was followed to create a set of guidelines for error message design in digital systems. The process followed the design science research methodology which adds to the validity of the results. The created artefact, in other words the design guidelines, presents an easy-to-use tool to evaluate error messages. With the presented guidelines, a small number of experts can evaluate error messages and find usability issues instead of having to perform costlier usability tests.

7.6 Future Research

The research only involved one iteration of the design science research process, which is typically more iterative. Therefore, future research can further evaluate the updated guidelines in subsequent iterations.

One question that could be answered in a subsequent iteration is: to what extent can the use of the proposed guidelines enhance a heuristic evaluation? This can be answered by completing two heuristic evaluations, one using the proposed guidelines for error message design, and one using existing heuristics regarding error messages as given in Molich and Nielsen (1990). A comparison of the two heuristic evaluations can then be done by following the same analysis process as presented in research on usability inspection methods, for example Nielsen and Molich (1990).

The evaluation of the results of the two heuristic evaluations would start with a list of all the usability problems that an error message has. This list would be determined from the literature and the researcher's own experience. The results from the heuristic evaluations could then be summarized into a list of which usability problems each study identified. The result of this would be a measure of how much more effective a heuristic evaluation using the proposed guidelines is. The analysis could also include a metric of effort, such as time spent on the heuristic evaluation. This would give a measure of the relative cost of using the guidelines.

The usability tests performed in the research were limited in that they failed to verify some of the guidelines and also tested the error messages in only one system. Future research can, therefore, include an in depth analysis of specific guidelines or an analysis of the guidelines in other scenarios.

Another area of concern is the ecological invalidity of using static images of error messages in the usability studies. Future research can validate the results of this research by performing usability tests on the error messages of systems in an environment that is a closer representation of the real world, or alternatively try to determine the best way to complete usability tests on error messages.

7.7 Reflection

After the completion of the research I was able to reflect on some of the choices I made and what I could have done differently to deliver better and more convincing results.

An area of concern was a possible over reliance on two papers on warning design theory, namely Laughery and Wogalter (2006) as well as Laughery (2006). Firstly, these two papers contain a thorough summary of the body of knowledge regarding the design of physical warnings. Secondly, there is a lack of other high quality sources regarding much of the content of these two papers. These two factors together, contribute to the need to rely on these two sources to inform much of the literature review.

In hindsight, I could have been more cognisant of the details of how a heuristic evaluation is conducted and how heuristics are used in usability inspection when I developed the initial set of proposed guidelines.

References

Adams, A., Bochner, S. and Bilik, L. 1998. The effectiveness of warning signs in hazardous work places: cognitive and social determinants. *Applied Ergonomics*, 29(4):247–254.

Al-Saleh, M., Al-Wabil, A., Al-Attas, E., Al-Abdulkarim, A., Chaurasia, M. and Alfaifi, R. 2012. Inline immediate feedback in Arabic web forms: An eye tracking study of transactional tasks. *Innovations in Information Technology (IIT), 2012 International Conference on*.

Anon. 2012. Census 2011 Census in brief. [Online]. Available http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf [Accessed 2016-03-18]

Bandura, A. 1982. Self-efficacy mechanism in human agency. *American Psychologist*, 37(2):122–147.

Bargas-Avila, J.A., Brenzikofer, O., Roth, S.P., Tuch, A.N., Opwis, K. and Orsini, S. 2010. Simple but crucial user interfaces in the world wide web: introducing 20 guidelines for usable web form design. In: R. Matrai, (ed.), *User Interfaces*, InTech.

Bargas-Avila, J.A., Oberholzer, G., Schmutz, P., de Vito, M. and Opwis, K. 2007. Usable error message presentation in the World Wide Web: Do not show errors right away. *Interacting with Computers*, 19(3):330–341.

Barlow, T. and Wogalter, M.S. 1991. Increasing the surface area on small product containers to facilitate communication of label information and warnings. *Proceedings of Interface*, 91(7):88–93.

Bartsch, S., Volkamer, M., Theuerling, H. and Karayumak, F. 2013. Contextualized Web Warnings, and How They Cause Distrust. In: M. Huth, N. Asokan, S. Čapkun, I. Flechais and L. Coles-Kemp, (eds.), *Trust and Trustworthy Computing*, Berlin: Springer.

Bias, R.G. 1994. The Pluralistic Usability Walkthrough: Coordinated Empathies. In: J. Nielsen and R.L. Mack, (eds.), *Usability Inspection Methods*, New York, NY, USA: John Wiley & Sons, Inc.

Bravo-Lillo, C., Cranor, L.F., Downs, J.S. and Komanduri, S. 2011. Bridging the Gap in Computer Security Warnings: A Mental Model Approach. *Security & Privacy, IEEE*, 9(2):18–26.

Bravo-Lillo, C., Komanduri, S., Cranor, L.F., Reeder, R.W., Sleeper, M., Downs, J. and Schechter, S. 2013. Your attention please: designing security-decision UIs to make genuine risks harder to ignore. *Proceedings of the Ninth Symposium on Usable Privacy and Security*, Newcastle, United Kingdom: ACM.

Davies, S., Haines, H., Norris, B. and Wilson, J.R. 1998. Safety pictograms: are they getting the message across? *Applied Ergonomics*, 29(1):15–23.

DeTurck, M.A., Chih, I.H. and Hsu, Y.P. 1999. Three Studies Testing the Effects of Role Models on Product Users' Safety Behavior. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 41(3):397–412.

Duarte, E., Rebelo, F., Teles, J. and Noriega, P. 2012. What should I do? A study about conflicting and ambiguous warning messages. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41(0):3633–3640.

Edworthy, J. and Dale, S. 2000. Extending Knowledge of the Effects of Social Influence in Warning Compliance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(25):107–110.

Egelman, S., Cranor, L.F. and Hong, J. 2008. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy: ACM.

Georgsson, M., Weir, C.R. and Staggers, N. 2014. Revisiting heuristic evaluation methods to improve the reliability of findings. *Studies in health technology and informatics*, 205:930–934.

Gregor, S. 2006. The Nature of Theory in Information Systems. *MIS Quarterly*, 30(3):611–642.

Gregor, S. and Hevner, A.R. 2013. Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2):337–356.

Harbach, M., Fahl, S., Yakovleva, P. and Smith, M. 2013. Sorry, I Dont Get It: An Analysis of Warning Message Texts. In: A. Adams, M. Brenner and M. Smith, (eds.), *Financial Cryptography and Data Security*, Berlin Heidelberg: Springer.

Harrell, W.A. 2003. Effect of two warning signs on adult supervision and risky activities by children in grocery shopping carts. *Psychological Reports*, 92(3):889–898.

Hevner, A.R., March, S.T., Park, J. and Ram, S. 2004. Design science in information systems research. *MIS Quarterly*, 28(1):75–105.

Hirschheim, R. and Klein, H.K. 1989. Four paradigms of information systems development. *Communications of the ACM*, 32(10):1199–1216.

Hollingsed, T. and Novick, D.G. 2007. Usability Inspection Methods After 15 Years of Research and Practice. *Proceedings of the 25th Annual ACM International Conference on Design of Communication*, SIGDOC '07, New York, NY, USA: ACM.

Hunn, B.P. and Dingus, T.A. 1992. Interactivity, information, and compliance cost in a consumer product warning scenario. *Accident Analysis & Prevention*, 24(5):497–505.

Isherwood, S.J., McDougall, S.J.P. and Curry, M.B. 2007. Icon Identification in Context: The Changing Role of Icon Characteristics With User Experience. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(3):465–476.

Jaynes, L.S. and Boles, D.B. 1990. The Effect of Symbols on Warning Compliance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 34(14):984–987.

Jeffries, R. and Desurvire, H. 1992. Usability Testing vs. Heuristic Evaluation: Was There a Contest? *SIGCHI Bull.*, 24(4):39–41.

Jeffries, R., Miller, J.R., Wharton, C. and Uyeda, K. 1991. User Interface Evaluation in the Real World: A Comparison of Four Techniques. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, New York, NY, USA: ACM.

Kahn, M.J. and Prail, A. 1994. Formal Usability Inspections. In: J. Nielsen and R.L. Mack, (eds.), *Usability Inspection Methods*, New York, NY, USA: John Wiley & Sons, Inc.

Kauer, M., Pfeiffer, T., Volkamer, M., Theuerling, H. and Bruder, R. 2012. It is not about the design - it is about the content! Making warnings more efficient by communicating risks appropriately. *Sicherheit 2012 - Sicherheit, Schutz und Zuverlässigkeit - 6. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft fuer Informatik e.V. (GI)*, volume P-195, Darmstadt: Gesellschaft fuer Informatik e.V. (GI).

Kelkar, T., Gadepalli, P. and Indurkhya, B. 2013. An Assistive Tool to Improve Usability of Error Dialogs. *Technologies and Applications of Artificial Intelligence (TAAI), 2013 Conference on*.

