UNIVERSITY OF OSLO
Department of informatics

The effectiveness of novice users in usability testing

Master thesis
60 credits

John Lester Sarmiento Gerardo

1 May 2007

Acknowledgements

Working with this thesis was both an exciting and frustrating learning experience, but also incredibly rewarding. This work would not be completed without the contribution from a number of people, and I would like to take this opportunity to thank those who were directly and indirectly involved in this project.

First and foremost, I would like to thank my two advisors who have guided me the last year.

Alma Culén advised me throughout this thesis through helpful discussions, reviews and feedback.

Kent Westgaard guided me from start to finish and shared his knowledge through preparatory training, discussions, and feedbacks.

Monika Maanum helped me realise my proposal and supported me throughout the project, Jostein Håvaldsrud gave the green light to work with this interesting case, and Håkon Nordby helped me with the recruitment process of the research.

Aslaug kept criticizing my opinions and pushed me towards the first of many steps towards completion of this degree.

The members of Bruri Council provided study rooms, healthy babs and beverages.

Last, but not least, I would like to thank mamma, pappa and Bernadette for unconditional support and an infinite supply of delicious chicken soup.

Abstract

Current research is fine-tuning the usability evaluation methods in order to aid current and future usability practitioners. Some of the main research topics are sample sizes and appropriate statistical methods for analysing usability test data. However, few studies have focused on participants with different experience levels on test samples.

This thesis aims to fill this gap and add knowledge to the ongoing research on users with different experience levels in usability testing. The purpose of this study is to investigate the effectiveness of using novice- and expert users in usability tests.

23 individuals participated in a series of usability tests. The participants were grouped into novice and expert users. The results from the tests were analysed and compared in order to find out if novice users encountered the same type of usability problems as expert users. An additional analysis was performed to study if novices provided a larger number of problems than expert users.

The analysis suggests that novice users are likely to reveal the same type of usability problems as expert users in usability tests. Furthermore, the analysis suggests that novice users are likely to uncover a larger number of usability problems than expert users.

These findings implicate both the researchers' and usability practitioners' community. First of all, this thesis supports previous research on similar topics. Secondly, usability practitioners will benefit from this knowledge during the planning and recruitment process of usability tests.

Table of contents

1 - Introduction **2**
2 - Literature review **6**
2.1 - Evaluation - an important part of interaction design 6
2.2 - Evaluation methods 7
2.3 - Current research on usability testing 9
2.3.1 - Classifying users 11
2.3.2 - Novice and expert users 13
3 - Aims and research questions **18**
3.1 - Comparing types of usability problems between novice- and expert users 18
3.2 - Comparing the number of usability problems between novice- and expert users 20
4 - Methods **22**
4.1 - Usability testing 22
4.2 - Subjects 25
4.3 - Software 27
4.4 - Data analysis 32
4.4.1 - Analysing type of usability problems between novices and experts 33
4.4.2 - Analysis of the amount of usability problems novices 35
4.5 - Ethical considerations 36
5 - Results and analysis **38**
5.1 - Task completion 38
5.2 - Confidence intervals 40
5.3 - Task time results 40
5.4 - Questionnaires 42
5.5 - Observed problems 43
5.6 - T-test of the task time results 44
5.7 - Usability problems 51
5.7.1 - Updating information boxes 51
5.7.2 - Editing My Functions 52
5.7.3 - Editing shortcut keys 53
5.7.4 - Changing information boxes 53
5.7.5 - Managing Contacts 53
6 - Discussion **55**
6.1 - Novices, experts and types of usability problems 55
6.1.1 - Task completion 56
6.1.2 - Confidence intervals 57
6.1.3 - Task time results 58
6.1.4 - Questionnaires 59
6.1.5 - Problem list 61
6.1.6 - Usability problems 62
6.2 - Novices, experts and the amount of usability problems 65
6.2.1 - Task time results and t-tests 65
6.2.2 - Questionnaires 66
6.2.3 - Usability problems 67
7 - Conclusion **69**
7.1 - Limitations 71
7.2 - Recommendations for future work 73
8 - References **75**
Appendix A - Task scenarios **A-1**

Appendix B - Questionnaires B-1

Figures

Figure 1. One of the companies meeting rooms was used as a test lab 23
Figure 2. MBS version 10. The menu to the left opens the system's various modules. 29
Figure 3. MBS version 10. Clicking "Kontakt" on the left menu opens the system's Contact module 30
Figure 4. MBS version 11. The menu to the left has been changed. This menu opens various focus areas. 31
Figure 5. Clicking "Kundeoppfølging", which means customer service, opens up the Customer Service focus area 31
Figure 6. Experts' task completion results..... 39
Figure 7. Novices' task completion results..... 39
Figure 8. Novices' task time results..... 41
Figure 9. Experts' task time results..... 41
Figure 10. Number of novice users who rated the tasks as either difficult or very difficult.... 42
Figure 11. Number of expert users who rated the tasks as either difficult or very difficult 43
Figure 12. The information box on the bottom requires that the button Update on the top is clicked in order to show the latest information 51
Figure 13. Several participants did not understand how to edit the list in My Functions..... 52

Tables

Table 1. Confidence intervals for tasks 10, 11 and 12 40
Table 2. Group statistics for task 4..... 45
Table 3. Student's t-test of task 4..... 45
Table 4. Group statistics for task 5..... 46
Table 5. Student's t-test of task 5..... 46
Table 6. Group statistics for task 7..... 47
Table 7. Student's t-test of task 7..... 47
Table 8. Group statistics for task 9..... 48
Table 9. Student's t-test of task 9..... 48
Table 10. Group statistics for group 13..... 49
Table 11. Student's t-test of task 13..... 49

1 - Introduction

The demand for usability tests on graphical user interfaces have grown over the past years (Preece et al., 2002). The method is not new; it has existed for some time. However, stakeholders of various interactive systems are recognizing the value of this evaluation method, and are increasing its use to find potential improvements.

There are different ways to design usability tests. There are a variety of different methods and samples that can be used in these tests depending on the overall goal. However, the strategic decisions need careful consideration and clear goals in order to effectively find results.

Usability practitioners are challenged by limited budgets and limited time (Gulliksen et al., 2006). Some consider usability tests as expensive activities and few have the necessary resources to run tests thoroughly. While the method has become recognised, its use has been limited by such constraints.

Usability practitioners need new knowledge in order to perform these activities with fewer resources. Researchers have begun addressing this need the recent time. Some topics that have been investigated are issues such as the number of users that should be included in usability tests (Lewis, 2006), what type of users that should be included (Faulkner and Wick, 2005), and how to analyse the results (Sauro and Lewis, 2005). The answer to these questions might provide usability practitioners with an understanding of how tests can be conducted efficiently, while effectively producing results.

The last years have seen some refinement of the usability test methods. Researchers have spent time figuring out how many participants that are required to uncover as many as 80 percent of the problems, while others have researched on different techniques that should be used when collecting data from the tests.

The purpose of this thesis is, however, to study if there are any classes of users who reveal more usability problems than others. Should the sample in usability tests include a mixed set of participants? Should the tests only be run with expert users? Or will novice users provide the most effective results?

Several studies have investigated various uses of novice and expert users in usability tests. This thesis has reviewed available research literature about this topic. Some of the results from recent literature shows that different types of users reveal different types of usability problems (Faulkner and Wick, 2005).

This thesis further investigates the issue on the use of novice and expert users in usability tests, and will provide the ongoing research and usability practitioners with additional knowledge. The aim is to learn more about the effectiveness of novice users in usability tests. Will they reveal the same type of usability problems as expert users? If so, then which of the user types are recommended for usability tests?

It is also desirable to learn if novice users reveal a larger amount of usability problems than expert users on redesigned graphical user interfaces. Will they reveal more usability problems than expert users? Or is it the other way around? Will expert users provide the largest number of usability problems?

Several usability tests were run in order to find the answers to these questions. A selection of handbooks about available user research techniques provided the necessary description of how to design usability tests (Dumas and Redish, 1999, Kuniavsky, 2003, Nielsen and Mack, 1994). The tests were performed by using the talk aloud technique in order to understand what the participants were thinking during the tests.

Meanwhile, the participants were monitored and measured. The tests measured which tasks the participants were unable to solve. The second type of data measured was how much time they spent on each task. Thirdly, all the observed problems during the tests were logged and noted to provide an understanding of the results.

In order to measure their subjective experience with the system, the participants also had to respond to a questionnaire. These were handed out before and during each usability test. The questionnaires that were handed out before the test provided a background profile of each participant. The second questionnaire was used reveal which tasks the participants were experiencing as difficult.

23 individuals participated in the usability tests. The participants were split into two groups and were asked to try out a redesigned system. The novice users were represented by

customers of another similar system. The expert users consisted of existing customers of an earlier version of the tested system.

The participants were asked to perform a number of tasks on a beta version of an ERP system for small- and medium-sized companies. The tests used a version that had not been released to the market. Half of the usability tests were run before it was released, while the other half were run after release.

The data from the usability tests were analysed through different methods. Simple descriptive analysis was performed on most of the data. The analysis was performed to give an overview of the important trends in the data from the task completion results, task times and the questionnaires.

Confidence intervals were calculated to see how severe the problems were. The calculations show the percentage of the rest of the population that were likely to experience the same problems in similar usability tests.

Furthermore, t-tests were conducted to compare the task time results between the groups. The t-tests were performed in order to find out if there were any statistical significant differences between the groups. The results were used to find the answers to some of the research questions.

The remaining thesis is divided into six chapters. The following chapter describes the related research. It covers the main topics in the research field and critically analyses published papers on novice and expert users in usability testing. Chapter 3 describes the aims of this study and the research questions. Chapter 4 describes how the study was conducted. The chapter thoroughly describes the problems that arose when deciding which methods should be used, which decisions were made, and why. Chapter 5 presents a summary of the results and analysis. Chapter 6 discusses the summary in chapter 5 and relates it to similar published results. Chapter 7 provides the conclusions of this thesis.

2 - Literature review

The following chapter describes current literature about usability testing. Chapter 2.1 presents some background history of usability evaluation in interaction design. Chapter 2.2 describes various usability evaluation methods. Lastly, chapter 2.3 summarizes some of the main research topics and critically reviews the research literature about experience levels in usability testing.

2.1 - Evaluation - an important part of interaction design

Interaction design is the process of "designing interactive products to support people in their everyday and working lives" (Preece et al., 2002). However, the way these products were designed has not always been the same. Interaction design evolved along with the users of technology.

Engineers designed the earliest hardware systems. The systems were easy to operate provided that one possessed the necessary knowledge about these systems. However, problems arose when people who were not engineers attempted to use them. To solve this problem, psychologists and computer scientists began designing graphical user interfaces (GUI) together. Psychologists provided knowledge about the human mind, while the computer scientists developed the GUI's. This marked the beginning of interaction design as a growing field. The field mixed computer science with other disciplines.

The use of computers increased for training and educational purposes during the 1980's. Psychologists and computer scientists' expertise were no longer sufficient; the new requirements demanded knowledge from other fields. It became necessary to include experts with educational background in order to develop educational software.

Technology developed further during the 1990's. The use of the personal computer increased and became useful for a range of different people, alongside the growing importance of internet and mobile phones as communication tools. The wide use of technology in everyday life required further expertise. Consequently, the field acquired expertise from fields such as graphic design and sociology in order to meet the changing demands.

Today, businesses such as software companies and mobile computing industries consider usable products as an advantage in an increasingly competitive market. Developing

information technology is cheaper and the quality has improved, forcing companies to search for other ways to differentiate their products. A typical sales argument could be that their products are “user-friendly”. Interaction design is now a recognized field, and is also becoming a growing commercial service.

The process in interaction design involves four activities; the requirements gathering, prototyping, designing and evaluation. This process is iterative similar to other software development life cycles. This means that the process is repeated in several iterations until time or resources limits are reached.

An important term in interaction design is the user centred approach (Norman, 2002). The term describes the process where the development of a product is based on putting the users in the front seat in order to gain more knowledge about what the users want, how they will use it, and if it is effective for the intended end-user.

There is an important activity in interaction design that can be performed in order to extract this knowledge; the evaluation of the product. Several methods and techniques for evaluating how users interact with systems exists in user centred development. These methods and techniques offer usability practitioners frameworks for evaluating GUIs and how potential users might use proposed systems.

The evaluation phase is relevant for this thesis, as it deals with the evaluation of GUIs. This can either be done with actual users, or with experts in the area. The purpose is to find out whether or not the GUI supports the usability criterions that were initially set for the product.

In short, the interaction between people and the product should be optimized, and one way of making them usable is to make sure the product is effective, efficient, safe, and easy to use.

2.2 - Evaluation methods

There are various methods for evaluating usability. There exists a selection of methods and techniques for evaluating user interfaces, like for example heuristic evaluations (Nielsen and Mack, 1994), focus groups (Kuniavsky, 2003), usability testing (Dumas and Redish, 1999), and questionnaires (Nielsen and Mack, 1994, Rubin, 1994).

Heuristic evaluations are performed by a group of evaluators who inspects if a GUI is consistent with heuristics or usability principles relevant for the system (Nielsen and Mack, 1994). The main goal of a heuristic evaluation is to uncover potential usability problems. They are conducted without involving users and can be completed in a couple of days.

However, the technique has been criticized for not being effective in revealing novice users' usability problems. A study showed that while heuristic evaluations effectively reveals expert users' usability problems, there are other usability evaluation methods, such as usability testing, that are more effective in revealing usability problems (Fu et al., 2002).

Focus groups are another way of evaluating a system. This is done by gathering a group of potential users where the discussion is controlled by a facilitator, and where the participants discusses and provide feedback concerning the product (Kuniavsky, 2003). However, its goal is not to infer what the target groups wants. Focus groups are useful tools for understanding the user's perceptions. The sessions can help finding their desires, motivations and values. It is a technique that helps understand what the end users desire and is therefore a useful tool in the early phases of a development project.

Another method that can provide insight about the end users is questionnaires. Performing a survey can gather information from a large population. It is also a less time consuming activity and cheaper to perform. However, they have some drawbacks. Questionnaires must be designed carefully in order to return reliable results. There are many pitfalls in the design of questionnaires that can affect the respondent's response. Questionnaires also lack the type of information that can be gathered through qualitative research. Questionnaires can tell you what is going on, but cannot explain why it is happening.

Lastly, there are usability tests, which is the main method in this thesis. The main purpose of these tests is to uncover usability problems. This is an important distinction. There exists various books that describes in detail how to perform usability tests (Dumas and Redish, 1999)

A usability test involves observing potential or actual users use a product (Dumas and Redish, 1999). This activity usually involves two roles; the participant and one or more evaluators. The evaluator gives the participant a set of tasks, which she is asked to perform. The evaluator observes the interaction between the participant and the interface. The observations are measured and noted. The results are then analysed so that the evaluator

can create an overview over potential usability problems. These usability problems describe the possible problems and flaws that are in the design that are likely to cause them.

Even though this method uncovers usability problems effectively, the technique has been criticized for a number of reasons. The following section provides some discussion on this theme.

2.3 - Current research on usability testing

Current research issues on usability testing are focused on fine-tuning and honing the evaluation technique (Faulkner and Wick, 2005). Some of these issues include various think-aloud techniques, sample sizes, evaluator effects, the use of inferential statistics to analyze data, and the use of novice and expert users in usability tests.

One of the issues that has received attention are the various think aloud methods used in usability testing. One recent paper has studied the use of Retrospective Think-Aloud method (RTA) in usability testing (Zhiwei et al., 2006). In contrast to the Think-Aloud method, where the participant speaks out loudly what she is thinking during the test, Retrospective Think-Aloud requires the participant to solve tasks on her own. The discussion on the issues that were experienced should be held after the task.

The authors of the paper argue that this method benefits usability practitioners in at least two ways. First of all, the think aloud technique asks participants to verbally express what they are thinking while solving a task. This aids the practitioners understanding of what the participants are thinking.

However, this can also create an unnatural setting. In reality, users of a system do not talk about what they are thinking while using a system. The RTA method is performed by asking participants to solve the tasks on their own, and talk about the issues after the task has been completed.

Secondly, the think aloud method may affect the collected data. Some people freely express their thoughts during the tests. Others are more reluctant to talk. This may affect the usability test results when for example measuring how much time the participants are spending on a task.

For instance, participants who talk less may spend less time performing tasks. They try to solve the tasks they are given on their own. Other extrovert participants who talk freely about their thoughts may spend more time on the same tasks. The reason might be that they spend the time they are allocated for the task to explain the evaluator her actions. The RTA method attempts to remove this difference by not allowing any interaction with the participant during the tasks.

Another issue in the research of usability testing methods is sample sizes (Lewis, 2006). The reason for studying sample sizes is because usability tests are time consuming and expensive. Usability tests are usually a small part of development projects and have low priority (Gulliksen et al., 2006). Studying effective sample size benefits usability practitioners by cutting the costs of their practice. Various studies have investigated the most effective sample sizes in order to cover these needs (Lewis, 2006).

Researchers have been arguing about sample size for several years, but Lewis (2006) argues that sample sizes depend on the context of the usability test. First of all, he recommends determining the variance of the dependent measures of interest. This is usually obtained from previous usability tests. The second requirement is to determine how precise the measurements should be.

These arguments suggest that there is no fixed sample size that can reveal a maximum amount of usability problems. The author of the paper suggests that the sample size should be estimated prior to each usability test.

Another aspect on usability tests that has been studied is the evaluator effect (Hertzum and Jacobsen, 2003). The authors of the study argue that different evaluators observe different problems. There can be several reasons for this, but they suggest that the differences are rooted in their previous experience in the usability field.

The authors' claims are based on an experiment with experienced and less experienced usability practitioners. Their findings indicate that experienced usability practitioners observe more critical usability problems. They are able to observe more problems because of their previous experience performing usability tests.

The authors also point to that less experienced practitioners observe less usability problems. Because of their lack of experience, they have a more difficult time observing

and noting serious problems. In order to track most the problems, it is recommended to use an experienced evaluator, or add an additional one next to the novice user in order to track most of the problems.

While many researchers have focused on refining the methods for gathering usability test data, other researchers have published suggestions about how the data should be analysed. One of the challenges usability practitioners faces when presenting their results is the demand for inferential statistical analysis of the usability test data (Sauro, 2006b, Sauro and Lewis, 2005). Usability practitioners with a background in statistics might be able to solve these problems. However, there are some practitioners that do not possess this competence.

Some researchers have published suggestions to how small sample sizes such as task completion data should be analysed in order to aid these practitioners (Agresti and Coull, 1998). The authors presented several methods to calculate the confidence intervals from test data and recommended different methods for different types of results. The methods are based on the authors' earlier research on confidence intervals for small sample sizes (Sauro and Lewis, 2005).

The authors recommend using the Adjusted Wald method for most situations. In short, this involves adding two successes and two failures to the task completion results, and thereafter calculating results using the traditional Wald method. The authors recommend this method because it is easy to use. However, the method can not be used for all situations, but this will not be discussed further in this thesis.

Another important issue in the research of usability testing is the classification of participants and usability problems. This is the central topic for this work. While other issues such as sample sizes, measurements and the use of inferential statistics in usability are equally important, they will not be covered in this thesis review. The use of expert and novice users in usability tests is the main theme of this thesis. The following chapter will critically analyse and synthesize the literature on this topic.

2.3.1 - Classifying users

Early literature by for example Nielsen (1993) makes a distinction between categories of users and their individual differences. He describes three main dimensions of users'

experience. These are the user's experience with a system, with computers in general and with the task domain. Furthermore, the user's experience level with a system has another dimension; they are either a novice or an expert.

A user's experience with a system is determined by how long and how much time a person has used a system. For example, a person who has used a web browser for less than a week can be classified as a novice user. On the contrary, an individual who has used the browser for more than a week may be classified as an expert user.

Experience with computers in general is also an important dimension when categorising a user's experience levels. There are for example differences between people who use the computer for work-related tasks, compared to people who use the computer for private purposes. A person who only uses the computer for work has another relationship with computers because it is used as a tool for work. People who are using the computer to manage personal pictures or to work on their hobbies may perceive and use the computer in a different way.

The user's domain knowledge is the final and equally important dimension. The GUI needs to be designed with the end user in mind. For example, software for users with a background in finance needs to be designed in a way that they understand. At the same time, the software needs to be designed differently for people without the same domain knowledge in order for the end user to be able to use the software.

While Nielsen describes six different categories (Nielsen, 1993), current research with different user types typically categorises participants as either novice or expert users (Dillon and Song, 1997, Faulkner and Wick, 2005, Fu et al., 2002, Goonetilleke et al., 2001).

While this might be the common method of grouping participants, categorising users depends on the context of the research (Faulkner and Wick, 2005). For example, in a study on the effects of training on performance, the research included novice users, and created expert users by giving them short training prior to the experiment (Goonetilleke et al., 2001). For this study, an expert user had less than a day's experience with the system. Measuring the user's experience level therefore depends on the context of the test, and varies from study to study.

Even though the various studies categorises users in different ways, most of them usually experiments with users who are new to the system, and with users who knows the system beforehand. The next chapter will present the current research on the use of novice and expert users during evaluation of usability.

2.3.2 - Novice and expert users

Current research classifies users in various ways. Some classify users as described in the previous chapter, with different levels of knowledge, use of computer and use of the systems. Others use their own, specified classifications of users.

However, most of the research literature distinguishes between users by classifying them as either novice or expert user. The meaning of novices and experts vary depending on the research. A novice is usually a user with no previous experience with a system, while the expert usually has previous experience with the system prior to the research. In other contexts, both novice and expert users have neither tried the system. In order to create the expert users in these situations, a group of individuals receive a short period of training prior to the research.

There have been few studies on this topic. One early study has investigated the difference between novice and expert users in computerized office work (Prumper et al., 1991). The authors of that paper studied how novice and expert users interacted with computers in office work. They observed 174 clerical workers from 12 different companies.

The findings from this study showed that expert users not necessarily revealed fewer errors than novice users. The expert users in their study reveal more problems than the novice users. This was contradictory to their initial belief that novice users would uncover the largest number of problems.

Other studies have focused on the performance levels between two types of users and two types of user interface (Dillon and Song, 1997). The study was performed from a different angle by performing an experiment with a search engine. The purpose was to investigate how novice and expert users performed on a text-based search engine, and another search engine with a graphical user interface.

The study compared the results from several angles. Their findings suggested that expert users performed faster than the novice users. The expert users in their study managed to complete the tasks faster than the novice users.

In addition, the results also indicated that novice users would fail more tasks than the expert users on textual user interfaces. While their results did not show any significant differences, their task completion rates showed that expert users completed more tasks than the novice users in the textual user interface. Their study suggests that expert users perform better than novice users in studies on their interaction with user interfaces.

Other studies further confirm these claims. In a separate paper, some researchers experimented with novice and expert users by measuring the response times for understanding various icon designs (Goonetilleke et al., 2001). The authors experimented with different icon designs and measured the response times for understanding the different icons. The researchers were trying to find out if training prior to the experiment had an effect on the response time contra the effect of no training.

They used 30 participants and grouped them into two types of users in the experiment similar to the one above; expert and novice users. Prior to the experiment, the participants were given a short period of training who would then represent the expert users. The other group received no training and represented the novice users.

The study found that expert users, who were given training prior to the experiment, had different mental models of the icons they were presented with, and performed well in an experiment where they were asked to associate verbs and objects with given icons. The novice participants on the other hand, who lacked the training prior to the experiment, had longer response times.

While previous studies focused on classifying users by novice or expert users, Fu *et al. (2002)* conducted a study with another set of names for the categories. These categories were the skill-based-, rule-based- and knowledge-based levels, and were used to distinguish the type of usability problems they uncovered using a heuristic evaluation compared to a user test.

While these categories were named differently than the previous studies, the category names also had some similarities. The study included 12 participants. In short, the skill-

based- and rule-based levels correspond to the expert users mentioned in the two studies mentioned previously, while the knowledge-based levels correspond to the novice users.

The study found that usability testing was effective in revealing usability problems with novice users. The results from the novice users proved to be more effective than with expert users. At the same time, heuristic reviews were found to be effective in finding the usability problems that the expert users were likely to experience.

The study concluded that usability tests were more effective than heuristic review in uncovering usability problems with novices. While it is ideal to include novice and expert users in usability tests, and in addition run a heuristic review graphical user interfaces, their paper suggests that novice users should be prioritised over expert users in usability tests. Heuristic review is a less resource spending activity to reveal the expert users' usability problems.

This is also supported by another study on usability testing with children. Donker and Reitsma (2004) found that novices experience significantly more usability problems than experts. The study was conducted by observing 70 children using educational software. The study showed that the novice children encountered more problems than the experts while using the software. In addition, the novices were less aware of the GUI's shortcomings.

Further research was conducted by comparing three different categories of users in a usability test on a time sheet application (Faulkner and Wick, 2005). The authors gathered 60 participants and grouped them into three categories. The categories had two dimensions; how much experience the participants had with computers in general, and how much experience the user had with the specific application they were testing with. By varying these two dimensions, a participant could either be a novice\novice, expert\novice or expert\expert. Novice\novice would then correspond to a participant who had little experience with computers in general and had no prior experience using the application. Expert\novice corresponded to users who had used a computer in more than a year, but had never tried the application that was being tested. Lastly, expert\expert users corresponded to users who had used computers in general for more than a year, and had prior experience with the intended application.

While the results showed that there were both similarities and significant differences between the group's performance and problems, the authors concluded that categorising in this way provided an additional layer of information to the usability problems, and therefore could tell the cause of the problem. While current usability tests may reveal the existence of a usability problem, they are unable to define why they exist. By screening and categorising the participants beforehand and analysing the results through these categories, a usability practitioner no longer has to depend on their cognitive interpretation of the problems to define who is likely to experience them. Categorising the users beforehand gives the evaluator information about the types of users that are most likely to experience the problems.

There has been little research on this subject. This literature review shows five studies on this topic. Of those four, three aimed at fine-tuning the usability test technique (Fu et al., 2002, Donker and Reitsma , 2004, Faulkner and Wick, 2005). The remaining two studies dealt with the performance between novice and expert users (Goonetilleke et al., 2001) and the effects of training (Dillon and Song, 1997).

Furthermore, in the study on various icon designs, the author describes how they created expert users (Goonetilleke et al., 2001). The study is limited by the fact that the expert users did not exist prior to the experiment. The study included a group of users that were different from what is typically an expert user in a real setting. All of the participants were novice users prior to their study. The expert users were created by adding some training.

The other study which classified usability problems by performance levels was limited in a different way (Fu et al., 2002). The article argues that user testing is more effective in uncovering usability problems than heuristic evaluation. However, their study only included 12 participants. While the authors do claim that 12 participants will uncover 98.8 % of the usability problems, they further explain that six persons participated in their user test, while the remaining six were usability experts who performed the heuristic evaluation. Furthermore, the subjects were internal employees and the tested application was an internal web-based training application. The research could therefore be limited by both their sample size and the participants' background.

The research on search engines (Dillon and Song, 1997) and icon designs indicate that novice and expert users perform differently. While their research is not directly related to

detecting usability problems in usability testing, they do show how the two experience levels perform differently; expert users perform better than novice users. However, one should not ignore that the study that the study only consisted of novice users, and created the expert users by giving them a small amount of training prior to the research.

Based on my examination of the literature, it seems that there is a lack of research on this area. There have been research on novice and expert users in relation to usability testing, but there are few comprehensive studies. In addition, most of the literature describes research that was conducted on web-based software. In other words, there is a gap in the research when it comes to assessing the differences between novice and expert users in usability testing. This thesis is an attempt to fill that gap.

3 - Aims and research questions

The following chapter presents the aims and research questions of this thesis. Chapter 3.1 presents the first aim which is to investigate if novice users reveal the same usability problems as expert users. Chapter 3.2 presents the second aim which is to investigate if novice users uncover a larger number of usability problems than expert users.

3.1 - Comparing types of usability problems between novice- and expert users

Several studies have researched on the effectiveness of novice- and expert users in usability tests. As seen in chapter 2.3.2, studies show that novice- and expert users encounter different types of problems. On the contrary, other studies show that different types of users detect the same type of problems.

However, the first overall aim of this thesis is to study this issue from another angle by investigating if novices and experts are able to reveal the same usability problems on redesigned GUIs. Will novice users reveal the same usability problems as expert users who have previous experience with the system? Or will novices fail to uncover the problems revealed by experts in usability tests?

The answers to these questions will contribute to the ongoing research. Some of the research has pointed out the benefits of including participants with different experience levels in usability tests (Faulkner and Wick, 2005). This thesis however, will look for benefits or drawbacks of only including novice users in usability tests. The findings from this study will therefore contribute to the research on this subject.

This topic is equally important for usability practitioners. Some usability practitioners are unable to perform usability tests because of the lack of resources. This is one of the reasons why there has been research on usability testing. As mentioned earlier, research has focused on issues such as the lowest number of users to include in usability tests, and which techniques that provides the most accurate results. Studying the effectiveness of novice and expert users in usability tests on redesigned GUIs, will help usability practitioners decide which types of users to include in tests. Hopefully, the same knowledge will also reduce the costs of their practice.

For example, if the purpose of a test is to uncover as many usability problems as possible, the answers from this thesis, combined with other studies, will provide usability

practitioners with state-of-the-art knowledge regarding the types of users that recommend recommended to include tests.

This thesis will focus on five variables in order to learn more about the similarities between novice- and expert users. These variables are empirical data collected from usability tests with both novice and expert users. The tests will be conducted with existing software, and its existing and potential users. Both qualitative and quantitative data will be collected for analysis and will also serve as supporting evidence for potential usability problems.

First of all, it will be necessary to study the tasks the users are unable to solve. Comparing the tasks that both groups are unable to solve will give the first clue to possible usability problems. Will the novice users fail the same tasks as the expert users? Or will the novice users manage to solve the tasks that the expert users are unable to pass?

Secondly, the confidence intervals for the task completion results will provide an estimate for how many users in the total population that are likely to fail the same tasks. While the task completion results shows how many participants in the test sample that failed to solve each task, the confidence intervals will calculate the percentage of the groups' population that are likely to fail the tasks. Are the novice users' population likely to fail the same tasks as the expert users' population?

Thirdly, it will be necessary to study which tasks the participants are spending the most time on. The task time must be measured in order to find out which tasks both user groups spend too much of time on. Will the novice users spend a large amount of time on the same tasks as expert users? Or are there tasks where the novice users are spending less time than experts?

The fourth variable that needs to be measured is the users' subjective experience of each task. While the previous data measures the participants' performance, their subjective experience will reveal if both groups are experiencing any parts of the GUI as difficult to use. In order to reveal which parts of the GUI they are experiencing as difficult to use, it is necessary to study which tasks both groups rate as difficult to solve. Will the novice- and expert users rate the same tasks as difficult? Or are there tasks where only the expert users are experiencing difficulties?

The fifth variable is a list of observed problems during the usability tests that will help identify what the problems are. While the quantitative results might indicate that there are flaws on the GUI, the problem list will help describe them.

3.2 - Comparing the number of usability problems between novice- and expert users

None of the reviewed literature has focused on the novice and expert users of redesigned GUIs. An interesting subject is the number of usability problems that novice- and experts users provide when testing a redesigned GUI. Which type of user is likely to uncover the largest amount of usability problems?

This is a central subject for the second aim of this thesis, which is to find out if novice users reveal more number of usability problems than expert users when conducting usability tests on redesigned GUIs. Will the novices reveal a larger amount of problems than experts? Or will the novices reveal less usability problems?

Two variables will be studied in order to answers these questions. The focus will be on the differences between the groups' results. First of all, it is necessary to look at the task time results. This makes it possible to track if the novices are spending more time than experts on any of the tasks. Will novices spend more time on tasks than the expert group? Or will both groups spend an equal amount of time on all tasks?

Secondly, it is necessary to look at which tasks the groups are rating with different levels of difficulty. Even though both groups may be able to solve a task, one of the groups might perceive the same task as difficult to solve. It is therefore necessary to analyse the results for both groups. Which tasks are rated with different levels of difficulty? Are there any tasks where the novices are rating the tasks as more difficult than experts? Will both groups rate them similarly? Or are there tasks where the experts rate them as more difficult than novices?

4 - Methods

The following chapter presents the methods that were used to perform the research. Chapter 4.1 presents how the qualitative and quantitative data was collected through usability testing. Chapter 4.2 provides an overview of the participants in the usability tests. Chapter 4.3 describes the software that was used for the tests. Chapter 4.4 presents how the analysis was performed. Chapter 4.5 discusses the ethical issues around the usability tests.

4.1 - Usability testing

Usability tests can be conducted through a variety of techniques. The tests have to be designed based on the overall goals. Some usability tests collects data through observations and are usually performed to provide quick feedback. Other tests are more comprehensive and systematically collects both qualitative and quantitative data. In addition, they perform statistical analysis to provide better supporting evidence for the potential of usability problems.

The data that was collected to uncover the usability problems in this study were a variety of qualitative and quantitative data. The data was collected from a series of observations and questionnaires (Figure 1).

Figure 1. One of the companies meeting rooms was used as a test lab

In order to perform the test and collect data, the participants needed tasks to solve. A test was conducted by asking each participant to solve 14 tasks (see appendix A). The company responsible for the software provided 9 tasks, while the remaining 4 were made by the author.

During the tests, the participants solved the tasks while the evaluator measured their performance. The tests were performed in short time, and it was therefore necessary to videotape the sessions to allow further reviews and analysis.

The quantitative data was collected through subjective and objective measures. The subjective measures were collected through questionnaires. During the tests, the participants received questionnaires after each task (see appendix B). They rated how easy or difficult it was to solve a task through a Likert-scale based questionnaire. The purpose was to analyse which tasks the participants experienced as difficult to solve.

Several objective performance data was measured during the tests. The first performance measure was the participant's effectiveness. The effectiveness was measured by observing

if the participants managed to solve the task or not. The purpose of this measure was to reveal which tasks the participants had difficulties solving.

Their efficiency was measured at the same time. The time the participants spent on each task was measured from when they were given the tasks, until they solved it, gave up or spent more than five minutes. The purpose of this measure was to uncover which tasks the participants spent the most time on.

The qualitative data was collected through observations. Whenever a participant experienced problems during the tests, they were noted down in a problem list. The problem list consisted of all the problems that were observed during the tests, and how many times they occurred.