Kim, S. and Wogalter, M.S. 2009. Habituation, Dishabituation, and Recovery Effects in Visual Warnings. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 53(20):1612–1616.

Kline, P.B., Braun, C.C., Peterson, N. and Silver, N.C. 1993. The Impact of Color on Warnings Research. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(14):940–944.

Krol, K., Moroz, M. and Sasse, M.A. 2012. Don't work. Can't work? Why it's time to rethink security warnings. *Risk and Security of Internet and Systems (CRiSIS), 2012 7th International Conference on*.

Laughery, K.R. 2006. Safety communications: Warnings. *Applied Ergonomics*, 37(4):467–478.

Laughery, K.R. and Wogalter, M.S. 2006. Designing Effective Warnings. *Reviews of Human Factors and Ergonomics*, 2(1):241–271.

Laughery, K.R., Young, S.L., Vaubel, K.P. and Brelsford, Jr, J.W. 1993. The Noticeability of Warnings on Alcoholic Beverage Containers. *Journal of Public Policy & Marketing*, 12(1):38–56.

Lazar, J., Feng, J.H. and Hochheiser, H. 2011. *Research methods in human-computer interaction*. John Wiley & Sons.

Maurer, M.E., Luca, A.D. and Hussmann, H. 2011. Data type based security alert dialogs. *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, Vancouver, BC, Canada: ACM.

Molich, R. and Nielsen, J. 1990. Improving a Human-computer Dialogue. *Commun. ACM*, 33(3):338–348.

Myers, M.D. 2013. *Qualitative Research in Business and Management*. London: Sage, 2 edition.

Nielsen, J. 1992. Finding Usability Problems Through Heuristic Evaluation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, New York, NY, USA: ACM.

Nielsen, J. 1994. Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. [Online]. Available https://www.nngroup.com/articles/guerrilla-hci/ [Accessed 2016-03-12]

Nielsen, J. 1995a. Characteristics of Usability Problems Found by Heuristic Evaluation. [Online]. Available https://www.nngroup.com/articles/usability-problems-found-by-heuristic-evaluation/ [Accessed 2016-03-11]

Nielsen, J. 1995b. How to Conduct a Heuristic Evaluation. [Online]. Available https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/ [Accessed 2016-01-01]

Nielsen, J. 1995c. Severity Ratings for Usability Problems. [Online]. Available https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/ [Accessed 2016-03-12]

Nielsen, J. and Molich, R. 1990. Heuristic Evaluation of User Interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, New York, NY, USA: ACM.

Orlikowski, W.J. and Baroudi, J.J. 1991. Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research*, 2(1):1–28.

Peffers, K., Tuunanen, T., Rothenberger, M. and Chatterjee, S. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3):45–77.

Polson, P.G., Lewis, C., Rieman, J. and Wharton, C. 1992. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies*, 36(5):741–773.

Rainer, B. and Stefan, K. 2010. Trained to accept?: a field experiment on consent dialogs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, Georgia, USA: ACM.

Rogers, W.A., Lamson, N. and Rousseau, G.K. 2000. Warning Research: An Integrative Perspective. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 42(1):102–139.

Rogers, Y., Sharp, H. and Preece, J. 2011. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, 3rd edition.

Seckler, M., Heinz, S., Bargas-Avila, J.A., Opwis, K. and Tuch, A.N. 2014. Designing usable web forms: empirical evaluation of web form improvement guidelines. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, Ontario, Canada: ACM.

Silver, N.C., Leonard, D.C., Ponsi, K.A. and Wogalter, M.S. 1991. Warnings and purchase intentions for pest-control products. *Forensic Reports*, 4(1):17–33.

Tan, M. and Teo, T.S.H. 2000. Factors influencing the adoption of Internet banking. *Journal of the AIS*, 1(1es):5.

Vaishnavi, V. and Kuechler, B. 2004. Design Science Research in Information Systems. [Online]. Available http://www.desrist.org/design-research-in-information-systems/ [Accessed 2015-05-12]

Wogalter, M.S., Allison, S.T. and Mckenna, N.A. 1989. Effects of Cost and Social Influence on Warning Compliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 31(2):133–140.

Wogalter, M.S., Barlow, T. and Murphy, S.A. 1995. Compliance to owner's manual warnings: influence of familiarity and the placement of a supplemental directive. *Ergonomics*, 38(6):1081–1091.

Wogalter, M.S., Begley, P.B., Scancorelli, L.F. and Brelsford, J.W. 1997. Effectiveness of elevator service signs: Measurement of perceived understandability, willingness to comply and behaviour. *Applied Ergonomics*, 28(3):181–187.

Wogalter, M.S., DeJoy, D. and Laughery, K.R., (eds.). 1999. *Warnings and risk communication*. CRC Press.

Wogalter, M.S., Godfrey, S.S., Fontenelle, G.A., Desaulniers, D.R., Rothstein, P.R. and Laughery, K.R. 1987. Effectiveness of Warnings. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 29(5):599–612.

Wogalter, M.S. and Silver, N.C. 1995. Warning signal words: connoted strength and understandability by children, elders, and non-native English speakers. *Ergonomics*, 38(11):2188–2206.

Young, S.L. and Wogalter, M.S. 1990. Comprehension and Memory of Instruction Manual Warnings: Conspicuous Print and Pictorial Icons. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 32(6):637–649.

Appendices

A Ethical Clearance Letter

Below is a copy of the letter of ethical clearance received from the Faculty Committee for Research Ethics and Integrity of the Faculty of Engineering, Built Environment and Information Technology of the University of Pretoria. The ethical clearance received was conditional on changes that had to be made to the individual interview process. These changes were made before data collection.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Engineering,
Built Environment and Information Technology

1956 – 2016
60
years of
Engineering Education

Reference number: EBIT/68/2015 8 March 2016

Mr H Pieterse
Department Informatics
School of Information Technology
University of Pretoria
Pretoria
0028

Dear Mr Pieterse,

FACULTY COMMITTEE FOR RESEARCH ETHICS AND INTEGRITY

Your recent application to the EBIT Ethics Committee refers.

1. I hereby wish to inform you that the research project titled "Towards Guidelines for Error Message Design in Digital Systems" has been approved conditionally by the Committee. The conditions are that the applicant:
   a) remove the questions about age and gender from the questionnaire/interview (these are not motivated in the context of the study)
   b) add into the informed consent form that audio recordings may be made (explicit consent needed for his)

   This approval does not imply that the researcher, student or lecturer is relieved of any accountability in terms of the Codes of Research Ethics of the University of Pretoria, if action is taken beyond the approved proposal.

2. According to the regulations, any relevant problem arising from the study or research methodology as well as any amendments or changes, must be brought to the attention of any member of the Faculty Committee who will deal with the matter.

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

The Committee wishes you every success with the research project.

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

EBIT Research Ethic Committee
Room 15-6, Level 15, Engineering Building 1
University of Pretoria, Private Bag X20
Hatfield 0028, South Africa
Tel +27 (0)12 420 3736
Email mari.ferreira@up.ac.za

B Heuristic Evaluation Guide

The guide supplied to experts for the heuristic evaluation is shown below. This guide includes details surrounding the initially proposed guidelines for error message design in digital systems and details of all the error messages that should have been evaluated. The experts were also given a evaluation form which is shown in Appendix C as well as a set of higher resolution images.

B.1 Introduction

As part of a research study into the design of effective error messages in Internet banking systems, a heuristic evaluation by field experts needs to be completed. This document includes all the information and material required to complete such a heuristic evaluation.

The aim of the research being performed is to create a set of guidelines for error message design. Such a set of guidelines was created by completing a literature review and compiling the information into a concise set of guidelines.

To evaluate the guidelines a heuristic evaluation of an existing Internet banking website needs to be completed using the proposed guidelines. Therefore, the purpose of the heuristic evaluation is:

- To evaluate how easy the proposed guidelines (heuristics) are to understand, and
- To determine how relevant the guidelines are i.e. to what extent do existing systems already adhere to the guidelines.

The rest of this document contains firstly an explanation of the design guidelines (or heuristics), and secondly the details of the evaluation itself.

B.2 The Heuristics

There are three supreme rules to adhere to when designing error messages. They are listed, in order of importance, as follows:

1. Never show an error message if the cause of the error can be designed out or guarded against.

2. Error messages must be noticeable.
3. The content of the error message must be easy to understand and should communicate the information required.

More information on the specifics and the source of the rules are shown in the sections below.

B.2.1 Supreme Rule 1: Hazard Control Hierarchy

The first rule: "Never show an error message if the cause can be designed out or guarded against", comes directly from the hazard control hierarchy. This rule states that a warning should only be shown when it is not feasible to design out the risk or guard against it (Laughery and Wogalter 2006, Laughery 2006, Bravo-Lillo et al. 2011). The same can be said for errors, where the error should only be shown when the action that produces the error cannot be designed out or guarded against. The result of this strategy will be a reduction in the number of error messages in a system. This will have a positive effect on habituation, which will result in more effective errors across the board. Therefore, this first rule is seen as the most important of the supreme rules.

An example of designing out the cause of an error would be removing a required field from a form that a system does not need. Guarding against an error would be not allowing invalid characters in a numeric field or using date pickers when filling in date fields.

B.2.2 Supreme Rule 2: Noticeability

The second rule has to do with how noticeable an error message is. This rule is more important than the third rule, because without actually noticing the error message, the user cannot be expected to understand the content and, therefore, it cannot affect the user's behaviour. This is in line with the C-HIP model discussed in section 2.2.1.

In order to make an error message noticeable, the guidelines below are suggested:

1. Make sure error messages are visually conspicuous. An error message can be made more noticeable by:
   (a) increasing the relative size,
   (b) using more contrasting colours, and

(c) increasing the font size and weight (Laughery and Wogalter 2006).

2. Error messages should be placed near the cause of an error in terms of location and time (Laughery and Wogalter 2006).

3. Where applicable, use immediate in-line validation (Al-Saleh et al. 2012). Immediate in-line validation is a design pattern where error messages for form fields are shown immediately after the field has been entered and showing the error in line with the form field. The opposite of this would be showing error messages at the top of the page only after a form submission has been attempted.

4. Use pop-overs or modal views only when it is necessary to stop the user from proceeding with the current task or where the error message is critical enough to interrupt the current task flow.

B.2.3 Supreme Rule 3: Understandability

The third supreme rule has to do with the understandability of the error content. Without the user being given the required information in an easy-to-understand format, the error message will not be successful in its communication or in its ability to affect the user's behaviour.

The following guidelines, in order of importance, can be used to create understandable error messages:

1. Use a kind and helpful tone, never blaming the user but rather assigning blame to system deficiency (Kelkar et al. 2013, Molich and Nielsen 1990).

2. Use uncomplicated language. Examples of what to steer away from includes long sentences, complicated sentence structures, complicated or compound words, and the unnecessary use of conjunctions (Harbach et al. 2013).

3. The error message content should include details about:

   (a) what the effects of the error is,

   (b) what the cause of the error is, and

   (c) how to recover from the error (Kelkar et al. 2013).

4. Steer away from technical terms that the target audience might not fully understand (Bravo-Lillo et al. 2011, Harbach et al. 2013). If it is not possible to remove technical terms from the error, make sure to keep possible misconceptions about the technical terms in mind when creating the content (Bravo-Lillo et al. 2011, Krol et al. 2012). An

example of this is misconceptions users have about phishing and computer viruses which lead to the misunderstanding of security messages (Krol et al. 2012).

5. Ensure that the error message is unambiguous (Duarte et al. 2012).
6. Make the error message as short as possible without compromising on the aforementioned guidelines.
7. Add icons where applicable, focusing on semantic distance (how close the icon is to the function it represents) or using well known icons (Kelkar et al. 2013).

The next section includes all the information needed to perform the heuristic evaluation.

B.3 Evaluation Procedure

The heuristic evaluation that should be performed includes the evaluation of ten error messages shown in an Internet banking website using the design guidelines described in the previous section. The evaluation of each error message can be done as follows.

1. Read the scenario description that will explain what was done to lead to the specific error scenario.
2. Read the details of the error that was produced which can also include possible causes.
3. Look at the image provided of the error as it is shown in the Internet banking website.
4. Complete the heuristic evaluation form.

The evaluation form provides a table of the heuristics with provided space to rate how well the error adheres to the heuristic. To make the rating easier, a true-false selector or a Likert scale is given. For the heuristics that are rated using a true-false selector an indication should be given whether the error message adheres to the heuristic or not. For the Likert scale rating an indication should be given to what extent the error message adheres to the heuristic. The heuristic in the evaluation form will be reworded into questions so that the rating scale makes sense.

For each heuristic a comment block is also provided that can be used to add optional comments on the reason for the answer and the heuristic itself. At the end of each section a comment block is also provided where general comments can be added. Please be as verbose as you want in your comments and do not worry about adding information that you think might be unnecessary. The more data that can be gathered the better the evaluation of the guidelines will be.

B.4 Error Evaluations

This section contains the details of each error to evaluate, including the scenario around the error, a description of the error, a screen shot of the error, and the evaluation form to complete. This document contains scaled down versions of the screen shots of the error messages for reference purposes. The full size screen shots are provided in separate files where the file name of the full size images are given at the bottom of each screen shot in this document.

B.4.1 Login failure

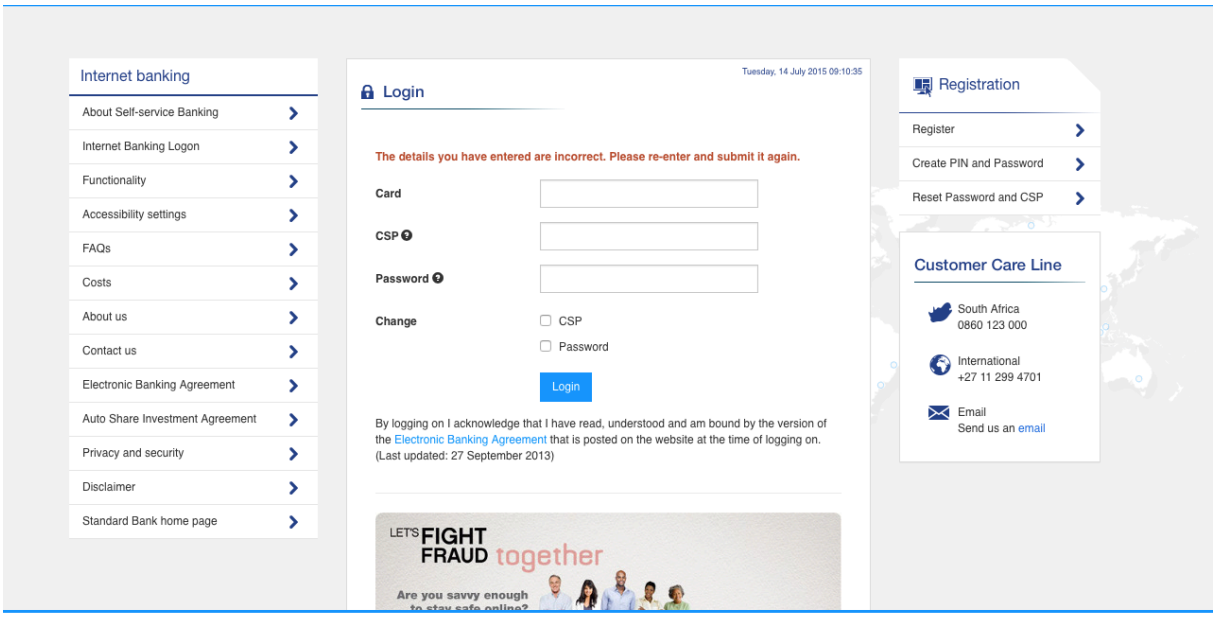
Scenario Description

You want to log into your bank's Internet banking portal. You fill in all your credentials and click on the login button. You are then presented with the screen below.

Error Detail

When incorrect login credentials are supplied the system will return an error. As part of proper security practise, it is not possible to know whether supplied credentials are valid before the login button is pressed or to specify what field was invalid.