While the methods used for this research provided a large amount of data, the measurements were not always accurate. For example, even though the efficiency was supposed to be objective, it was actually a subjective measure because it was the evaluator's responsibility to start and stop the timer. Some of the participants started on the tasks while they were being read out orally. The timing would therefore not be 100 % accurate.

While keeping track of the time, the evaluator also had to keep a running list of problems. These problems were identified by the evaluator, meaning that the problems that were observed were the problems that were put in the problem list. That also means that the problem list depends on what the evaluator perceives as a problem. The results might have been two different sets of problems if there had been two evaluators in the test. This is, as mentioned in chapter 2.3, known as the evaluator effect (Hertzum and Jacobsen, 2003). According to another study, an evaluator's experience affects the amount of problems that are revealed during usability tests. The test could therefore have revealed more problems if the evaluator had more experience, and if there were more than one evaluator. However, this was not possible for this project, because of limited resources.

Some books also recommend measuring frustration during the tests. By measuring this, an evaluator can measure which tasks that led to frustration. This was however not appropriate for this study. While some of the participants reacted with frustration during the tests, other participants did not express it as much even when they had problems solving the same tasks. Possible explanations might be that the participants had different

personalities; some were extrovert, while others were introvert. Other possible explanations might be that some of the participants did not want to express anything negative during the tests.

Frustration was initially measured during the tests, but was stopped after a while. The level of frustration the participants expressed differed between the participants, resulting in useless data.

4.2 - Subjects

This study was originally conducted with 24 participants. In order to answer the research questions, it was necessary to plan ahead before the participants were recruited. The participants needed to have the required profile to be able to provide the answers to the research questions. First of all, the nature of the research questions required participants without prior experience with the new desktop.

Secondly, the research questions required two groups of users with different experience levels. One group needed to represent the novice users while the other group would represent the experienced users. In addition, it was desirable with an equal amount of male and female participants to reduce biased results.

The third requirement was to include 12 participants in each group. The reason for this number was a previous study which showed that 98 % of the usability problems would be uncovered with 12 participants in a usability test (Fu et al., 2002). In addition, it is always desirable to have as many users as possible to uncover as many problems as possible and to make sure the research is valid.

The last requirement for the selection of participants was that they had similar domain knowledge; they should all be users with experience using ERP systems in small- or medium sized company.

The company's research and development-, and marketing department were responsible for the recruitment process. They decided to invite existing customers of the products Mamut Business Software (MBS) and daTax Totaløkonomi through email. Several customers replied and were added to a list of possible participants. The customers in the list were crosschecked in the company's customer database to confirm that they matched

the profile. The customers who matched the profile were contacted in order to book a date for each usability test.

The group of MBS customers represented the expert users. They were existing customers from small- and medium sized companies. At the time, they were using MBS version 10, and were going to receive version 11. This group consisted of 12 existing customers of MBS, six of them females and the remaining six males.

The existing customers of daTax Totaløkonomi had not used MBS prior to the test, and represented the novice users. The novices consisted of 12 participants who used daTax Totaløkonomi for business purposes. They had been customers for several years, and had the same domain experience as the expert users.

However, it was later revealed that one of the novice users was a beta customer. He had previous experience with both products and did not fit the profile. The results from this participant were therefore removed.

The usability tests were conducted with a total of 23 participants. While this is beyond the recommended number of participants for regular usability tests, other similar research projects included 60 participants (Faulkner and Wick, 2005), which is more than twice as much as in this study. The total amount of participants in this study is therefore a possible weakness.

Other weaknesses with the participants were the novice group. One participant revealed that he had used an older version of MBS, but before version 10. He matched the profile because the version he used was older than MBS version 10. However, the structure of the system's modules has been the same for some time, meaning that he would probably have a different mental model of the system than the others in his group. His performance measures could therefore be biased.

The target group was customers who were using daTax for business purposes, but this was only partially achieved in this study. Some of the participants revealed during the test that they did not use the software for business purposes, but for personal use. Others explained that they were senior citizens. However, by the time they showed up for the test, it was already late, and it was not possible to send them back or invite more participants because

of the incentives they were promised. Some of the participants did therefore not match the profile 100 %.

While recruiting the participants, it was also intended to recruit an equal amount of participants of each sex. This was achieved with the expert group; they consisted of six females and six males. On the other side, an equal amount of participants of each sex was not achieved for the novice group. This group consisted of 10 males and two females. The reason was a larger amount of respondents from male users than females. It was therefore not possible to have an equal amount of each.

While this study could have included more participants, the 23 participants did provide answers to the research aims. The usability problems were discovered, and the two user groups provided answers concerning their impact on usability tests.

In addition, even though the novice group initially did not match the profile as intended, they did represent the novice users well. With the exception of one person, who had used an older version of MBS, the group did represent potential customers and provided feedback about how they might have interacted with the system by themselves.

4.3 - Software

As mentioned in chapter 2.3, earlier studies have focused on the differences between novice and expert users in usability testing. The authors of these studies performed usability tests on a variety of information systems. Some of these GUIs were early prototypes with little functionality. Other studies included complete systems.

Some of the systems that were used in the studies mentioned above did not have any actual users. In order to perform their experiments with expert users, they created the expert users by providing them with a short period of training prior to the experiment. These were some of the limitations of the studies. Few based on the use of real-world software.

As mentioned earlier in the research questions in chapter 2, the purpose of this study was to investigate which users group that provides the most effective results in usability tests after redesigning a system. The purpose is to find out which user group one should focus on when introducing a new graphical user interface for the users.

This study required an actual information system used by several users in order to find the answers to these questions. The study required that the information system had been used by several users for years during their daily work. It needed to be a system that was well established and had gone through several official versions.

Furthermore, this information system should undergo major redesigns. The redesign should have such an impact on its users that it changes the way they are working. Their regular tasks should be performed in a different and hopefully a better way.

At the time of the study, the existing customers of MBS had access to version 10 of the system. This version included a desktop with a menu on the top and another menu on the left as shown in figure 2. The large space on the right side of the interface consisted of an information desk providing various articles and access to the product's support service.

The menu on the top of the window gave access to most of the systems functionality. These included links to the system's various parts such as the system's settings, maintenance functions, and help files. This menu was similar to other Microsoft products such as Word and Excel.

MBS consisted of several modules. An example of these was the Contact-module, which represented the Customer Relationship Management part of the system. This module allowed users of an organisation to manage their Contacts. Examples of such management tasks include adding information about new vendors or updating the information of an existing contact.

Another module is the Order-module. This part of MBS allows the user to perform tasks such as creating new sales orders on existing customers. This exemplifies the ERP part of the system. While one can create a customer in one part of the system, another part allows one to retrieve this customer and register an order on this person. As soon as the order has been created, the information is immediately available for those who are working with the customer information. The system's different modules are connected in this way in order to support these business processes. MBS supports this by integrating several processes through one system.

As shown in figure 2, the menus on the left side of the desktop are links to the various modules. Figure 3 shows what happens after clicking on Contact on the left menu. The Contact module of the system opens up in a new and smaller window to the right.

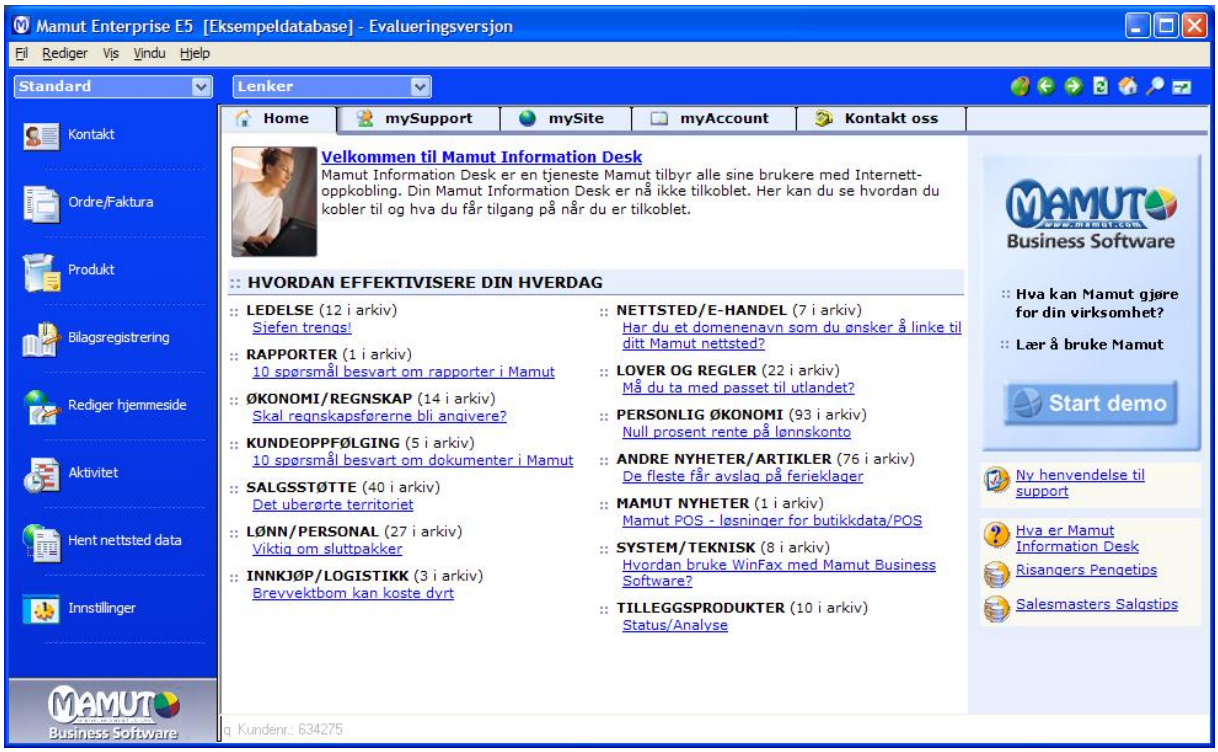


Figure 2. MBS version 10. The menu to the left opens the system's various modules.

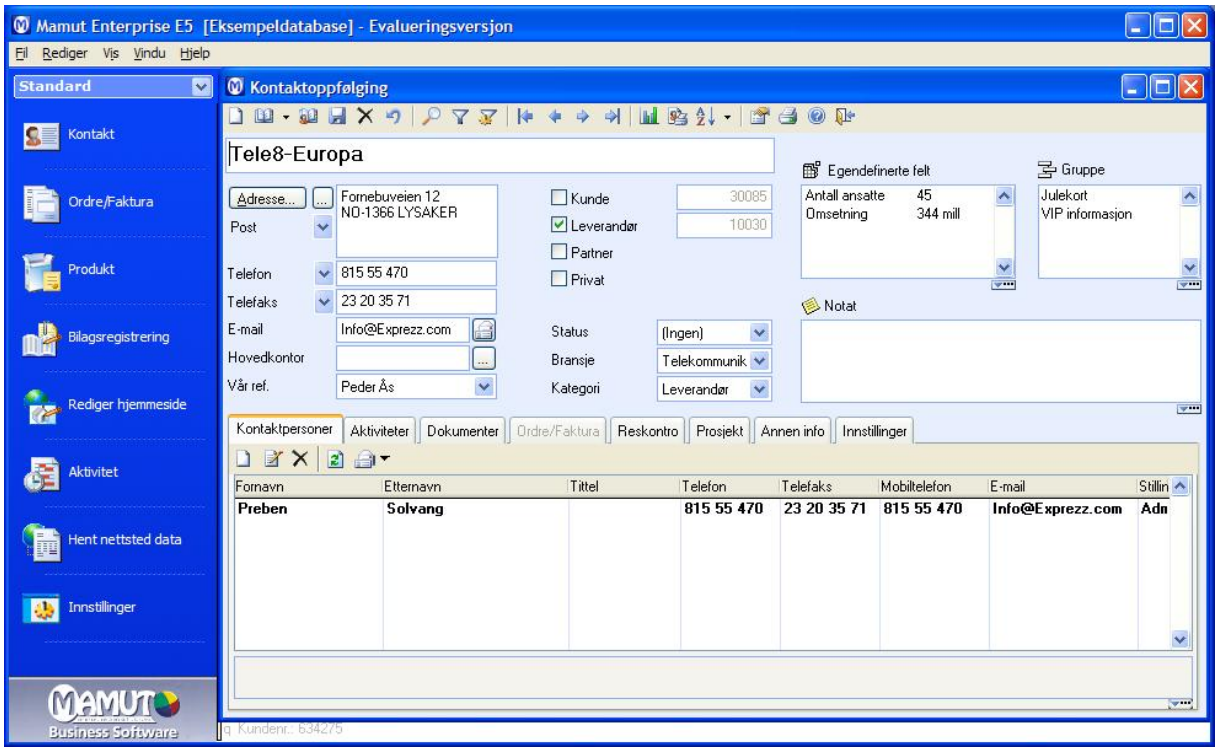


Figure 3. MBS version 10. Clicking "Kontakt" on the left menu opens the system's Contact module

However, this interface changed in version 11. While the top menu remained unchanged as shown in Figure 4, several changes were made to the menu on the left and the information desk. The links on the left menu no longer referred to MBS' modules; they were replaced by terms of various activities. For example, the previous menu allowed the user to open up a specific module like the Contact module. The new menu was changed. For example, if a user was following up a customer, clicking this option on the menu would open up a card instead of a module, as shown in (Figure 5).

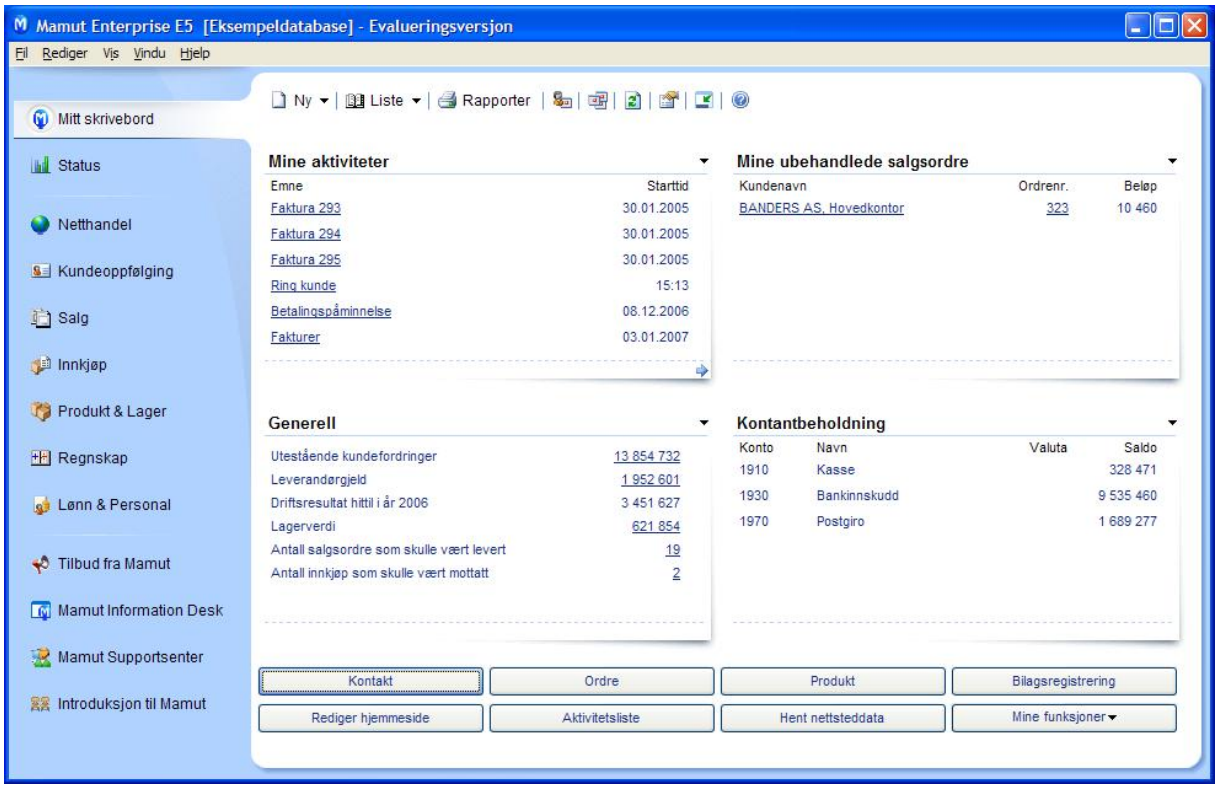


Figure 4. MBS version 11. The menu to the left has been changed. This menu opens various focus areas.