Screen



1-FailedLogin.pdf

B.4.2 Automatic Logout

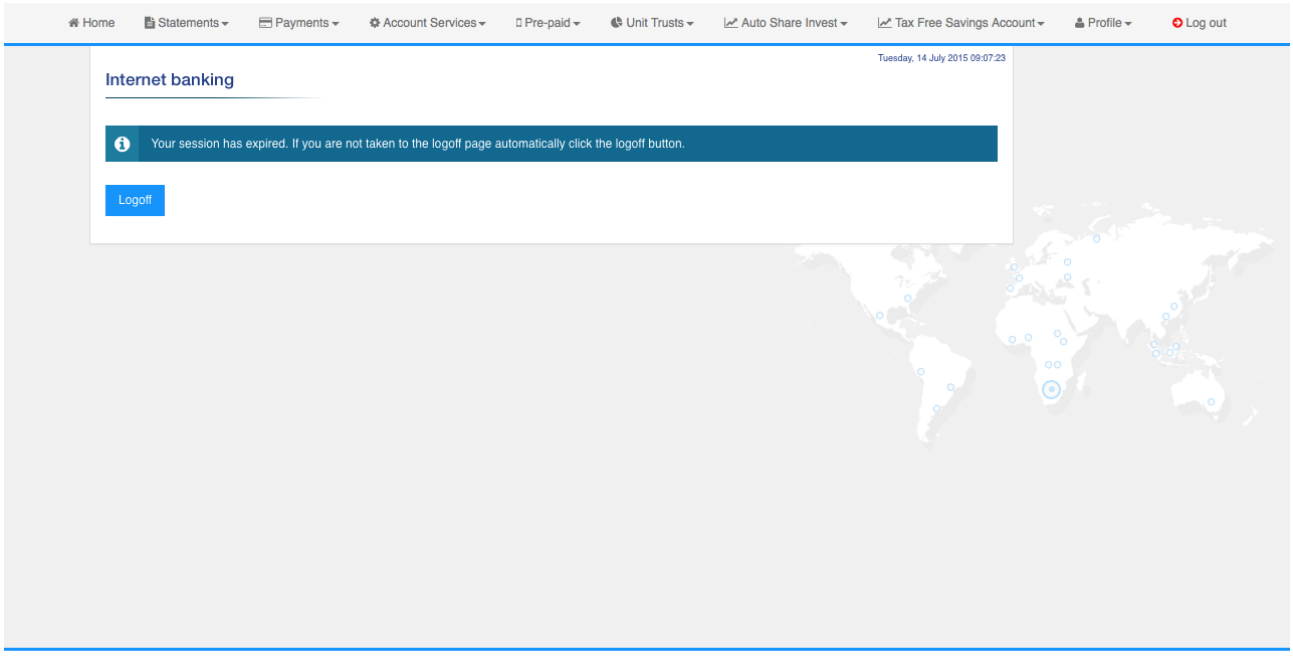
Scenario Description

While you are busy adding your friend as a new beneficiary you realise you do not know the type of account your friend has. You quickly call him, but the conversation takes a bit longer that you expected. When you return you complete the form, but when you submit the form you are shown the screen below.

Error Detail

As a security precaution a session on the Internet banking website will terminate itself after a period of inactivity. As soon as a request is made to the website (page navigation or form submit) and the session has been terminated due to inactivity, the server automatically redirects the user to the page below.

Screen



1-SessionTimeout.pdf

B.4.3 Payments Above Monthly Limit

Scenario Description

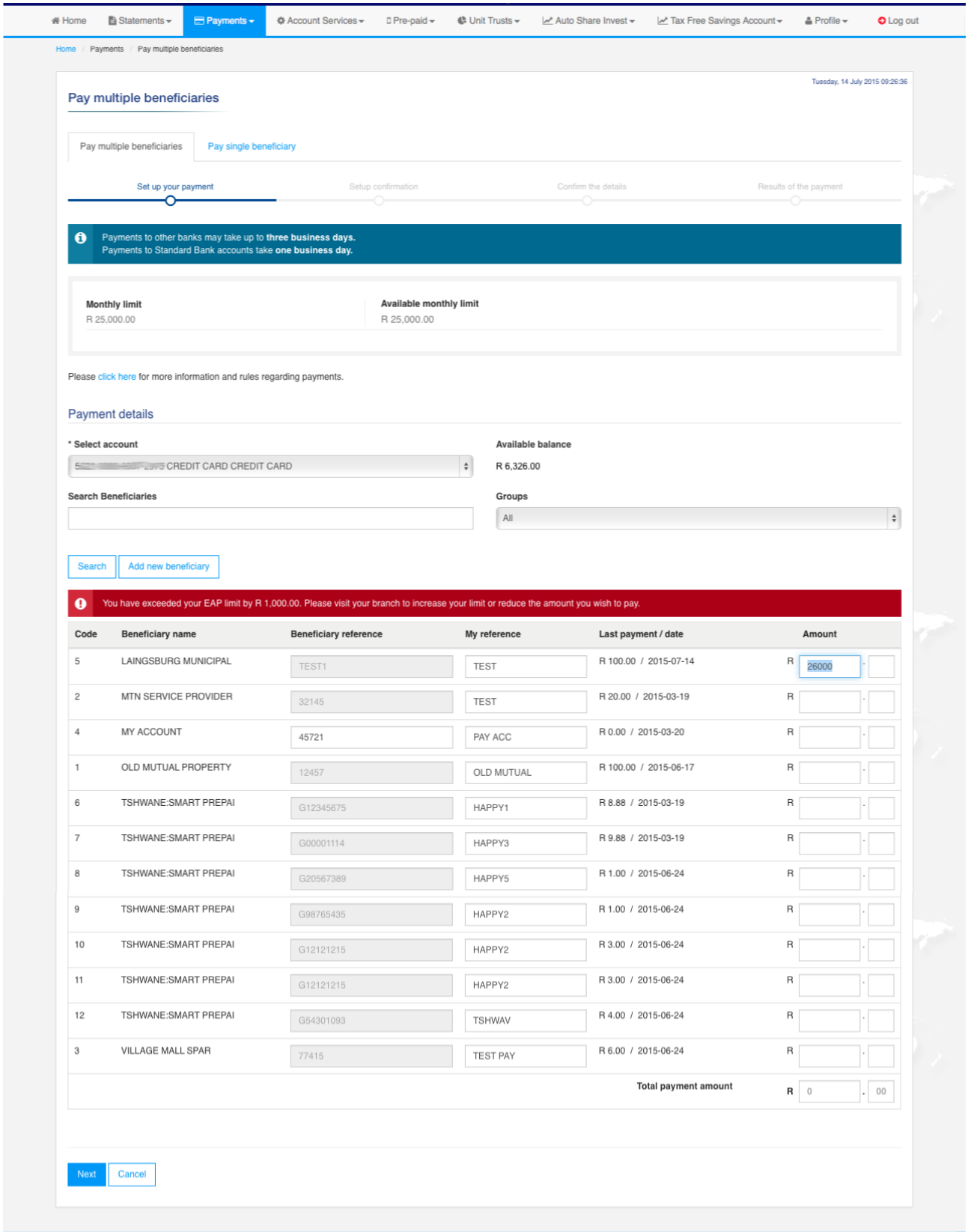
You saved up some money over the past few months, and you finally have enough money to buy your friends car. Your first payment to your friend is quite large, but when you try to pay him you are presented with the following screen.

Error Detail

As part of the security policies of a bank, you specify a monthly payment limit. You are then not allowed to transfer more money than this limit per month. You can make changes to this limit in one of the following manners:

- By visiting your local branch
- By calling 0860 123 000 and following the voice prompts
- By sending a secure message to your branch in the profiles section of the Internet banking website. This request will processed after two working days.

Screen



3-LimitExceeded.pdf

B.4.4 Incorrect One-Time Pin

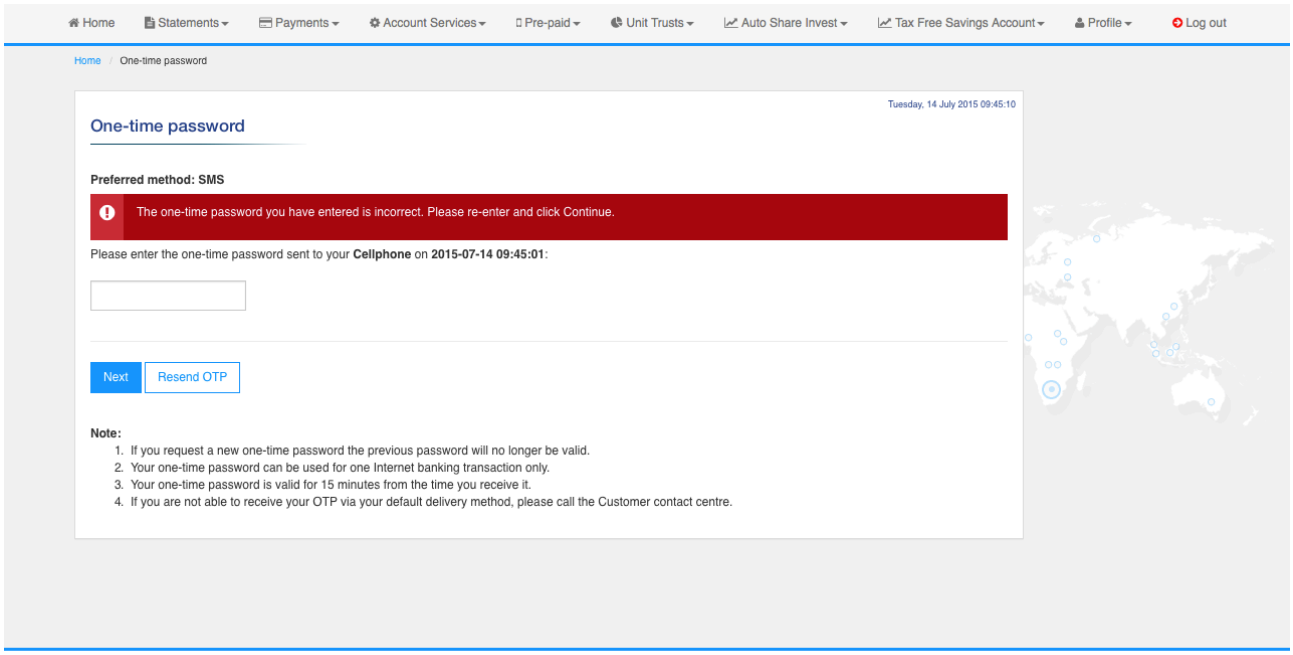
Scenario Description

You are making a payment to a once of beneficiary, and after filing in the details of the payment you are asked to fill in your one-time pin. You receive the one-time pin on your cellphone and you fill it in and try to continue, but you are presented with the following screen.