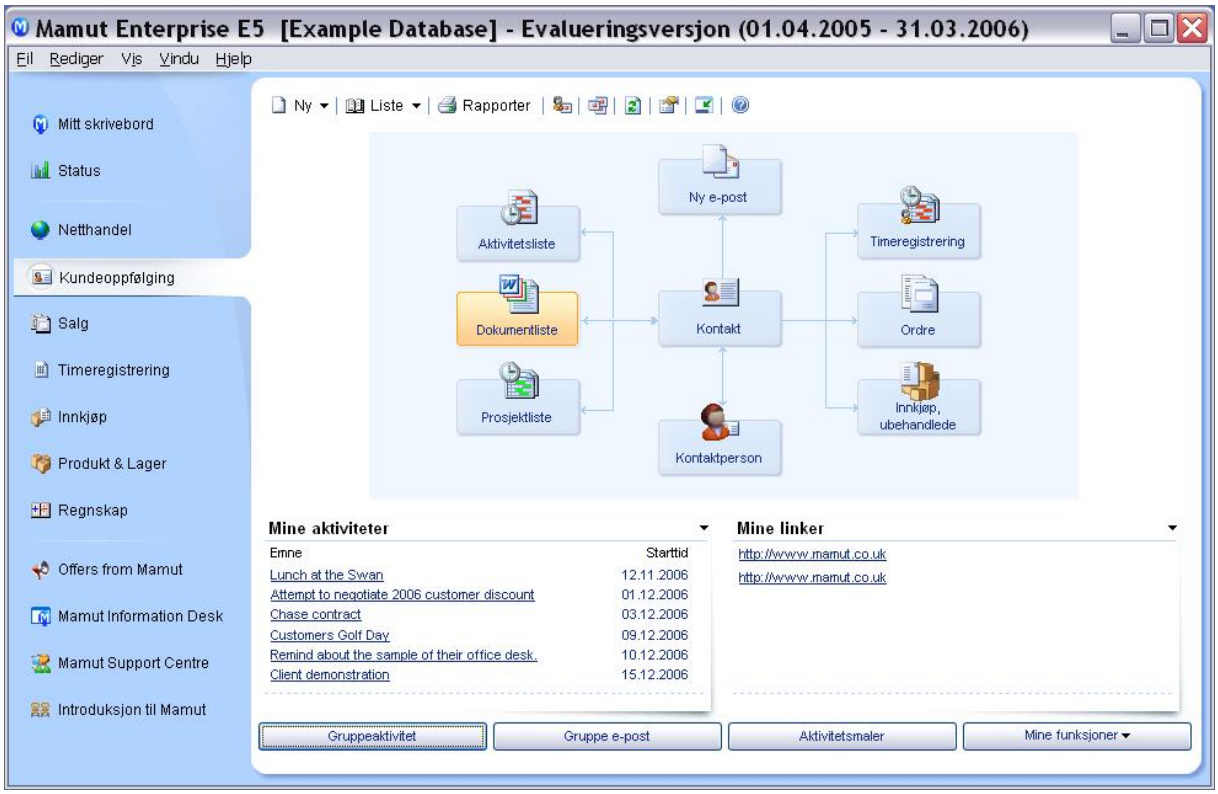


Figure 5. Clicking "Kundeoppfølging", which means customer service, opens up the Customer Service focus area

These cards consist of a selection of buttons that are connected to the chosen activity. These buttons are links to MBS' various modules. The buttons also shows how the modules are connected to each other. As shown in figure 5, the buttons are connected by arrowed lines. The purpose of this overview was to help users conceptualise their mental models of the system.

In addition, the cards consisted of various information boxes. The purpose of these boxes was to summarize specific data in order to give the user an overview over their organisation's current status. For example, instead of manually printing out a report showing the financial status of the organisation, a user could take a glance at one of the information boxes that provided this data. Such information was previously only available deep in the system. This information is now available on the system's desktop.

An interesting possibility would be to perform an additional usability test with the expert users and version 10 of the software. Such a test would provide a benchmark and make it possible to measure if the new designs made existing users perform more effectively, or if it created more obstacles. While this would be an interesting approach, this study was limited by its resources. There was not enough time to extend the study.

However, the purpose was to find out which user group provides the most effective results after redesigns, and this was accomplished with MBS version 11. Testing the system's redesigned GUI was sufficient in order to answer the research question.

4.4 - Data analysis

The collected data required analysis in order to identify the usability problems. There exist various methods for analysing data from usability tests; from simple descriptive statistical analysis methods to inferential statistical analysis. The choices depend on the overall goals and needs. The following chapter describes how the analysis for this project was conducted.

The analysis was conducted in two parts. The first part studied similar usability problems between the groups in order to find out if novices were able to uncover the same type of problems as experts. This required an analysis of the task completion, confidence intervals, task time results, questionnaires and problem list. The main objective of the second part of the analysis required an analysis of the differences between both groups to see if novices

revealed a larger number of usability problems than experts. This part of the analysis studied the task time results for differences and looked though the results from the questionnaires.

4.4.1 - Analysing type of usability problems between novices and experts

Dumas and Redish (1999) describes how the analysis of a usability test should be performed. This part of the process requires a thorough analysis of several data sources through triangulation. Five different data sources were used in this usability test before they were triangulated.

The first step was an analysis of the effectiveness. This was conducted through a simple descriptive analysis summarising how many participants that failed the tasks. The purpose was to find out which tasks the participants had most problems solving. The analysis was performed by counting how many participants that did not solve the same tasks.

There is however a problem with only presenting how many participants that did not solve a task. It would be difficult to infer how the rest of the population would perform if the results showed that 6 out of 12 participants did not solve a task. Do 6 out of 12 mean that there is a chance that 50 % of the population will be unable to solve the task? Or do those six individuals only represent six users of the total population?

The second step therefore analysed the confidence intervals in order to provide an additional layer of information. The confidence interval gives a percentage of the population who are unlikely to solve the tasks in a similar test (Sauro and Lewis, 2005).

The confidence interval was calculated for a selection of tasks. The selection was based on tasks where several participants failed. As recommended by Sauro and Lewis (2005), the Adjusted Wald method was used for task completion results containing both failures and passes, while the Laplace method was used for task completion results where either all users failed a task.

The third step analysed the efficiency. This step also used descriptive statistics. The analysis was performed calculating the mean task time for each group and each task. The criteria that were set for the efficiency analysis was that if the participants spent less than 180 seconds on a task, then the performance was good. If they spent more than 180 seconds, then the task required too much time. The analysis therefore also counted how

many participants who spent less than 180 seconds, and how many participants who spent more than 180 seconds for each task.

The fourth step analysed the participants' subjective experience of the system. The results from the questionnaire were tabulated in a spreadsheet. The purpose of the analysis was to uncover which tasks the participants experienced as difficult to solve. In order to do so, the average answers were calculated, and the tasks that were taken into consideration were the tasks that where rated as either difficult or very difficult to complete. The analysis counted how many participants for each task that rated them as difficult to solve.

The usability test provided a long list of problems, many of them which were recurring problems. The analysis of the observed problems required two steps. First, the problems were checked for recurrence. This step was necessary to identify the unique problems. Afterwards, they were categorised into which parts of the design they were discovered in.

After summarising the results from the participants, they were triangulated in order to compare the novice- and expert users' usability problems. The comparison made it possible to find out if novices encountered the same type of usability problems as experts.

There are several things that could affect the quality of the analysis. First of all, there were limitations regarding the statistical analysis of the results. Dumas (1999) argues that the use of inferential statistics, which involves analysing the probability of a usability problem occurring in the rest of the population, depends on the evaluator's skills with statistics. The author of this study had limited experience with statistics which could possibly affect the results.

While previous experience might have had an impact on the quality of the analysis, all the methods that were used in this study were from the literature. The literature has described how usability practitioners have faced problems finding the appropriate method for analysing inferential statistics. Sauro (2005) recently published an article with a solution to this problem, by introducing the use of confidence intervals to analyse the results from a usability problem.

Even though there are no standard methods for analysing data from usability tests, the author believes that the analysis in this study is sufficient. By triangulating different data

sources, a usability problem is less likely to be simply an observation because they are supported by various supporting evidence.

4.4.2 - Analysis of the amount of usability problems novices

In order to find answers to the second aim, the usability problems had to be analyzed to find out if novice users encountered a larger number of usability problems than experts. During the planning of the analysis for this research question, it was initially going to be performed similar to the prior analysis. However, while some of the data provided some important information, other results did not. This part therefore only analysed data that indicated problems.

The first step involved reanalysing the task time results. The trends showed which tasks the novices were spending more time on than experts, and provided a clue to novice usability problems. However, there were possible weaknesses by simply comparing the mean task times for each group. The groups included up to12 participants each. With as few as 12 participants in each group, the task time of one participant might greatly affect the mean. It was therefore necessary to analyse the differences in the results in a different way.

A two-sample Student's t-test assuming unequal variance was performed on the mean task times for a selection of tasks. The mean task times were tested for statistical significance to provide another layer of information about the differences between the groups. The difference is considered significant if the p-value is lower than 0.05. The test was performed using a statistical software package called SPSS.

Secondly, the results from the participants' subjective experience were compared in order to find out if the novices found certain tasks more difficult than the experts. The comparison was performed by counting the amount of participants from each group who rated tasks as either difficult or very difficult.

The last step involved going through the list of observed problems in order to find an explanation to the findings in the task time results, t-tests and questionnaires.

The task completion results were initially analysed for this part of the analysis, but did not provide any results. However, while the task completion results did not provide any clue, the remaining data strengthened the possible usability problems related to the second research question.

4.5 - Ethical considerations

The usability tests were videotaped in order to allow further analysis at a later time. However, research that is recorded on video is only allowed with the participants consent.

All the participants were therefore informed about the video recordings. They were informed through the invitations sent out by email, and prior to each test. They were also informed that the recordings would be anonymous, and that they would only be used to improve the systems usability.

5 - Results and analysis

The following chapter presents the results from the usability tests. Chapter 5.1 presents the task completion results. Chapter 5.2 describes the confidence intervals of a selection of the tasks. Chapter 5.3 provides the results of the task times. The results show the average time for each task and each group. Chapter 5.4 summarizes the results from the questionnaires. Chapter 5.5 list some of the problems that were observed during the tests. The problems are presented in a problem list. Chapter 5.6 shows the results from a t-test of the task time results. Lastly, chapter 5.7 presents the usability problems that were discovered after triangulating the results in chapter 5.7.1 to 5.7.5.

5.1 - Task completion

The following chapter presents the task completion results. The participants were measured by determining if they passed or failed each task. The summary of the results shows how many participants in each group that passed or failed each task.

The results show that at least 4 participants in the expert group failed to complete tasks 2, 10 and 12 (Figure 6). Task 2 involved editing an information box, task 10 involved understanding how information boxes are updated, and task 12 involved editing the button My Functions (see appendix A).

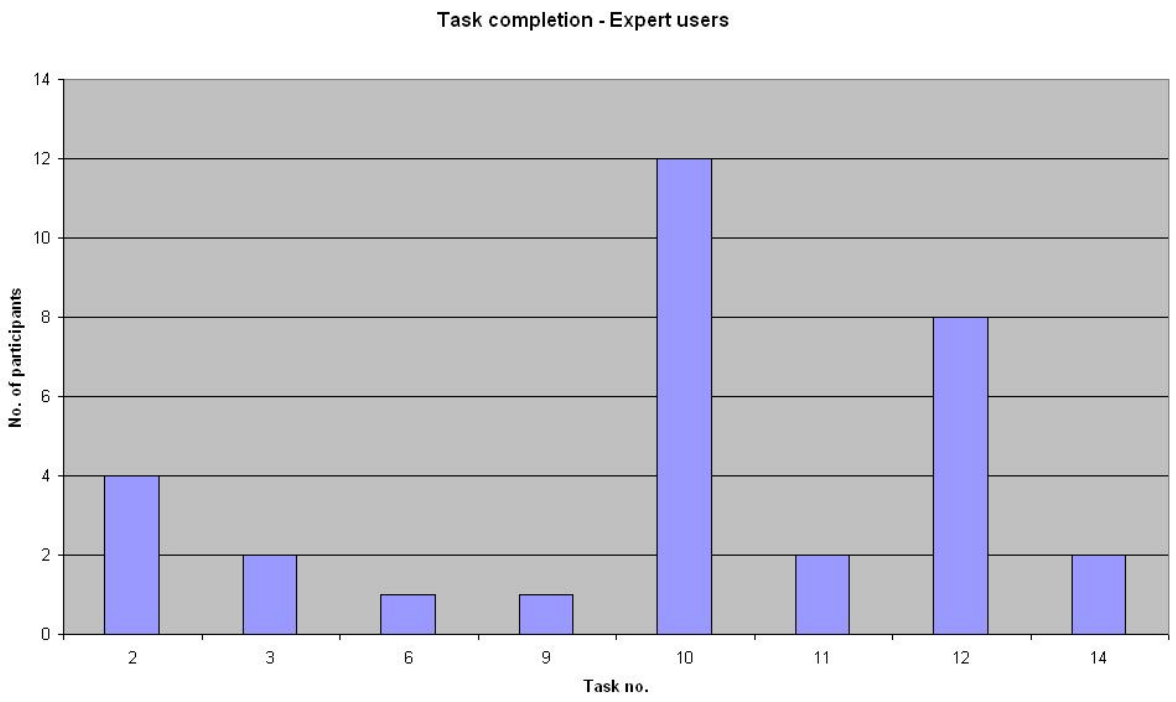


Figure 6. Experts' task completion results

The results from the novice group show that at least 4 participants failed tasks 3, 9, 10, 11, and 12 (Figure 7).

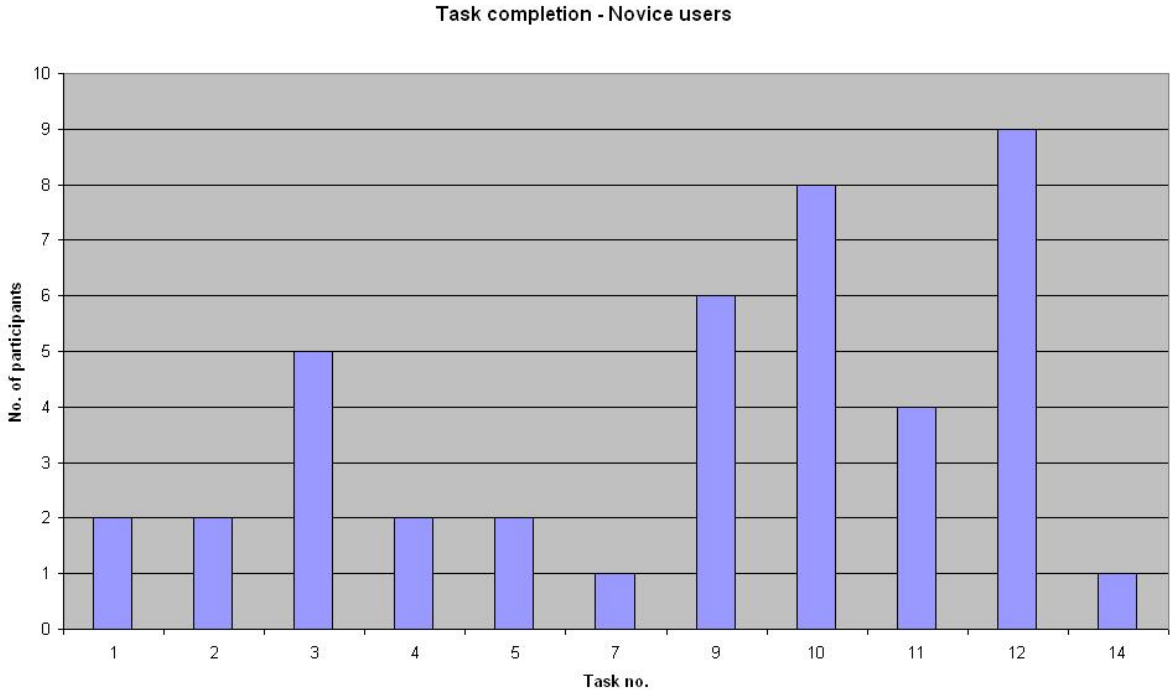


Figure 7. Novices' task completion results

A comparison between the groups' results shows that 4 participants in the expert group were unable to find out how to change the content of an information box in task 2. 1 participant in the novice group failed the same task. 1 participant in the expert group and 5 participants in the novice group failed task 9. In task 10, 12 out of 12 experts failed to solve the task, while 7 novices failed the same task. 2 experts and 4 novices failed task 11. Lastly, 8 participants in the expert group failed task 12, and 9 participants failed the same task.

5.2 - Confidence intervals

The following chapter presents the results from the confidence intervals. A calculation of the confidence intervals for tasks 10, 11 and 12 show the percentage of the user groups' population that are likely to fail the same tasks in a similar test (Table 1). The confidence intervals are calculated by using the Adjusted Wald method and the Laplace method.

Task no.	User group	Low	High	Margin of error	Confidence level
10	Novice	0.4289	0.9080	0.2395	95 %
11	Novice	0.1498	0.6481	0.2491	95 %
11	Expert	0.0350	0.4600	0.2125	95 %
12	Novice	0.5115	0.9601	0.2243	95 %
12	Expert	0.3880	0.8645	0.2382	95 %

Table 1. Confidence intervals for tasks 10, 11 and 12

42.89% to 90.90% of the novices, and 92.9% of the expert users are likely to fail task 10 (p = 0.95). Furthermore, the study is 95% confident that 14.98% to 64.81 of the novice users' population, and 3.50% to 46% of the expert users' population are likely to fail task 11 (p = 0.95). Lastly, the results show that 51.15% to 96.01% of the novices and 38.80% to 86.45% of the experts are likely to fail task 12 (p = 0.95).

5.3 - Task time results

The following chapter describes the average time each groups spent on each task. The time was measured from when the tasks were handed out and until the participants completed the task or gave up.

The results show that the novice group spent an average of at least three minutes on task 3-10 and 12-13 (Figure 8). Similarly, the expert group spent an average of at least three minutes on tasks 2, 3, 10 and 12 (Figure 9).

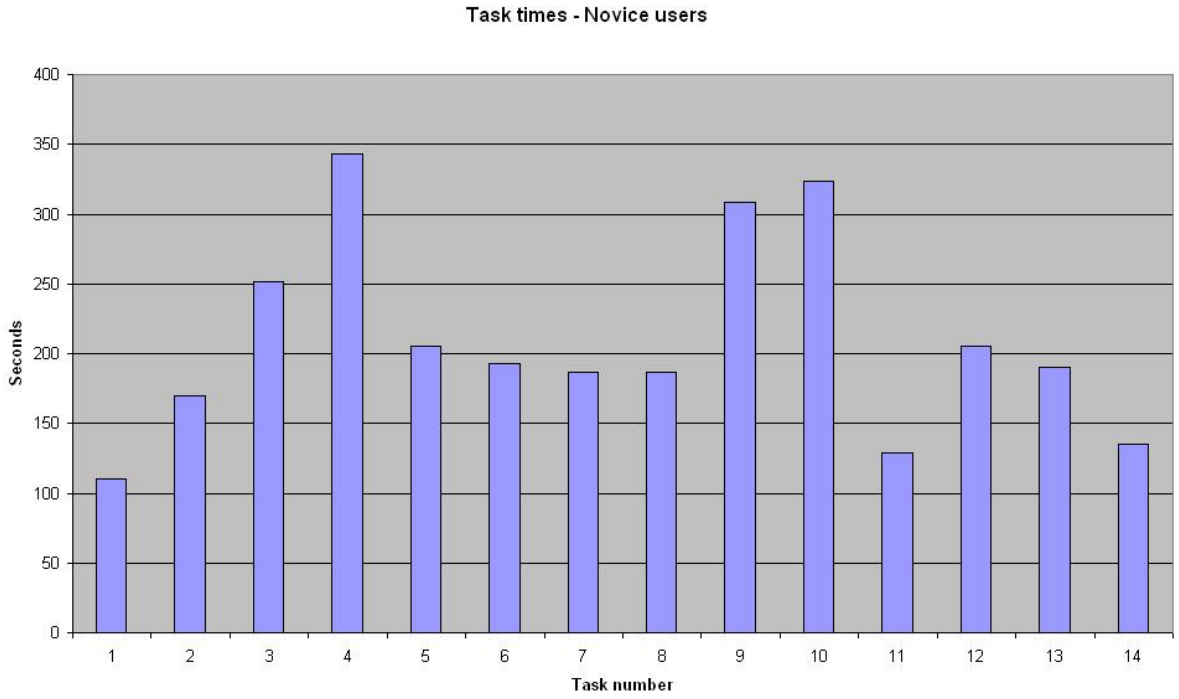


Figure 8. Novices' task time results

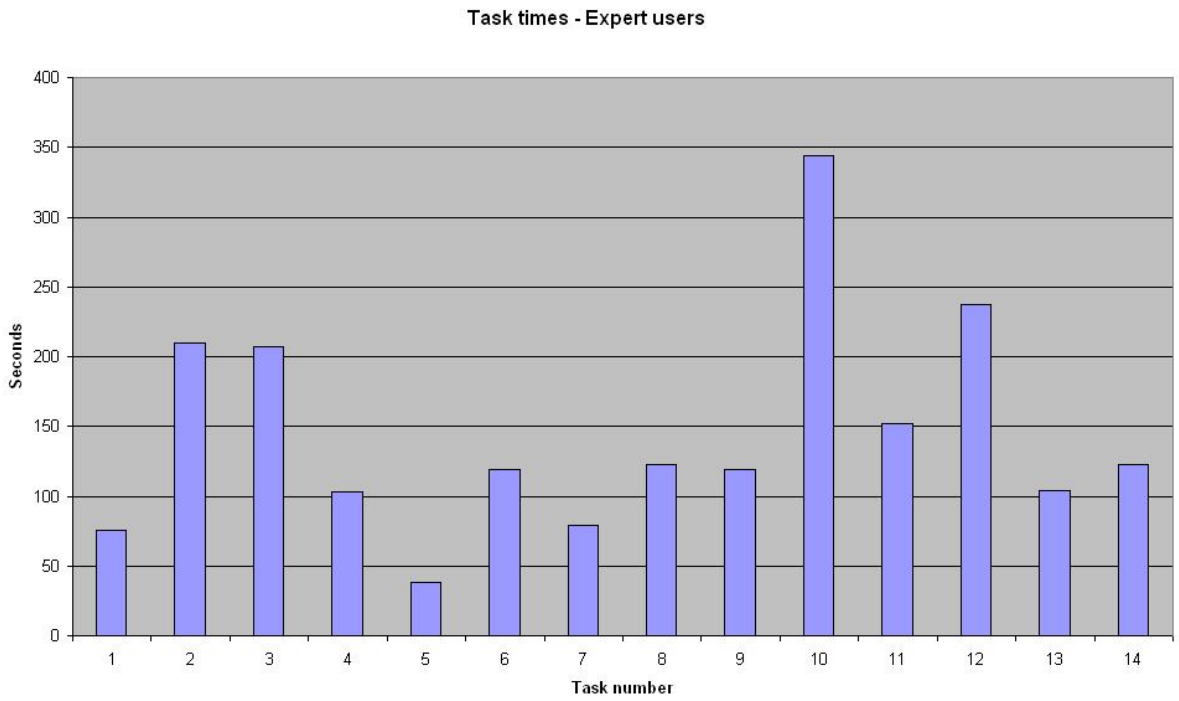


Figure 9. Experts' task time results

5.4 - Questionnaires

The following chapter gives the results obtained from the questionnaires. The results show which tasks the participants rated as difficult.

The results show that the average novice users found tasks 2-5 and 9-12 as difficult to complete (Figure 10). There were 6 participants rated task 2 as difficult, 4 participants on task 3, 7 participants on task 4, 4 participants on task 5, 6 participants on task 9, 6 on task 10, 4 participants on task 11, and 10 participants on task 12.

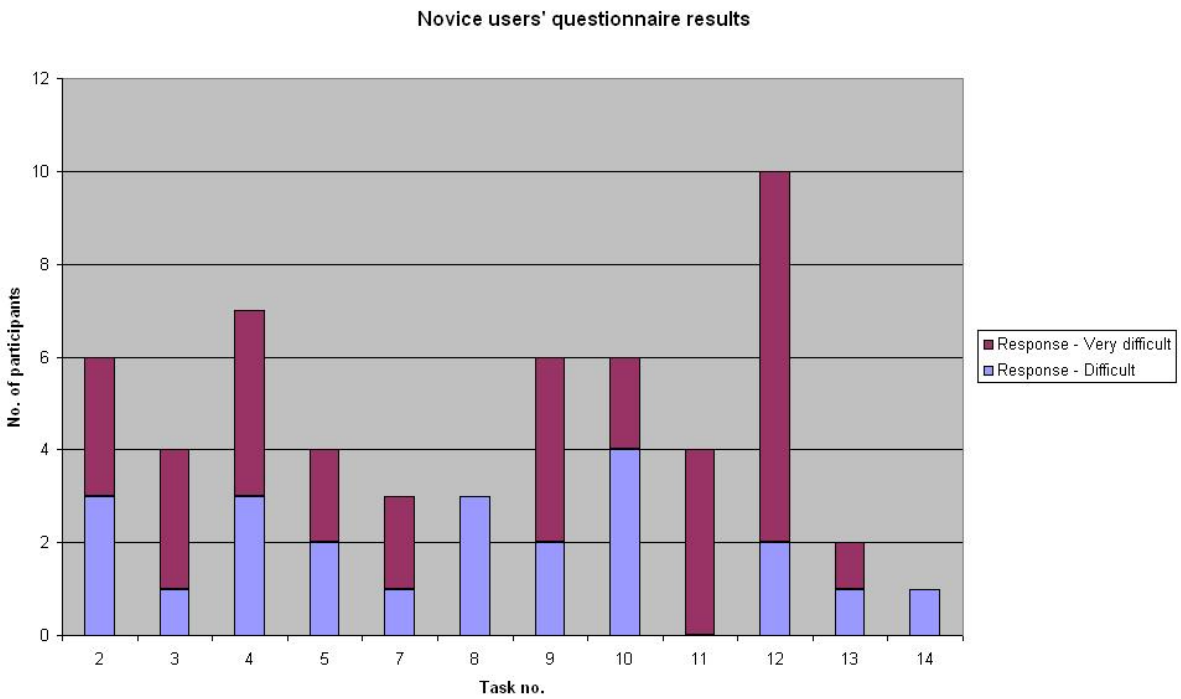


Figure 10. Number of novice users who rated the tasks as either difficult or very difficult

The results from the expert group found that they experienced task 2, 3, 10 and 12 as difficult to complete (Figure 11). 5 participants rated task 2 as difficult, 4 participants on task 3, 9 participants on task 10, and 10 participants on task 12.

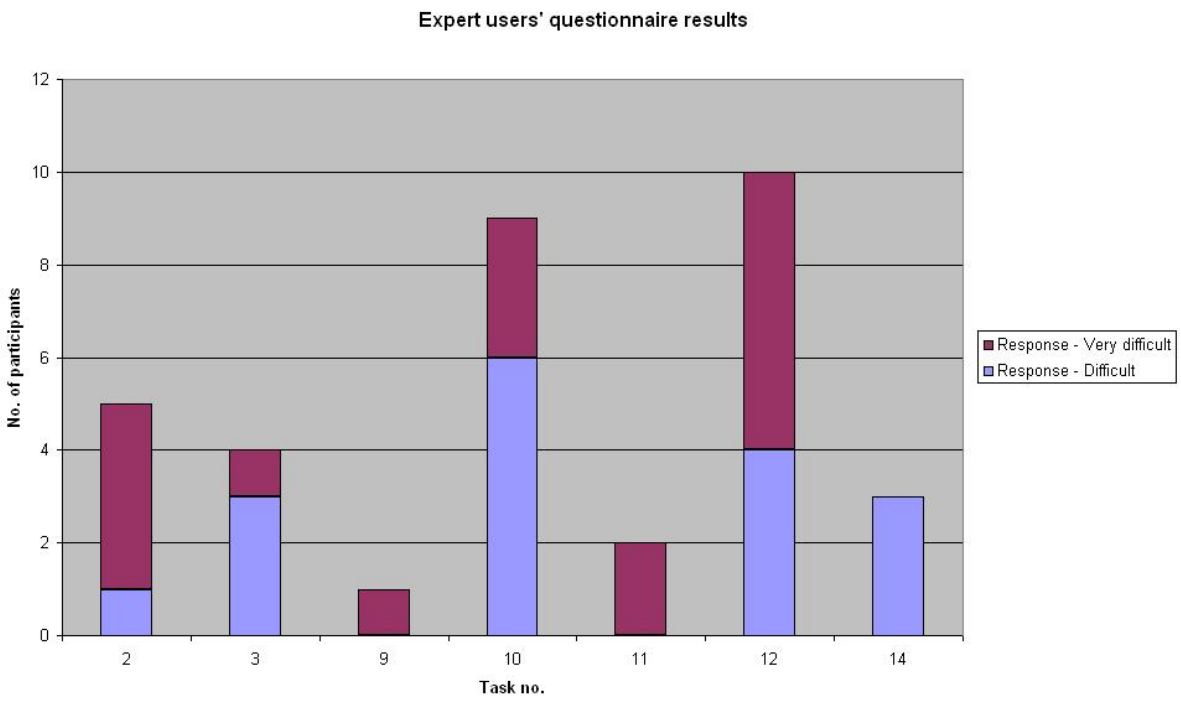


Figure 11. Number of expert users who rated the tasks as either difficult or very difficult

5.5 - Observed problems

Our usability tests identified 116 problems. After closer inspection and classification, we discovered 29 distinct problems. The following list presents the unique problems that were observed at least 5 times:

1. The user did not understand that information boxes required updating
2. The user has problems manipulating My Functions
3. The user does not understand the relationship between Contact and Vendor
4. The user experiences problems scrolling down the Contact-list
5. The user does not understand that the shortcut keys can be right-clicked
6. The user does not understand how to change the content of the information box
7. The user is experiencing that tabbing between field in the Contact-module is not natural

5.6 - T-test of the task time results

The following chapter describes the results from the t-test analysis of the participants' task time results in chapter 5.3.

The task time results shows differences between the participants' average task times in tasks 4, 5, 7, 9 and 13. A t-test was performed in order to reveal any statistically significant differences between the groups.

In task 4, the novices mean task time (M = 343.09, SD = 152.36, N = 11) was significantly different from the experts' mean task time (M = 103.25, SD = 57.06, N = 12), t(13)=-239.80, p = 0 (Table 2 and Table 3).

In task 5, the novices mean task time (M = 205, SD = 191.67, N = 11) was significantly different from the experts' mean task time (M = 38.42, SD = 24.69, N = 12), t(10)=-166.58, p = 0.01 (Table 4 and Table 5).

In task 7, the novices mean task time (M = 186.36, SD = 129.93, N = 11) was significantly different from the experts' mean task time (M = 79.42, SD = 59.60, N = 12), t(14)=-106.95, p = 0.03 (Table 6 and Table 7).

In task 9, the novices mean task time (M = 308.64, SD = 161.78, N = 11) was significantly different from the experts' mean task time (M = 118.83, SD = 78.17, N = 12), t(14)=-189.80, p = 0 (Table 8 and Table 9).

In task 13, the novices mean task time (M = 190, SD = 87.20, N = 11) was significantly different from the experts' mean task time (M = 104.42, SD = 49.28, N = 12), t(16)=-85.60, p = 0.01 (Table 10 and Table 11).

Group Statistics

	Experience	N	Mean	Std. Deviation	Std. Error Mean
Time	1	12	103,25	57,06	16,472
	0	11	343,09	152,362	45,939

Table 2. Group statistics for task 4

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Upper	Lower
Time	Equal variances assumed	12,154	0,002	-5,087	21	0	-239,841	47,152	-337,899	-141,783
	Equal variances not assumed			-4,914	12,548	0	-239,841	48,803	-345,66	-134,022

Table 3. Student's t-test of task 4

Group Statistics

	Experience	N	Mean	Std. Deviation	Std. Error Mean
Time	1	12	38,42	24,689	7,127
	0	11	205	191,67	57,791

Table 4. Group statistics for task 5

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Upper	Lower
Time	Equal variances assumed	25,621	0	-2,99	21	0,007	-166,583	55,712	-282,443	-50,724
	Equal variances not assumed			-2,861	10,304	0,016	-166,583	58,228	-295,807	-37,36

Table 5. Student's t-test of task 5

Group Statistics

	Experience	N	Mean	Std. Deviation	Std. Error Mean
Time	1	12	79,42	59,609	17,207
	0	11	186,36	129,933	39,176

Table 6. Group statistics for task 7

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Upper	Lower
Time	Equal variances assumed	7,081	0,015	-2,575	21	0,018	-106,947	41,534	-193,322	-20,572
	Equal variances not assumed			-2,499	13,765	0,026	-106,947	42,789	-198,867	-15,027

Table 7. Student's t-test of task 7

Group Statistics

	Experience	N	Mean	Std. Deviation	Std. Error Mean
Time	1	12	118,83	78,169	22,565
	0	11	308,64	161,784	48,78

Table 8. Group statistics for task 9

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Upper	Lower
Time	Equal variances assumed	2,979	0,099	-3,633	21	0,002	-189,803	52,244	-298,45	-81,156
	Equal variances not assumed			-3,531	14,149	0,003	-189,803	53,746	-304,964	-74,643

Table 9. Student's t-test of task 9

Group Statistics

	Experience	N	Mean	Std. Deviation	Std. Error Mean
Time	1	12	104,42	49,276	14,225
	0	11	190	87,17	26,283

Table 10. Group statistics for group 13

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Upper	Lower
Time	Equal variances assumed	5,193	0,033	-2,932	21	0,008	-85,583	29,191	-146,288	-24,878
	Equal variances not assumed			-2,864	15,507	0,012	-85,583	29,885	-149,101	-22,065

Table 11. Student's t-test of task 13

5.7 - Usability problems

The following chapter describes the usability problems that were uncovered during the test. The usability problems are the result of a triangulation of the data from chapter 5.1 to 5.5.

5.7.1 - Updating information boxes

Users will have problems understanding that information boxes needs to be updated manually in order to show the latest information during their first attempt (Figure 12).