Error Detail

As part of the process for specific Internet banking transactions users are required to fill in a one-time pin. This pin is sent to their cell phones or email addresses and helps to verify their identity and guard against fraudulent transactions.

Screen



3-OTPFailed.pdf

B.4.5 Insufficient Funds

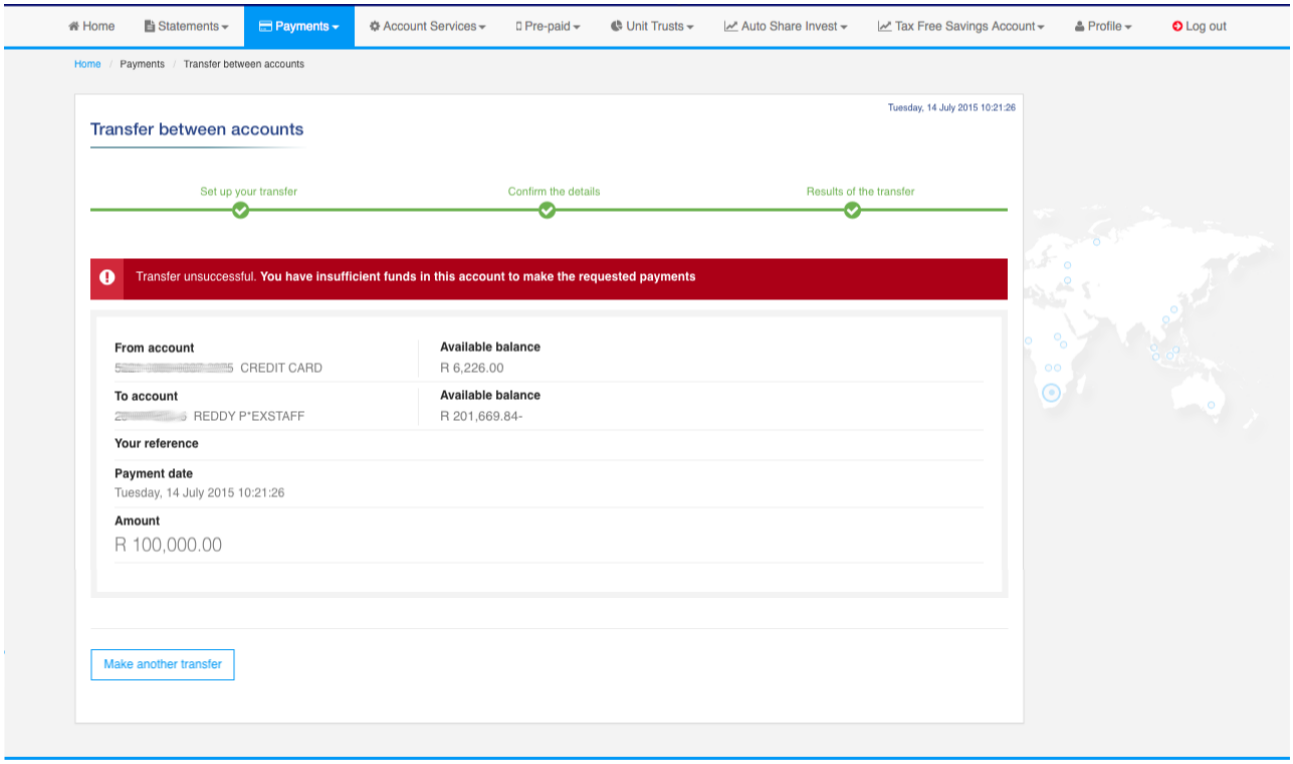
Scenario Description

You want to transfer some money between two of your accounts. You fill in the details and confirm that it is correct. On the last screen you are presented with the following screen.

Error Detail

As part of the system requirements, you are not allowed to transfer more money out of an account than is available in it.

Screen



5-InsufficientFunds.pdf

B.4.6 New Beneficiary

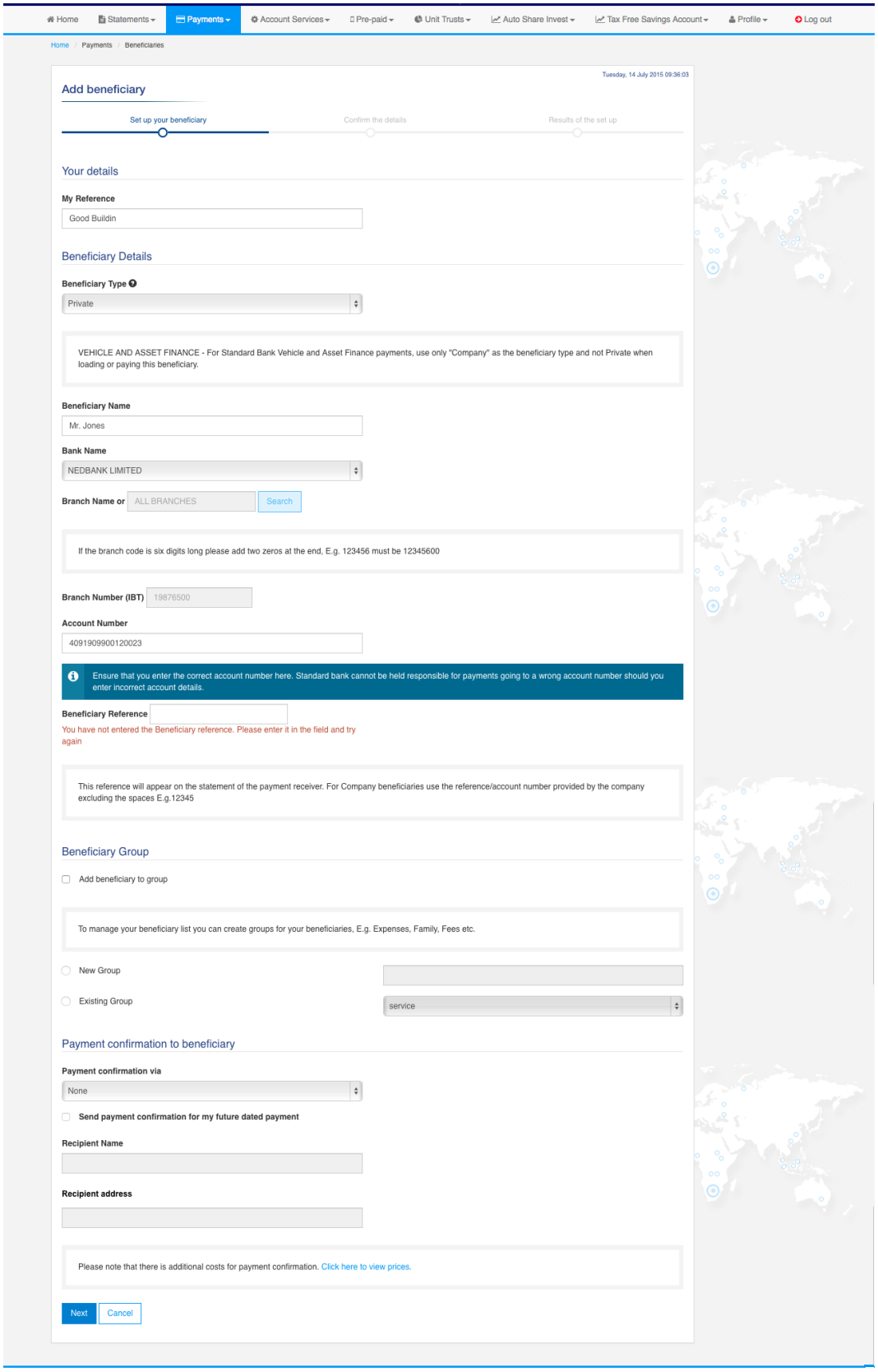
Scenario Description

You want to add a new beneficiary to your Internet banking profile. You fill in the account details in the appropriate fields and then try to continue with the process, but you are presented wit the following error.

Error Detail

Part of the required fields of a new beneficiary is the beneficiary reference, which will be displayed on the statement of the receiver of the payment.

Screen



6-NewBenificiary.pdf

B.4.7 Back-End System Communication Error

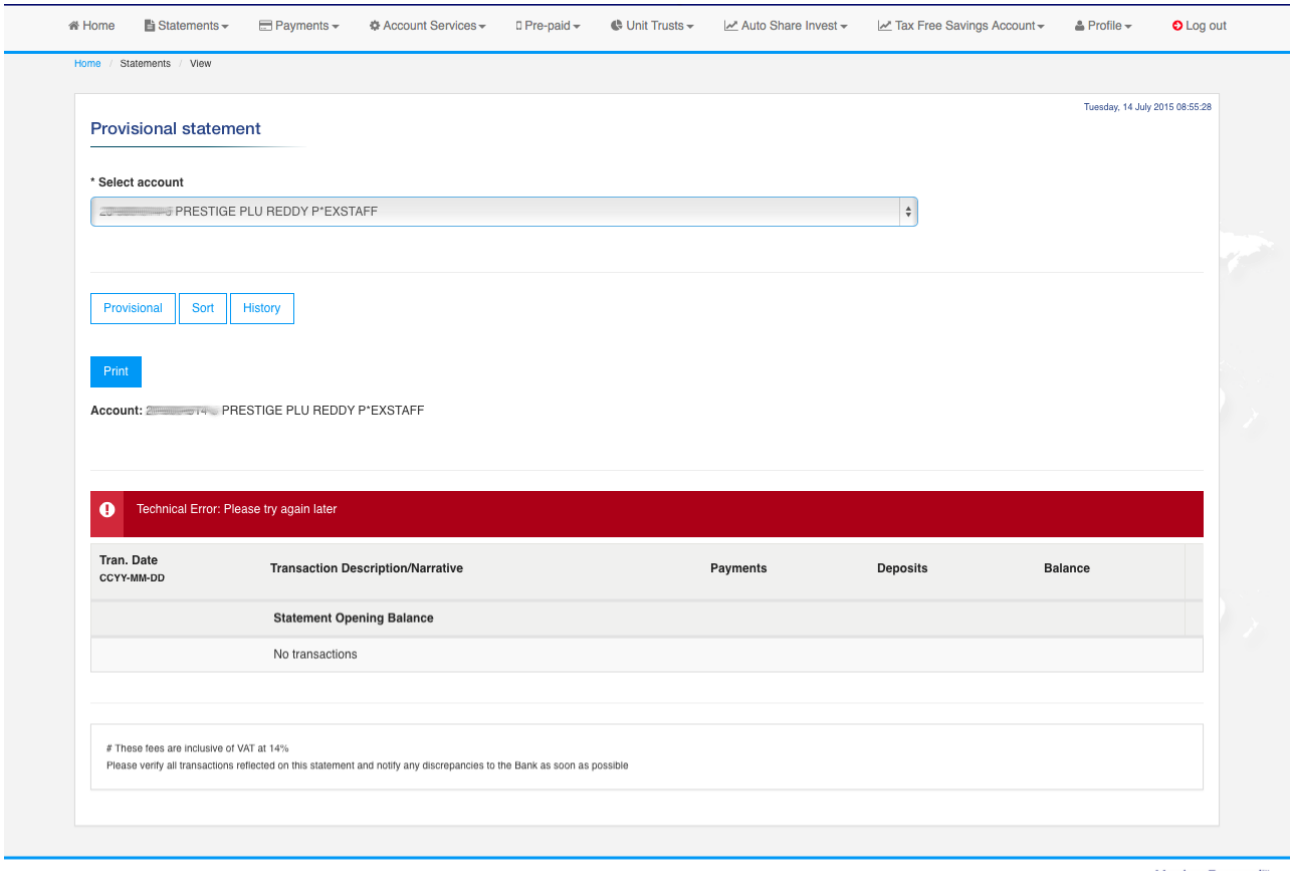
Scenario Description

You are logged in on your bank's Internet banking website, and you want to see a provisional account statement of your transactions for the past few days. You click on the provisional statement link in the menu, select you current account, and then you are presented with the error below.

Error Detail

This error is presented to the user when the Internet banking system failed to communicate successfully with the back-end system that is responsible for generating account statements. The reason for the communication is not known.

Screen



7-LimitExceeded.pdf

B.4.8 Repeating Payment

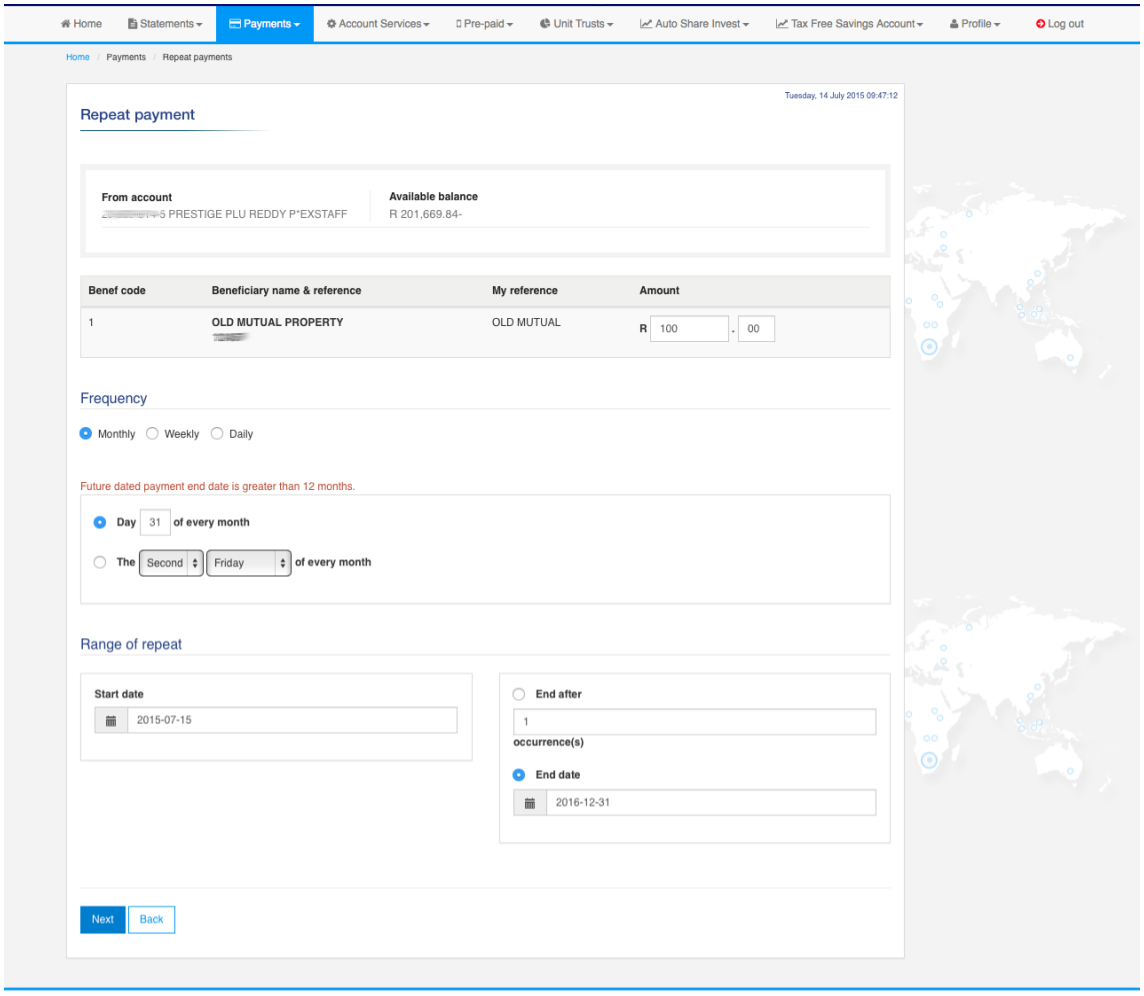
Scenario Description

You just moved in to a new apartment and you want to set up a repeating payment for your rent. You are planning to stay there fore a while so you set up the payment to start now to the end of next year. After filling in the form and selecting the start and end dates with the date picker, you try to continue but you are presented with the following screen.

Error Detail

As part of the system limitations and policies, you are not allowed to add repeating payments for more than a year in the future.