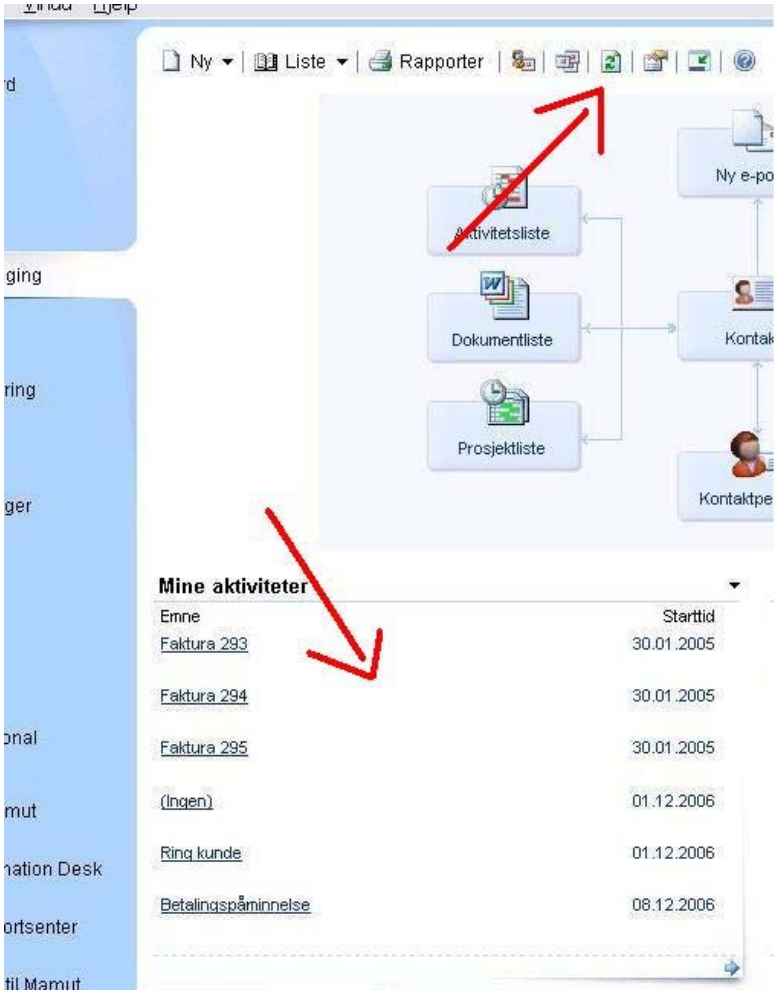


Figure 12. The information box on the bottom requires that the button Update on the top is clicked in order to show the latest information

- 12 out of 12 expert users failed the task
- 8 out of 11 novice users failed the task
- 92,9 % of the expert users are likely to fail the task (p = 0.95)

- 42 % of the novice users are likely to fail the task (p = 0.95)
- Both groups spent an average of more than 3 minutes on the task
- 9 out of 12 expert users rated the task as difficult to solve
- 6 out of 11 novice users rated the task as difficult to solve

5.7.2 - Editing My Functions

Users will have problems editing My Functions during their first attempt (Figure 13).

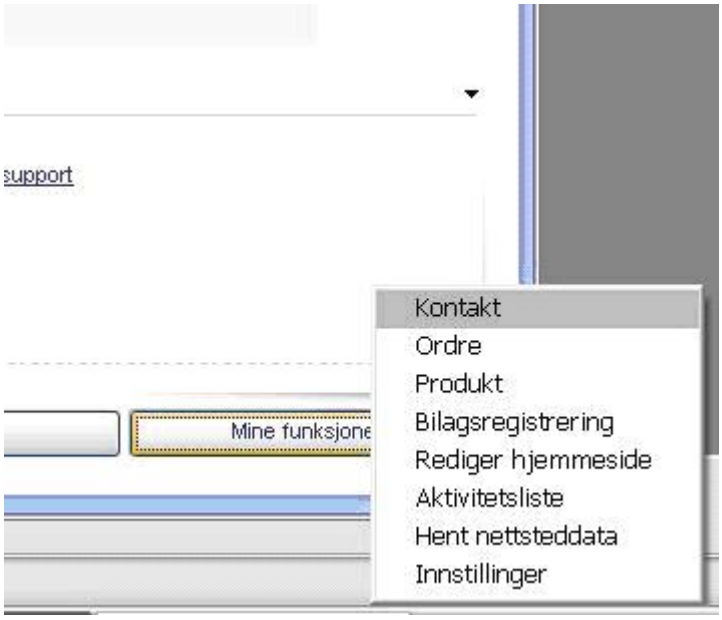


Figure 13. Several participants did not understand how to edit the list in My Functions

- 8 out of 12 expert users failed the task
- 9 out of 11 novice users failed the task
- At least 39 % of the expert users' population are likely to fail the same task (p = 0.95)
- At least 51 % of the novice users' population are likely to fail the same task (p = 0.95)
- Both groups spent an average of more than 3 minutes on the task
- 10 out of 12 expert users rated the task as difficult

- 10 out of 11 novice users rated the task as difficult

5.7.3 - Editing shortcut keys

Users will have problems editing the shortcut keys during their first attempt. The shortcut keys are the rows with buttons in the bottom of the focus areas, as shown in figure 5

- 2 out of 12 expert users failed the task
- 4 out of 11 novice users failed the task
- At least 15 % of the novice users' population are likely to fail the task (p = 0.95)
- 2 out of 12 expert users rated the task as difficult to solve
- 4 out of 11 novice users rated the task as difficult to solve

5.7.4 - Changing information boxes

Users will have problems understanding how to change the information boxes during their first attempt. These boxes are similar to the information boxes in chapter 5.6.1.

- 4 out of 12 expert users failed the task
- 1 out of 11 novice users failed the task
- At least 14 % of the expert users are likely to fail the task (p = 0.95)
- Expert users spent an average of at least 3 minutes on the task
- 5 out of 12 expert users rated the task as difficult
- 6 out of 11 novice users rated the task as difficult

5.7.5 - Managing Contacts

Users will have problems managing the contacts in the system. These are tasks surrounding the contact module as described in figure 3. These were tasks such as creating new contacts, and retrieving information about contacts.

- Novice users spent an average of at least 3 minutes on tasks 4-7, 9 and 13.

- A t-test of tasks 4, 5, 7, 9 and 13 shows a significant difference between the expert and novice users' task times
- 9 out of 11 novice users rated it as difficult to create a contact (task 4)
- 4 out of 11 novice users rated it as difficult to find a contact (task 5)

6 - Discussion

The following chapter discusses the results and address the research questions. Chapter 6.1 interprets and discusses the findings that indicate that novice users reveal the same type of usability problems as expert users. Chapter 6.2 discusses the findings that indicate the novice users uncover a larger amount of usability problems than expert users.

6.1 - Novices, experts and types of usability problems

As mentioned earlier, the purpose of the first research question was to investigate the possibility of detecting the experts' usability problems through the novice users' results. The purpose was to find out if the novices would encounter the same type of usability problems as experts. It was necessary to compare five variables between the groups in order to answer the first research question. These variables are the same data that are used to define hypothetical usability problems in regular usability tests. As presented in the results in chapter 5, the data that was used for this study included the results from task completion, confidence intervals, task time, questionnaires and list of observed problems.

Some researchers believe novice and expert users reveal different problems (Prumper et al. 1991). As mentioned in the literature review in chapter 2.3.2, experts and novices do not necessarily experience the same problems. They found that expert users provided more feedback about the tested system than novice users. According to the authors, their study indicates that novice and expert users do not necessarily experience the same type of problems.

Their hypothesis is, however, based on more than just usability problems. Their research differs from other similar studies because of their broader focus of the use of novice and expert users. While the usability problems were taken into account in their research, they also included functional problems with the systems. These were software bugs that may have been easier for expert users to detect. Novices might be unable to encounter the same functional problems.

The hypothesis that novice users are unable to uncover the same type of usability problems as experts is supported by few studies. While it might be true that novices are less likely to reveal the same amount of usability problems and functional problems as experienced users, few studies indicate that novice users are unable to uncover the same type of usability problems as expert users.

Another possible hypothesis is that novices experience the same type of usability problems as experts. It would seem likely that novice- and expert users face similar usability problems when both user groups are introduced to a completely new GUI. This hypothesis is supported by the literature and the findings in this study.

This can be demonstrated by three of the usability problems in chapter 5.7. An analysis of the novices' and experts' results detected three major usability problems. As shown in that section, these problems involved updating information boxes, editing the shortcut keys, and changing My Functions. These problems were discovered after triangulating the four data sources in chapter 5.1 to 5.5, and comparing them between the groups. The remaining chapter will discuss these similarities.

6.1.1 - Task completion

First of all, the task completion results in chapter 5.1 shows that several participants in both groups were unable to solve task 10. This task involved adding an activity to an information box (see appendix A). As described in chapter 5.7.1, both novices and experts failed to understand that the information boxes needed to be updated manually. Most participants understood that the activity needed to be created through the system's activity module. This was true particularly of expert users.

Other participants tried to create a new activity directly in the information box. This was not an available function. After realising this, the participants opened the activity module in order to create a new activity. This was the first right step towards the completion of their goal.

Many of the participants managed to create the activity, but they did not manage to make it visible in the information box. It was necessary to click the button Update, as shown in figure 12, in order to complete the task. This step was not clear for several participants and was the main reason why they failed the task.

Four novices and two experts failed task 11, showing that some of the novices and experts were unable to manipulate the shortcut keys. The participants needed to right click the shortcut keys in order to edit them and solve the task, but this was not obvious for all.

The results also show that several participants from both groups failed to complete task 12. This task involved changing the button My Functions (see appendix A). Many participants

believed they could manipulate the button directly by right clicking it similar to previous tasks. The system, however, does not allow right clicking My Functions.

After realizing this, many participants tried to look through the system's settings. This was the correct step. It was necessary to navigate through several windows in the system's settings in order to find the right setting and complete the task. Some of the expert users who were familiar with the system's settings managed to complete the task. On the other hand, most of the users failed to find the setting.

6.1.2 - Confidence intervals

The task completion results can be viewed from another perspective to give more meaning. It is possible to analyse the percentage of the groups' total population that are likely to fail the same tasks by calculating the confidence intervals. Chapter 5.2 provides the percentage of the novices' and experts' total population with the same background as the participants in this study that are likely to be fail the same tasks. The results indicate that a large percentage of both groups' total population are likely to fail tasks 10 and 12.

The confidence intervals for task 10 show that a large percentage of both groups are likely to experience problems updating the information boxes on the new desktop. The results from chapter 5.1 indicate that 92.9 % of all expert users will fail the task in a similar situation ($p = 0.95$). Likewise, the results also illustrate that at least 43 % of the novice users' population are likely to fail the same task ($p = 0.95$). These results indicate that a large percentage of both user groups' population are likely to fail the task in a similar test.

The results from the confidence intervals in chapter 5.2 indicate that both novices and experts are likely to fail the task involving changing My Functions. According to the results, at least 39% of the expert users' population are likely to fail editing My Functions ($p = 0.95$). Similarly, at least 51% of the novice users' population are likely to experience the same problems ($p = 0.95$). This indicates that a nearly equal percentage of both users' population are likely to fail task 12 in a similar test.

The results from the confidence intervals from the novice user's task completion results in chapter 5.2 may seem imprecise. For example, the results from the expert users in task 12 points to a 24 % margin of error. However, in this example, the confidence intervals also communicates that at least 51% of the novice users are likely to fail the task in the same

situation. This is possible by looking at the lowest and highest points in the confidence intervals.

It may seem like there are no similarities between the groups' confidence intervals in task 10. The reason why the results differ by at least 50 % is because two different methods were used to calculate the confidence intervals. As mentioned in chapter 2.3 and 4.4, the Adjusted Wald Method is recommended for most situations where the sample size is less than 150 (Sauro, 2006a). However, in order to use the method properly, the method requires that some users pass, while others fail. This was the case for the novice users in task 10. As mentioned earlier, 7 novice users failed, while 4 passed.

However, all participants in the expert group failed to solve task 10. Because 100 % of the group failed, it was necessary to use the Laplace method. The Laplace method is recommended for results where 100 % of the participants either fail or pass (Sauro, 2006a).

The confidence intervals for task 10 can therefore not be compared simply by comparing the percentage between the groups. The results for task 10 are similar by the fact that a fairly large percentage of both user groups are likely to fail the task.

This can be illustrated by for example assuming that the total population of novice and expert users each consisted of 100000 users. 42 % of the expert users' population equals 42 000. Likewise, 92.9 % of the novices' population equals 92 900 users. Both numbers can be used as sufficient arguments for a potential usability problem.

The confidence intervals for the other results were not included because the lowest point of the confidence interval reached below 0 %. The results were therefore not useful to argue for other possible usability problems.

6.1.3 - Task time results

The task time results in chapter 5.3 illustrates that both novice- and expert users spent most time trying to find sales information (task 3), understand the problem with information boxes (task 10) and how to change My functions (task 12). Both groups spent more than three minutes or 180 seconds to solve these tasks.

When the participants were asked to add information in to an information box in task 10, few had problems figuring out how to add the information into the system. The high task

time for this task was caused by the participants numerous attempts to view the information in the information boxes. Some of the participants expressed that they thought they didn't create the activity properly when they actually did.

Both users groups also spent much time trying to understand how to edit My Functions in task 12. The approach was similar for both groups. In this task, the amount of time was caused by the participants search for the right setting. They had to browse through several windows before either finding it or giving up.

The task time clearly shows which tasks the participants were having problems with, but not all tasks were designed well. For example, task 3 shows that both groups spent more than 3 minutes on the task. However, the design of the task description was not precise enough to lead the user into a certain place, which ultimately led to that most users were trying to solve the task in different ways. The amount of time spent on the task was caused by an unspecific task scenario.

6.1.4 - Questionnaires

The results from the questionnaires show that some of the novices and experts rated similar tasks as difficult. The results in chapter 5.4 show that at least four participants from both user groups rated tasks 2, 3, 10, 11 and 12 as difficult or very difficult.

Both user groups rated task 2 as a difficult task. This task is connected to the usability problem described in chapter 5.7.3. The results show that nearly half of both user groups rated the task as difficult. This even applied for the novice users, the group where most of the participants solved the task.

Four participants from each group rated task 3 as difficult. As mentioned earlier, the task scenario for this task was imprecise, causing many participants to look for the answer in different places in the system. The purpose of the task was to lead them into a specific focus area, but the system itself has many different ways of retrieving the same information.

As expected, many participants rated task 10 as difficult to solve. This is similar to the indication found in the task completion results, confidence intervals and task time results. Several participants commented that the task was not difficult to understand, but the task

was difficult to complete. They were frustrated by the fact that they knew they were doing it the right way, but the outcome was not as expected.

Similarly, some novices and experts rated task 11 as difficult. These participants failed to understand how to solve the task within short time and perceived this as difficult.

Lastly, a majority of the participants rated task 12 as a difficult task. Several of them expressed frustration over lack of simplicity in editing the button similarly to the prior task. In task 11, some participants managed to edit the shortcut keys next to My Functions by right clicking the button and choosing the option to edit the button. However, this function was not available for My Functions. This could possibly have affected their expectations about task 12.

Some of the participants rated the tasks as easy even though they were clearly having problems with them. Several participants expressed that even though they failed the tasks, the tasks were probably easy to pass after a couple of attempts. Therefore, there is a possibility that the results are not reflecting the participant's true experience, but rather the results that the participants expected.

Another reason why the results may not be reflecting the participant's real experience of the system is that they might feel the need to perform well (Dumas and Redish, 1999). On several occasions, some of the participants asked if they were solving the tasks correctly. They forgot that there were no right or wrong way to complete the tasks. The purpose was to gather critical feedback.

Another possible explanation might be that some of the participants did not want to give negative feedback. There were some participants who were satisfied with the system. Some of them would rather provide positive feedback to express their satisfaction. Other participants took part in the test because they wanted to express their frustration with the earlier version. Some of this frustration was transferred to the questionnaires. The questionnaires gathered subjective feedback and were therefore prone to the varying experiences some of the participants had in the past. The feedback did therefore not always reflect what they were experiencing during the time of the tasks.

Lastly, the design of the questionnaires should also be taken into account. The questionnaires in this test were standardised and based on general tips from a handbook.

Sufficient time should be spent to design proper questionnaires. However, the questionnaires in this project provided with sufficient information even with the limited time and resources that was allocated to this project.

6.1.5 - Problem list

The observed problems in chapter 5.5 describe most of the problems that have been discussed so far. While the results in chapter 5.1 to 5.4 show that problems existed in the design, the observed problems in chapter 5.5 help interpret what the problems were.

Some of the problems can be connected to the other results. These were problems concerning the systems information boxes, shortcut keys and My Functions. The most observed problem was the participants' attempts to retrieve updated information from the information boxes. As mentioned earlier, most participants understood what needed to be done in order to insert the data. The problem was to get the data out again.

After observing several participants attempt to understand how to retrieve the information, it became clear that the problem could be explained in two ways. The first explanation came from the participants. When they were asked what they expected to be the output, they explained that they expected the information box to continually show information as it was updated. As soon as they put in the necessary information, they expected the information box to update automatically. They did not expect it to show any old information.

The second explanation is that they simply did not notice the update button. Figure 12 shows that the button is grouped together with 9 other buttons. It also does not have text next to the icons as the two left most buttons. This might make it difficult to understand its important role in connection with the information boxes.

Several participants were observed having trouble changing information boxes. The participants usually spent some time trying to understand how the information boxes could be changed. While some participants understood that the small arrow on the top right of the information box, as seen in figure 12, there were some participants who did not manage to understand this. A possible explanation why some participants experienced problems trying to solve the task is that the concept of information boxes was new for both user groups.

Some participants were also observed having problems editing shortcut keys that were situated to the left of My Functions. The easiest way to edit the shortcut keys is to right click them. The participants who did not manage to edit the shortcut keys explained after the test that they were not used to that the system allowed anything be directly manipulated through right clicking. This was a common comment from both user groups. Right clicking was not an available option in version 10 of the system. Novice users were also unfamiliar with the right clicking option as this was not available in their product either.

Modifying My Functions was also a problem that occurred several times during task 12. As mentioned earlier, the most common explanation to why the problem occurred was that the participants expected that My Functions could be changed in the same manner as the shortcut keys to the left. Their second attempts were usually to go through the system's settings where most participants eventually got lost.

Based on the observations, the problem was caused by the system's lack of functionality to manipulate the list in My Functions directly. In order to edit the list according to ones own needs, it is necessary to locate the specific setting. Observations showed that several participants were unable to locate this specific setting.

The observed problems might have been affected by the evaluator effect mentioned in chapter 2.3 (Hertzum and Jacobsen, 2003). One evaluator was used in the test and might have affected the list of observed problem. A study by Hertzum and Jacobsen (2003) suggests that different evaluators observe different problems. The cause of the different observations is explained by the evaluators experience with usability tests. There is therefore a possibility that this study has missed some of the problems. However, the weakness in the observations is strengthened by the findings in the other data sources. It is therefore still possible to reveal critical problems by looking at the other results.

6.1.6 - Usability problems

The usability problems in chapter 5.7 are outcomes of a triangulation of the results in 5.1 to 5.5. The following describes the usability problems that both the expert and novice users experienced. The results show that both groups encountered three similar usability problems.

The most obvious usability problem was related to task 10, which involved updating the information boxes as described in chapter 5.7.1. The results were similar for both the novice- and expert group. Several participants failed to complete the task, a significant portion of both user groups' are likely to fail the same task, a large amount of time was spent trying to solve the tasks, and several rated the task as difficult to complete.

Chapter 5.7.3 describes a problem related to task 11. Some participants failed to complete the task when they were asked to edit a button. This problem was most likely caused by their previous experience with other programs. Various applications from Microsoft Office have the same functionality. However, they are not explicit and require to be explored. Users who are unfamiliar with this type of directly manipulating an element are likely to fail the same task.

Another usability problem was the manipulation of My Functions in task 12, as described in chapter 5.7.2. Similar to the usability problems mentioned above, the results from the novices and experts show that several participants failed task 12, a large percentage of both user groups' population are likely to fail the task, a significant amount of time were spent on the tasks, and they were rated as difficult to complete.

The results in chapter 5.7.4 also describe a usability problem related to task 2, which was only uncovered by the expert users. This problem involved changing the content of an information box. The results show that even though both groups rated the task as difficult to solve, there were fewer novice users than expert users who failed the task, and that the novice users spent less time on the task than expert users.

However, a possible explanation to why the novices rarely experienced the usability problem in task 2 is that the expert users tested the system several months before the novice users. After the usability test with the expert users, a second test was run at least 3 months later. There were two important releases during these months. First of all, the tested system was released to the market. It was therefore available for evaluation for all markets.

Secondly, the novice users received a new version of their product. One of the highlights in this version was the use of similar information boxes. The novice users could therefore have been exposed to these elements prior to the tests and consequently affected the usability test results.

The comparison between the groups usability problems are based on the isolated results. As mentioned earlier, usability problems based on outliers are not included in the comparison, because they were based on intuitive guesses. They were potential usability problems with little supporting evidence in the various measurements. Such problems are detected by chance and do not depend on the users experience level.

The overall results and analysis suggests that novice users will reveal the same type of usability problems as expert users. Both novices and experts encountered problems related to tasks 10, 11 and 12. Thus, a usability test with only novice users would most likely point to the same problems as in a test with expert users.

Using novice users in a usability test on a redesigned GUI will most likely uncover the same usability problems as expert users. A possible explanation is that the GUI is new for both novices and experts. The differences lie in their past experience with the interactive system and the mental model of how the system works.

The expert users are expected to perform better than novice users (Goonetilleke et al., 2001). The experts' performance in this study is better due to their prior experience with the system. They are familiar with the underlying system architecture and are familiar with its various functions.

However, the GUI's design is usually critically flawed whenever an expert is experiencing a usability problem (Donker and Reitsma, 2004). If an expert is unable to work her way around a problem, a novice user is less likely to solve the problem unless he knows how to work around it from previous experience.

The results from this study confirm the findings similar findings from literature review (Faulkner and Wick, 2005). The authors from that study argued that the results from a usability test with both novices and experts would reveal many similar usability problems, and would also help categorise the problems. Similarly, the results from this study show that the usability problems that the expert users were experiencing were also revealed by the novice users.

As mentioned in chapter 2.3, other authors argue that heuristic reviews effectively reveal experienced users' usability problems, while usability tests effectively reveals novice users' usability problems (Fu et al., 2002). The authors claim that usability problems that

are experienced by expert users can be uncovered through heuristic reviews. This further removes the need of expert users in usability tests.

It is therefore likely that the expert user's usability problems will be detected through the use of novice users in usability tests. If a usability practitioner has to prioritise between recruiting novices and experts for a usability test after a major redesign, the practitioner will gain more by recruiting novice users.

6.2 - Novices, experts and the amount of usability problems

The following chapter discusses the results in order to answer the second research question. The second research question of this thesis asked if novices uncover a larger amount of usability problems than experts.

As mentioned earlier, some believe that novices reveal less number of usability problems than experts (Prumper et al., 1991). Chapter 5.7.3 described a usability problem that might have supported that study. This problem was only detected by the experts in task 2. However, as mentioned earlier, there is a possibility that the novice users successfully completed task 2 because of their prior experience with another product. There is therefore no sufficient evidence in this thesis that supports the previous studies suggesting that the use of novices in usability tests will detect less usability problems than experts.

Contradictory to these arguments, several findings indicate that novices provide a larger amount of usability problems than expert users. This is supported by the usability problems described in chapter 5.7. The novices in this experiment provided more problems with the redesigned GUI than the experts. One usability problem was connected to four tasks. These were tasks 4, 5, 7 and 9 (see appendix A) and were related to three fundamental processes in the ERP system.

As described in chapter 5.7.5, this problem involved the management of contacts. Even though the task completion results in chapter 5.1 shows that most participants in both groups managed to solve the tasks, the following section discusses how the remaining results show otherwise.

6.2.1 - Task time results and t-tests

The task time results in chapter 5.3 shows different performance levels in the tasks mentioned above. The trends in figure 8 and 9 indicate that novice users spent more time

than expert users on tasks 4, 5, 7 and 9. Even though the novices had the necessary domain knowledge to pass the tasks, they still experienced problems with the system. According to the observations in chapter 5.5, when they were asked to create a vendor, they assumed that they would find a function in the system that created them. However, the system requires additional knowledge about its concepts.

First, one needs to create the contact and specify whether that contact is a customer or a vendor. The novice users did not understand this concept. While the novices were able to navigate themselves correctly towards task completion, most of them faced difficulties during navigation. Consequently, the novices spent a considerable amount of time trying to understand how this worked which resulted in a higher task time than the experts.

This explanation is further strengthened by reanalyzing the task time results. The Student's t-test in chapter 5.6 shows that the task time difference between the novices and experts were statistically significant ($p < 0.05$).

The experts managed to solve the tasks within reasonable time even though they were new to the GUI. When the experts faced problems, they recovered much faster than the novices, much like what Prumper *et al.* (1993) describes in their study. This is also supported by other studies such as Dillon and Song (1997), which showed how expert users spent less time than novice users on search engines, and the study by Goonetilleke et al. (2001) on different response times on icon designs.

However, there may be several weaknesses with the task time results. The measurement could have been more accurate. For instance, some of the participants talked more than others. At times, they would comment on various issues during the tasks because they were told to think aloud. They were not 100 % focused on the tasks. The performance time could have been more accurate by asking them to solve tasks by themselves and take some time afterwards to talk about possible problems. This is, as mentioned in chapter 2.3, also known as the RTA method (Zhiwei et al., 2006).

6.2.2 - Questionnaires

The results from the questionnaires in chapter 5.4 further confirmed that novice users experienced additional usability problems compared to the expert users. While the experts

rated the tasks as easy to solve, there were a larger number of novice users who rated tasks 4, 5, 7 and 9 as difficult.

6.2.3 - Usability problems

A triangulation of the results suggested one potential usability problems that were only detected through the novices. This was a recurring problem in the tasks mentioned above. As described in 5.7.5, the results suggest that novice users experience difficulties understanding how to manage their contacts. During their first encounter with the system, novice users are expected to experience usability problems when performing contact management, unlike the expert users who are likely to face some problems, but are still able to recover from them and manage to solve related tasks.

The results suggest that novice users reveal more usability problems than expert users on redesigned GUIs. Along with the findings in chapter 6.1, which suggested that novices were able to uncover the same type of problems as experts, the findings in this chapter indicates that novices encounter additional problems.

A possible explanation might be the participants' system knowledge, as described in chapter 2.4 (Nielsen, 1993). The mental model of the system differs between the users (Donker and Reitsma, 2004). In this study, the novice- and expert users had different mental models of how contact management worked in the tested system. The expert users are more likely to be able to manage their contacts than the novice users because of their prior experience with the earlier version of the system.

Based on the literature (Fu et al., 2002) and the findings in this study, novices are likely to uncover a larger number of usability problems than experts in usability tests on redesigned GUIs. With the exception of the problem in chapter 5.7.4, the novices revealed 4 usability problems. On the other hand, the experts in this study only detected 3 of those 4 problems. This indicates that novices provide more usability problems than experts.

The findings do not suggest that usability practitioners should perform usability tests with only novice users. If the goal is to uncover problems for a certain target group, then the test should, without a doubt, perform tests with the users in the target group. However, if a usability practitioner is asked to uncover as many problems as possible, with limited

resources, and must to choose between different types of users, then the usability test will be more effective with novice users.

The suggestions in this study are appropriate for summative usability tests rather than formative tests. While the summative tests seek to confirm potential usability problems, formative tests attempts to explore previously unknown problems. This thesis detected several usability problems based on outliers. However, these were not a part of the analysis because of the lack of supporting evidence. Most of the outliers were discovered through the list of observed problems, and not through the other collected data.

Usability practitioners will benefit from recruiting novice users in usability tests because of the high probability of revealing both the novices' and experts' usability problems. This knowledge might be a useful tool for usability practitioners who are planning the recruitment process for their usability tests.

7 - Conclusion

The following chapter concludes the findings in this study. The chapter includes the study's conclusions, theoretical implications, limitations and suggestions for future research.

Current research is fine tuning the usability testing method by studying topics such as different talk aloud techniques, classification of usability problems and sample sizes. This study investigated the use of novice users in usability testing. The primary aim was to investigate if novice users uncovered the same type of usability problems as expert users.

The study was performed by running a usability test with both novice and expert users on a redesigned ERP system. The purpose was to compare the results from the novices and the experts in order to find the most effective group.

A set of 14 tasks was designed to simulate typical business processes which the participants were asked to solve. The first series of usability tests was performed with a group of expert users, while the second series of tests was performed with the novice users.

The tests produced results from multiple data sources. The first data source was the task completion rate. These results gave indications about which tasks both groups were unable to solve. The second data source was the task efficiency. These results measured how much time both groups were spending on each task. The results showed which tasks that required a considerable amount of time. Lastly, the participants' subjective experience with the system was measured in order to find the tasks that both groups were experiencing difficulties with.

The analysis of the results for the first research question was performed by analyzing the results from the task completion results, confidence intervals, task time results, questionnaires and the observed problems. The results from these analyses showed which usability problems both groups were experiencing. The results and analysis indicated that novice users revealed the same type of usability problems as the expert users.

These findings are in contrast with earlier studies (Prumper et al., 1991) which claim that novice users not necessarily reveal larger amount of problems than expert users. They argued that even though it seems reasonable that novice users encounter more problems, their experiment proved otherwise. On the other hand, the majority of recent studies claims

that novice users indeed provides additional information about several types of usability problems (Faulkner and Wick, 2005). First of all, the study shows how different types of users perform differently. Secondly, the study shows that the inclusion of both experts and novices provides an additional layer of information about the usability problems, because different users will experience different problems due to their unmatched mental model of the system.

This thesis therefore concludes that novice users reveal the same usability problems as expert users in usability tests on redesigned GUIs. The findings in this study support the hypothesis that novice users detect the same type of usability problems as expert users.

The second aim was to investigate if the use of novice users in usability testing provided more usability problems than expert users. This was done by examining which group that produced most of the usability problems.

The analysis first revealed the usability problems between the groups. The results showed that the novice users were experiencing usability problems concerning some of the concepts in the system such customers and vendors. The experts did not encounter these problems.

An analysis was performed on the participants task times and questionnaire results. The results from the comparison of the task times showed a significant difference between the groups in the tasks related to the usability problems mentioned above. The results from comparing the subjective ratings showed that the novice users were experiencing more difficulties with the same tasks. The analysis indicates that novice users detect more usability problems than expert users. Even though both groups revealed similar usability problems, the novice group revealed additional usability problems that the expert group never experienced.

A possible explanation is that the groups had different mental models of the ERP system. Even though none of the groups had used the new interface prior to the tests, the expert users had experience with the previous version. Both the new and old version had the similar architecture. On the other hand, the novice users had previous experience from other similar products, but these systems did not have the similar architecture.

The findings were consistent with previous studies in the same area (Donker and Reitsma, 2004). As mentioned earlier, several studies argue that expert users will be able to perform tasks faster than novice users. Experiments also showed that training with the system had effects on the participant's performance times (Dillon and Song, 1997). Other studies argue that novice users will reveal other types of problems (Fu et al., 2002). This study showed similar results.

This thesis therefore concludes that novice users are more likely to reveal more usability problems than expert users in a usability test on a redesigned GUI. Expert users are likely to reveal serious usability problems on new interfaces. However, novice users are likely to uncover the same usability problems, and are also provide additional usability problems that are caused by their different mental model of the system. Usability practitioners should therefore include novice users when performing usability tests on redesigned systems for effective usability test results.

The findings from this study implicate both the research field of usability evaluation methods and the community of usability practitioners. The research on usability methods is currently fine-tuning the usability test method in order to make them more effective and efficient. The findings in this study contribute with knowledge on the topic on different types of users in usability tests.

The usability practitioners' community benefits from this knowledge by providing them with an extra tool during recruitment of participants for usability tests. Many usability practitioners perform tests on limited budgets and will not be able to recruit the same amount of participants as in this study. Because of these limitations, they may have to choose between difficult options when deciding which types of users to recruit. The knowledge from this study will make such decisions easier. If the goal of their usability tests are to reveal as many usability problems as possible from a wide range of users, this study recommends recruiting novice users from the appropriate target group for a maximum effect.

7.1 - Limitations

Some of the limitations in this study include the evaluator, the participants, and the goal of the usability tests. First of all, the usability tests were evaluated by one person. The results may therefore be subject to the problems of the evaluator effects mentioned in the literature

(Hertzum and Jacobsen, 2003). The observed problems in usability tests are affected by the evaluator's experience. The literature describes several studies where different evaluators observe different problems. This study may therefore have overlooked some problems.

The evaluator in this study had limited experience with usability testing prior to the experiment. The usability tests were run for the first time by the evaluator. The same study shows that experience also have an effect on the problems that the evaluator observes. An experienced evaluator could also possibly have designed a more effective test and revealed more usability problems.

Another limitation is the possibility that the findings in this study are biased. The evaluator in this study was currently an employee by the company who developed the tested system product. However, the limited time, competence and budget made it difficult to include an external evaluator. An external evaluator might have provided more objective results.

Another limitation in this study is related to the number of participants in each group. Originally, there were 24 participants. However, a later discovery revealed that one of participants was a beta-customer and had used system for at least six months. The participant's results were therefore removed. The results that were analysed came from the remaining 23 participants. This led to an uneven sample with 12 expert- and 11 novice users. The sample could have been extended by adding another participant, but there were no more incentives to offer the participants.

A sample size of 23 participants is small compared to other similar studies. Other research projects included up to 60 participants (Faulkner and Wick, 2005). Such sample sizes are more reliable. In order to match other studies, one should include more participants than in this study. This increases the reliability of the results.

The study tried to recruit an equal number of participants of each sex. The expert group consisted of 6 males and 6 females. The novice group however consisted of one female and the rest were males. It was desirable to have an equal amount of each sex in each group. There were however few females in the novice group who signed up for the experiment and was therefore a possible weakness. The study did not represent an equal amount of participants from each sex.

Measuring the task times was not always accurate. One reason was that the study used the talk aloud method as described in chapter 2.3, which involved asking the participants to talk during the usability test. Some of the participants talked more than others during the tests and most likely increased their performance times, which consequently increased the overall average times. The average task times were therefore also vulnerable to the small samples size.

The task time's accuracy could have been improved by using the retrospective talk aloud technique (Zhiwei et al., 2006). This allows more precise measuring because the participants are allowed to solve tasks at their own pace, and allows further discussion after the task has been completed.

Despite the limitations, one of the strengths of the study was that the usability tests were conducted with a commercially available ERP system. The experiment was therefore as close as possible to a realist usability test situation. The results from this study are therefore valid for usability practitioners performing similar evaluations on commercially available software products.

Other strengths with this study were that the participants were actual customers. The findings are based on results from actual users of ERP systems. These users increased the validity of the study.