Screen



8-RepeatingPayment.pdf

B.4.9 Transaction History

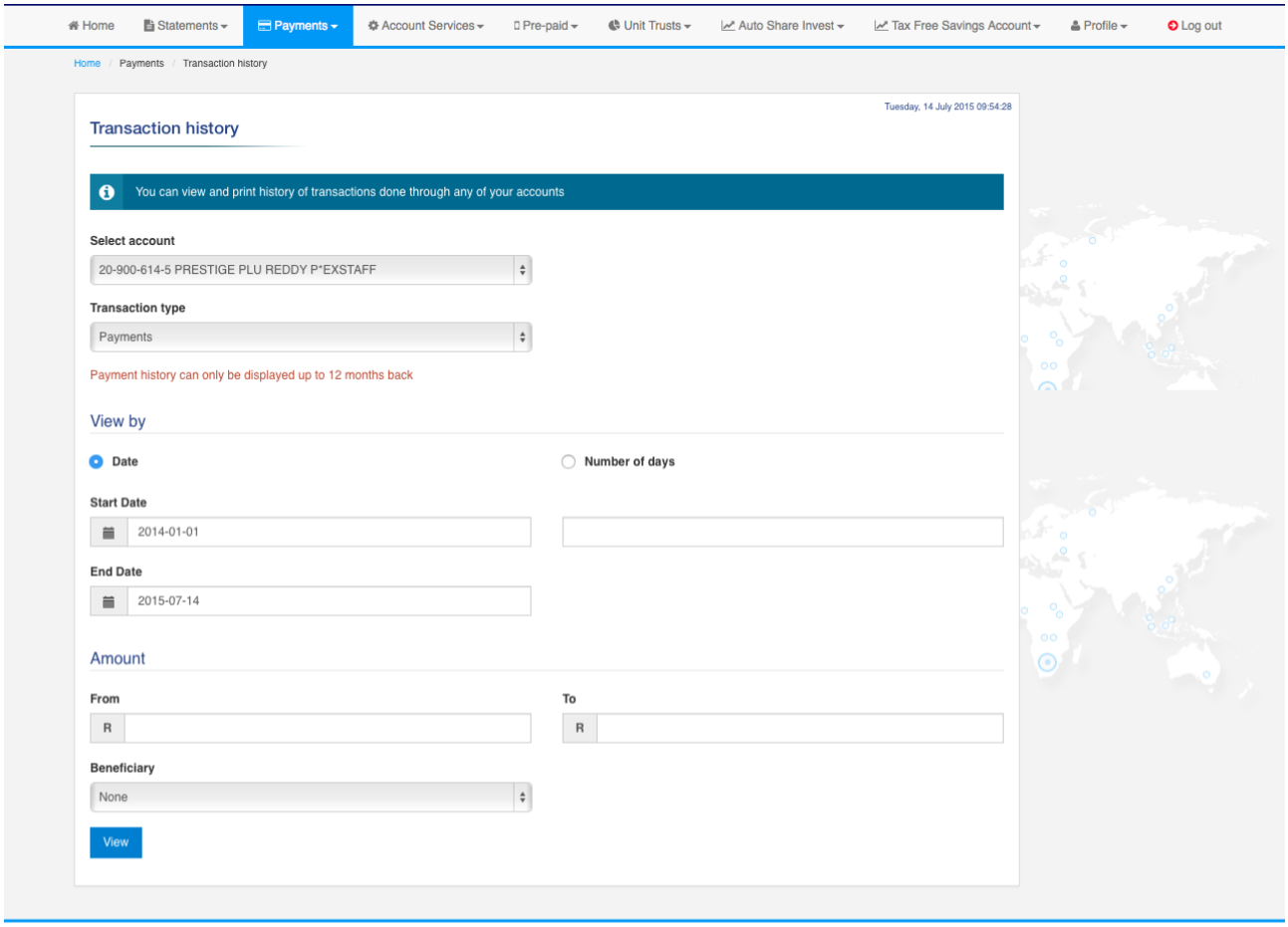
Scenario Description

You are looking for a specific payment you made last year between your credit card and savings account. You can't remember the exact details or when exactly you made it so you try to find it using the transaction history feature. You select the appropriate account and select the start date as the start of last year with the date picker. When you try to view the results you see the following screen.

Error Detail

As part of a system limitation, you cannot view payment history form more than a year ago.

Screen



9-TransactionHistory.pdf

B.4.10 Account Statement Email

Scenario Description

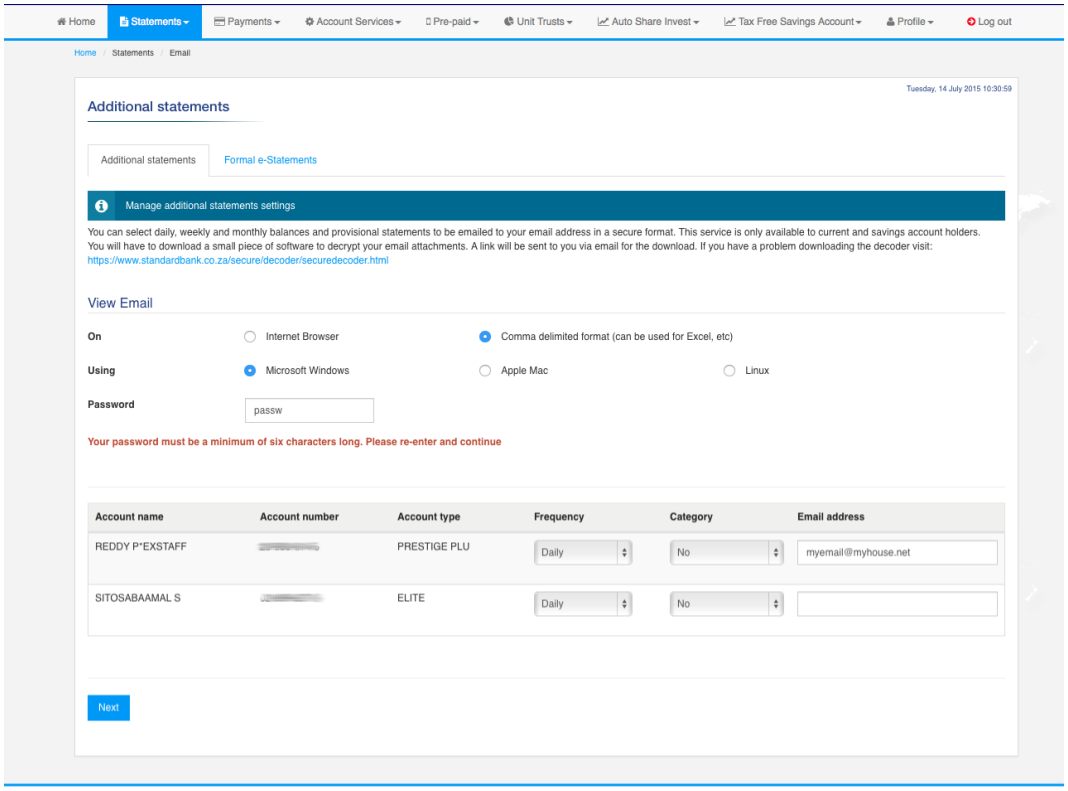
You want to take out a new cellphone contract which required 3 months worth of bank statements. You do not have a printer at home, so you decide you will email yourself the statements and print them at work tomorrow. You log onto your Internet banking profile and try to email yourself an account statement. As part of the process you have to enter a password which will be used to encrypt the email. You fill in the form details and try to continue but you are presented with the first screen shown below.

The error indicated that your password must be longer, so you increase the length and try to continue again, but you are presented with the second screen below.

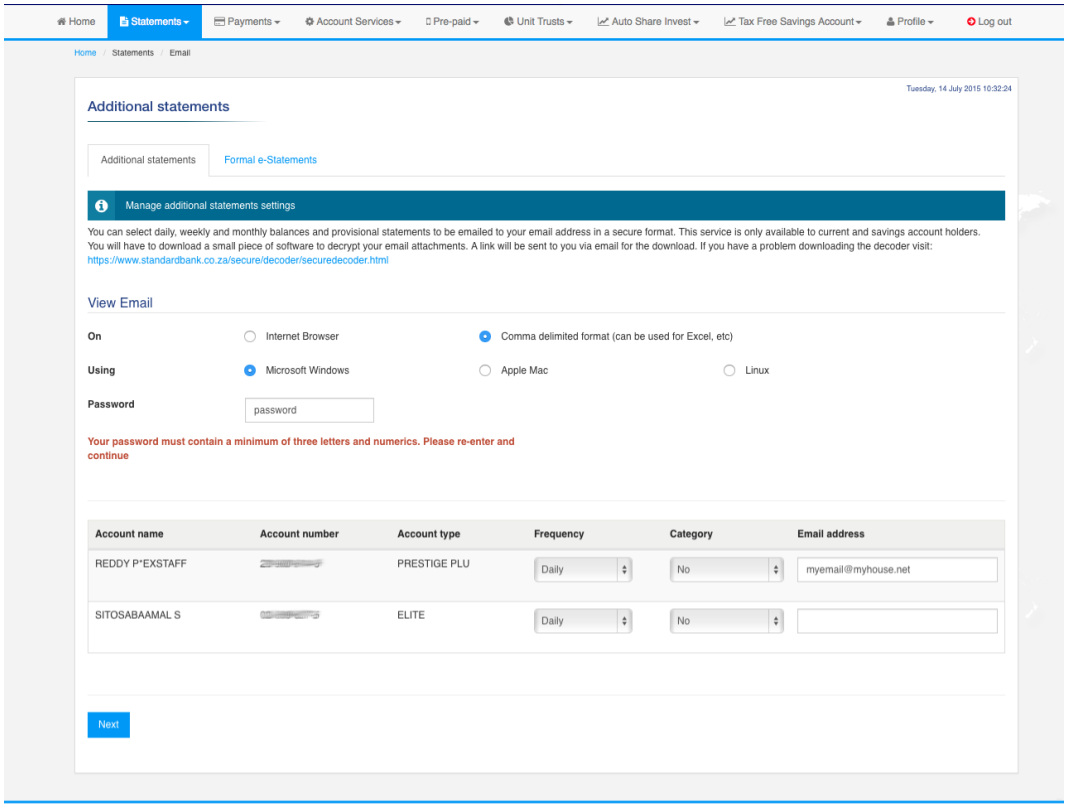
Error Detail

As part of the security requirements of the system all email statements are encrypted and you have to enter the password for the file before it can be sent. The chosen password is required to be longer than 6 characters and include al least three letters and one number.

Screen



10-EmailStatementA.pdf



10-EmailStatementB.pdf

C Heuristic Evaluation Form

As part of the heuristic evaluation performed in the research, the experts were supplied with an evaluation form. This form was intended to assist the analysis of error messages and contained a question for each of the design guidelines and either a Likert rating scale of a yes/no selection to assist in the. This was supplied together with the heuristic evaluation guide shown in Appendix B

Supreme Rules (Error Evaluation)

No.	Description	Score	Comments
1	Can the cause of the error be designed out or guarded against?	○ yes ○ no	
2	Do you agree that the error message is noticeable?	○ strongly disagree ○ disagree ○ neutral ○ agree ○ strongly agree	
3	Do you agree that the error message is understandable and communicates the necessary information?	○ strongly disagree ○ disagree ○ neutral ○ agree ○ strongly agree	
	Comments:		

Supreme Rule 2 - Guidelines (Error Evaluation)

No.	Description	Score					Comments
		strongly disagree	disagree	neutral	agree	strongly agree	
2.1	Do you agree that the error message is visually conspicuous successfully using size and contrast?	○	○	○	○	○	
2.2	Do you agree that the error is close enough to the cause of the error in terms of both time and location?	○	○	○	○	○	
2.3	If applicable, do you agree that immediate in-line validation is used appropriately?	○	○	○	○	○	
2.4	If applicable, do you agree that the use of pop-up errors is appropriate?	○	○	○	○	○	

Comments:

Supreme Rule 3 - Guidelines (Error Evaluation)

No.	Description	Score	Comments
3.1	Do you agree that the tone of the error message is kind, helpful, and does not blame the user?	○ yes ○ no	
3.2	Do you agree that the language is uncomplicated?	strongly disagree ○ disagree ○ neutral ○ agree ○ strongly agree ○	
3.3.a	Does the error message include details about the effects of the error?	○ yes ○ no	
3.3.b	Does the error message include details about the cause of the error?	○ yes ○ no	
3.3.c	Does the error message include details about how to recover from the error?	○ yes ○ no	
3.4	Do you agree that the error message does not include technical terms that a typical user might not understand or is not explained properly?	strongly disagree ○ disagree ○ neutral ○ agree ○ strongly agree ○	

Supreme Rule 3 - Guidelines (Error Evaluation)

No.	Description	Score					Comments
3.5	Do you agree that the error message is unambiguous?	strongly disagree ○	disagree ○	neutral ○	agree ○	strongly agree ○	
3.6	Do you agree that the error message is as short as possible without compromising on the other guidelines?	strongly disagree ○	disagree ○	neutral ○	agree ○	strongly agree ○	
3.7	Do you agree that the icon use is appropriate i.e. 1. If icons are used do you agree that the icons are appropriate and necessary? 2. If no icon is present, do you agree that an icon is not necessary?	strongly disagree ○	disagree ○	neutral ○	agree ○	strongly agree ○	

Comments:

D Individual Interview Guide

The individual interviews were conducted on a semi structured basis. The interview guide shown in this appendix was used to guide the interviews.

D.1 Introduction

This document should be used to guide the individual interviews as part of a study on error message design in Internet banking systems. Each interview must complete each of the following steps:

1. Complete the informed consent form with the study participant.
2. Ask the participant each of the question in the background and demographics section and fill in the answers.
3. Read the interview introduction shown in section D.3 out load to the participant.
4. Follow the interview steps out lined in section D.4.

D.2 Background and Demographics

No.	Question	Answer
1	Is English your home or first language?	◯ Yes ◯ No
2	What is your profession?	
3	How often to you use an Internet banking website?	◯ never ◯ less than once a month ◯ once a month ◯ once a week ◯ daily

No.	Question	Answer
4	How often do you use a computer?	◯ less than once a week ◯ multiple times a week ◯ on a daily basis
5	How often to you browse the Internet	◯ less than once a week ◯ multiple times a week ◯ on a daily basis

D.3 Interview Introduction

The following should be read to the study participant:

> Thank you very much for your time and your willingness to help me. The purpose of this interview is for you to **help** us to determine how to improve Internet banking web sites.
>
> How this will work is, I will give you a scenario of an action that you performed on an Internet banking website. I will then show you an image of an Internet banking website after the given scenario. I will then ask you to tell me what you see, what you are thinking and how you feel the website handled the scenario.
>
> I would ask you to not worry about what you tell me. There is no right or wrong answer so feel free to say whatever comes into mind.
>
> Finally, would you mind if I do an audio recording of this session?

D.4 Screen Evaluations

Follow the steps below to evaluate each of the error scenarios included in sections 3.1 to 3.10 below.

1. Read the scenario out loud for the participant.
2. Show the screen shot to the participant
3. The following conversation points should be used to guide the interview
   (a) How well do you think the website handled this scenario?

(b) What would you do next, and why?
Note down whether

i. The error was noticed

ii. The warning content was understood

iii. Any assumptions the user made

(c) Can you think of anything else you would want the website to do?

E Updated Heuristic Evaluation Form

This appendix contains a questionnaire that can be used to evaluate error messages. This questionnaire was created from the guidelines for error message design for digital systems shown in Chapter 6.

Supreme Rule 1 - Necessary

Answer the following questions as best you can. The purpose of the questions is to lead thought and facilitate ideation treated accordingly.

No	Description	Comments
1.1	What is the basic or underlying cause of the error?	
1.2	Is this due to a limitation imposed by the system?	
1.3	Can the system conceivably be updated or changed to remove the limitation?	
1.4	In a perfect world, how should the system function to prevent the error from occurring?	

No	Description	Comments
1.5	Is the error due to an invalid user interaction?	
1.6	Can the user be prevented from performing the invalid interaction?	

Following the above analysis what is the answer to the following questions?

No.	Description	Answer	Comments
1 (a)	Is it possible to design out the error message?	○ yes ○ no	
1 (b)	Is it possible to guard against the error message?	○ yes ○ no	

Supreme Rule 2 - Noticeable

Answer the following question to determine how noticeable the error message is.

No.	Description	Score	Comments
2.1	Do you agree that the error message is visually conspicuous, successfully using size and contrast?	strongly disagree ○ disagree ○ neutral ○ agree ○ strongly agree ○	
2.2	Do you agree that the error placed close enough to the location of its cause?	strongly disagree ○ disagree ○ neutral ○ agree ○ strongly agree ○	
2.3	Do you agree that the error is displayed soon enough after it is caused?	strongly disagree ○ disagree ○ neutral ○ agree ○ strongly agree ○	
2.4	Is the error message styling consistent with other error messages of its class?	○ yes ○ no	
2.5	If applicable, do you agree that immediate in-line validation is used appropriately?	○ yes ○ no	

No.	Description	Score	Comments
2.6	If applicable, is the combination of summarising errors and form field errors used correctly?	◯ yes ◯ no	
2.7	If applicable, do you agree that the use of a modal error appropriate?	◯ yes ◯ no	

Supreme Rule 3 - Understandable

Answer the following question to determine how understandable the error message is.

No.	Description	Score	Comments
3.1	Do you agree that the tone of the error message is kind, helpful, and does not blame the user?	◯ yes ◯ no	
3.2	Do you agree that the language is uncomplicated?	strongly disagree ◯ disagree ◯ neutral ◯ agree ◯ strongly agree ◯	
3.3.a	Does the error message include details about the cause of the error?	◯ yes ◯ no	
3.3.b	Does the error message include details about how to recover from the error?	◯ yes ◯ no	
3.4	Do you agree that the error message does not include technical terms that a typical user might not understand, or that is not explained properly?	strongly disagree ◯ disagree ◯ neutral ◯ agree ◯ strongly agree ◯	

No.	Description	Score					Comments
3.5	Do you agree that the error message is unambiguous?	strongly disagree ○	disagree ○	neutral ○	agree ○	strongly agree ○	
3.6	Do you agree that the error message is as short as possible without compromising on the other guidelines?	strongly disagree ○	disagree ○	neutral ○	agree ○	strongly agree ○	
3.7	Do you agree that the icon use is appropriate i.e. 1. If icons are used do you agree that the icons are appropriate and necessary? 2. If no icon is present, do you agree that an icon is not necessary?	strongly disagree ○	disagree ○	neutral ○	agree ○	strongly agree ○