7.2 - Recommendations for future work

There are several suggestions for future work in this field. A similar experiment should be conducted with a larger amount of participants and another type of system in order to confirm the finding's reliability. Additional research on the same topic might provide interesting results.

There is an increasing demand for usability testing on web sites. Instead of conducting another study on information systems for local machines, a possible experiment could be to run usability tests on a web-based system. The research should include at least 70 participants and an equal amount of males and females. Furthermore, the tests should be evaluated by at least two experienced evaluators. This will increase the accuracy of the results. In addition, the retrospective think aloud technique should be used. As mentioned

earlier, this technique increases the accuracy of the results. Data such as task times will be closer to reality. Such an experiment would take this study up to another level.

Another suggestion for future work is to continue developing current design principles. Current studies show that these principles help evaluators reveal experts' usability problems (Fu et al., 2002).An interesting direction could be to include existing knowledge on novices' mental models into new design principles. Such design principles would benefit GUI designers and reduce the need for usability testing.

8 - References

AGRESTI, A. & COULL, B. A. (1998) Approximate is better than "Exact" for interval estimation of binominal proportions. *The American Statistician,* 52**,** 119-126.

DILLON, A. & SONG, M. (1997) An empirical comparison of the usability for novice and expert searchers of a textual and a graphic interface to an art-resource database. *Journal of digital information,* 1.

DONKER, A. & REITSMA, P. (2004) Usability testing with young children. *Proceeding of the 2004 conference on Interaction design and children: building a community***,** 43-48.

DUMAS, J. S. & REDISH, J. C. (1999) *A practical guide to usability testing,* Exeter, Great Britain, Intellect Books.

FAULKNER, L. & WICK, D. (2005) Cross-user analysis: Benefits of skill level comparison in usability testing. *Interacting with Computers,* 17**,** 773-786.

FU, L., SALVENDY, G. & TURLEY, L. (2002) Effectiveness of user testing and heuristic evaluation as a function of performance classification. *Behaviour & information technology,* 21**,** 137-143.

GOONETILLEKE, R. S., SHIH, H. M., ON, H. K. & FRITSCH, J. (2001) Effects of training and representational characteristics in icon design. *International Journal of Human-Computer Studies,* 55**,** 741-760.

GULLIKSEN, J., BOIVIE, I. & GORANSSON, B. (2006) Usability professionals--current practices and future development. *Interacting with Computers,* 18**,** 568-600.

HERTZUM, M. & JACOBSEN, N. (2003) The evaluator effect: a chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction,* 15.

KUNIAVSKY, M. (2003) *Observing the user experience: a practitioner's guide to user research,* San Francisco, Morgan Kaufmann Publishers.

LEWIS, J. R. (2006) Sample sizes for usability tests: mostly math, not magic. *interactions,* 13**,** 29-33.

NIELSEN, J. (1993) *Usability engineering,* California, Morgan Kaufmann.

NIELSEN, J. & MACK, R. L. (1994) *Usability inspection methods,* New York, John Wiley & Sons Inc. .

NORMAN, D. A. (2002) *The design of everyday things,* New York, Basic Books Inc.

PREECE, J., ROGERS, Y. & SHARP, H. (2002) *Interaction design : beyond human-computer interaction,* New York, John Wiley & Sons Inc.

PRUMPER, J., FRESE, M., ZAPF, D. & BRODBECK, F. C. (1991) Errors in computerized office work: differences between novice and expert users. *SIGCHI Bulletin,* 23**,** 63-66.

RUBIN, J. (1994) *Handbook of usability testing : how to plan, design, and conduct effective tests,* New York, Wiley.

SAURO, J. (2006a) Confidence Interval Around a Completion Rate. Retrieved April 17, 2007, from http://www.measuringusability.com/wald.htm

SAURO, J. (2006b) The user is in the numbers. *interactions,* 13**,** 22-25.

SAURO, J. & LEWIS, J. R. (2005) Estimating completion rates from small samples using binominal confidence intervals: comparisons and recommendations. *Proceedings of the Human Factors and Ergonomics society 49th annual meeting***,** 2100-2104.

ZHIWEI, G., SHIRLEY, L., ELISABETH, C. & JUDITH, R. (2006) The validity of the stimulated retrospective think-aloud method as measured by eye tracking. *Proceedings of the SIGCHI conference on Human Factors in computing systems***,** 1253-1262.

Appendix A - Task scenarios

Usability test av MBS11 - Oppgavesett

Dato:

Oppgave 1

Du ønsker å se hva dine planlagte aktiviteter er for i dag. I tillegg ønsker du å finne ut hvem som skal delta i møtet kl. 10 00.

Hvor finner du aktivitetene dine?

Du har et møte kl. 10 00. Hvordan finner du mer informasjon om hvem som skal delta i møtet?

Oppgave 2

Du har en rute på skrivebordet som heter Kontantbeholdning. Du ønsker å endre denne ruten til Mine ubehandlede innkjøp.

Hvordan endrer du innholdet?

Oppgave 3

Du er daglig leder i en bedrift og ønsker å få en oversikt over hvordan salget går.

Hvordan får du opp en oversikt over salget i bedriften?

Hvor mange fakturaer har forfalt til betaling?

Oppgave 4

Du har fått en ny leverandør av kontorrekvisita. Du har ikke handlet med denne leverandøren før så de må registreres. Leverandøren oppgir følgende informasjon:

- Navn: Nordmann Ark
- Adresse: Kirkeveien 20, 0110 Oslo
- E-mail: sara@nordmann.no
- Telefonnummer: 2201 0510

Hvordan registrerer du leverandøren?

Oppgave 5

Du har fått en henvendelse fra en kunde som heter Pedersen Consulting. Du ønsker mer informasjon om denne kunden.

Hvordan finner du frem kundens adresse?

Hva er kundens adresse?

Oppgave 6

Du har avtalt med Pedersen Consulting om at du skal ringe dem opp rundt september. Sett av en halvtime til samtalen.

Hvordan oppretter du en ny aktivitet om at du skal ringe Pedersen Consulting den 1.9.2006, kl. 11 00?

Oppgave 7

En kunde fra Bygg AS ringer inn og mener fakturaen han har mottatt ikke stemmer. Fakturaen ble sendt 25.04.06. Du ønsker å lete opp denne fakturaen for å se hvilke og hvor mange produktlinjer som har blitt satt inn.

Hvordan henter du opp denne ordren for å få sjekket produktlinjene på nytt?

Oppgave 8

Du har et prosjekt som heter Project X. Du ønsker å se hvordan prosjekt funksjonene henger sammen.

- Gå inn på Prosjekt i fokusmenyen til venstre
- Forklar hvordan du forstår oversikten du har fått opp
- Hvordan går du videre til innkjøpsmodulen?

Oppgave 9

Du ønsker å finne et dokument som du tidligere har opprettet for en kontaktperson hos Etcetera AS.

1. Hvordan finner du kontaktpersonen hos Etcetera AS?

2. Gå tilbake til Kundeoppfølgings-området.

3. Hvordan oppretter du en ny ordre på Etcetera AS

Oppgave 10

Du har en rute på skjermen som heter Mine Aktiviteter og du ønsker å legge til en aktivitet til denne ruten.

Emnet er: ”Møte med potensiell kunde”
Starttid er 28.09.06 kl. 14 30

Hvordan legger du til denne aktiviteten under Mine aktiviteter?

Oppgave 11

Du ønsker å endre en knapp på den nederste knapperaden.

Hvordan ville du endret knappen **Kontaktutvalg** til **Lagerbeholdning**?

Oppgave 12

Du har en rad med knapper nederst i vinduet. Helt til høyre har du en knapp som heter Mine Funksjoner. Ved å trykke på knappen får du opp en liste. Du ønsker å legge til Lønnsregistrering til denne listen.

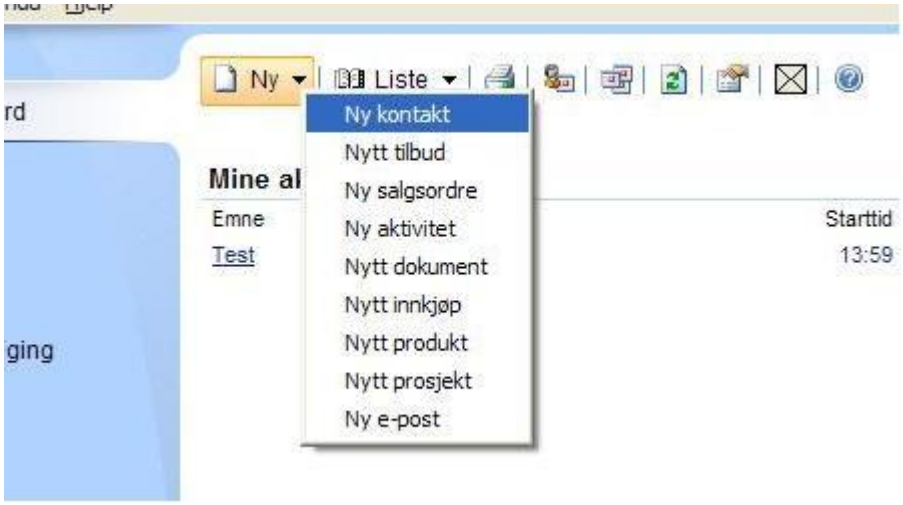
Hvordan legger du til Lønnsregistrering til Mine Funksjoner-knappen?

Tilleggsopplysninger

Følgende er en beskrivelse av noen de nye funksjonene i MBS11. Det er blant annet Ny- og Liste-knappen på verktøylinjen, og flytdiagrammet

Verktøylinjen – Ny-knappen

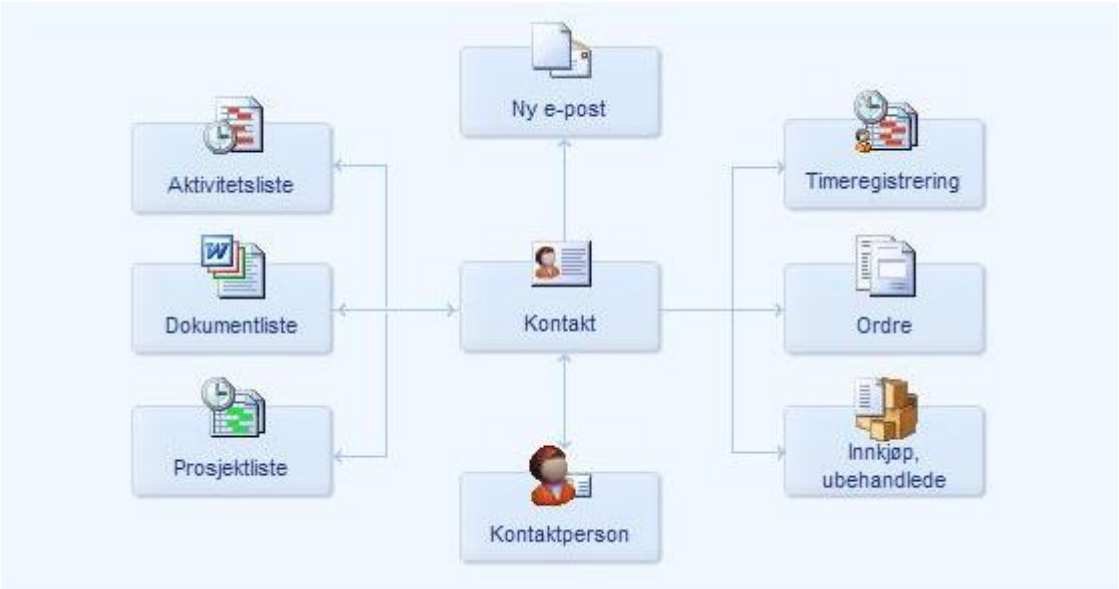
Øverst på skrivebordet har ligger det en verktøylinje med en Ny-knapp. Fra denne knappen er det mulig å opprette for eksempel nye kontakter, nye ordre osv.



Fokusvalg – Kundeoppfølging

Under fokusmenyvalget Kundeoppfølging ligger det et flytdiagram som viser hvordan ulike funksjoner henger sammen, og hvordan informasjonen flyter mellom dem. Her kan man for eksempel se hvordan Aktiviteter kan tilknyttes kontakter og hvordan en ordre kan baseres på en kontakt.

Knappene kan klikkes på for å gå rett inn i funksjonene.


Ny e-post
Aktivitetsliste
Timeregistrering
Dokumentliste
Kontakt
Ordre
Prosjektliste
Innkjøp, ubehandlede
Kontaktperson

Oppgave 13

Du ønsker å registrere en ny leverandør av datautstyr. Leverandøren oppgir følgende informasjon:

- Navn: Complete Data
- Adresse: Pilestredet 45, 0203 Oslo
- E-mail: daglig@complete.no
- Telefonrnummer: 2230 7849

Hvordan registrerer du leverandøren?

Oppgave 14

En av kundene dine har ikke mottatt invitasjon til et møte. Du ønsker å finne ut hvem din faste kontaktperson er hos Tog Ekspress Oslo og vil skrive ut invitasjonen på nytt.

Hvordan finner du kontaktpersonen hos Tog Ekspress Oslo?

Hvordan finner du invitasjon til denne kontaktpersonen?

Appendix B - Questionnaires

Pretest spørreskjema

Deltagernr: 17
Dato: 29.11.06

1. Kjønn (kryss av):
   __ Kvinne
   __ Mann

2. Alder (kryss av):
   __ 20-24 år
   __ 25-34 år
   __ 35-49 år
   __ 50-60 år

3. Hva arbeider du med? (Kryss av for alt som gjelder)

   ___ Salg
   ___ Logistikk
   ___ E-handel
   ___ Annet: _____
   ___ Prosjektledelse
   ___ Regnskapsføring
   ___ Markedsføring

4. Hva slags type stilling arbeider du i?

   ___ Heltidsstilling
   ___ Deltidsstilling

5. Hvilke av de følgende produktene bruker du i forbindelse med arbeidet ditt?

   __ Datax
   __ Mamut Business Software

6. Hvis du krysset av for Datax i spørsmål 5, vennligst svar på følgende spørsmål:

   Hvor lenge har du arbeidet med Datax?

   ___ Mindre enn 6 måneder
   ___ 6 måneder til 2 år
   ___ Mer enn 2 år

7. Hvis du krysset av for Mamut Business Software i spørsmål 5, vennligst svar på følgende spørsmål:

7 a). Hvor lenge har du brukt Mamut Business?

___ Mindre enn 6 måneder
___ 6 måneder til 2 år
___ Mer enn 2 år

7 b). Hvilke moduler arbeider du mest med? (Kryss av for maks. tre moduler)

___ Kontakt
___ Aktivitet
___ Dokument
___ Prosjekt
___ Email
___ Netthandel
___ Reskontro
___ Produkt
___ Tilbud\Ordre\Faktura
___ Regnskap
___ Lager
___ Timeregistrering
___ Innkjøp

7 c). Hvor mange brukere arbeider med Mamut Business Software på arbeidsplassen din?

___ En bruker
___ To eller flere brukere

Spørreskjema – Oppgave 1

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å finne aktivitetene dine? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

2. Hvor lett eller vanskelig var det å oppdage ruten med Mine Aktiviteter? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema - Oppgave 2

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å endre ruten Kontantbeholdning til Mine ubehandlede innkjøp? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 3

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å finne status for salg? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 4

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å opprette leverandøren? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 5

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å finne utfyllende informasjon om kunden? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 6

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å opprette aktiviteten? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 7

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å finne ordre\faktura-modulen? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 8

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å forstå hensikten med diagrammet? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

2. Hvor lett eller vanskelig var det å forstå at man kan gå til innkjøpsmodulen via diagrammet? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 9

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å finne kontaktpersonen og å opprette en ordre ved bruk av diagrammet? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 10

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å legge til en ny aktivitet under Mine aktiviteter? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 11

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å endre snarveisknappen? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 12

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å legge til en snarvei i Mine Funksjoner-knappen? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 13

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å opprette leverandøren? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema – Oppgave 14

Deltager nr: **17**

Dato: **29.11.06**

1. Hvor lett eller vanskelig var det å finne kontaktpersonen og å finne invitasjonen? (sett ring rundt svaret ditt)

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

Spørreskjema - Oppsummering

Deltager nr: 17

Dato: 29.11.06

Dette spørreskjemaet gir oss tilbakemelding om hva du synes om Mamut Business Software 11. Sett ring rundt det som utrykker best hva du synes om produktet. Skriv ned eventuelle kommentarer under hvert spørsmål.

1. Hvor lett eller vanskelig var det å ta i bruk det nye skrivebordet i Mamut Business Software 11 (sett ring rundt svaret ditt)?

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

2. Hvor lett eller vanskelig var det å ta i bruk den øverste knapperaden?

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

3. Hvor lett eller vanskelig var det å bruke rutene som inneholdt informasjon?

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____

_____
_____

4. Hvor lett eller vanskelig var det å bruke den nederste knapperaden?

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

5. Hvor lett eller vanskelig var det å bruke diagrammene med knappene?

1 Veldig Lett	2 Lett	3 Verken lett eller vanskelig	4 Vanskelig	5 Veldig vanskelig

Andre kommentarer:_____
_____
_____
_____

6. Hvor nyttig eller unyttig var det å kunne gjennomføre oppgaver på forskjellige måter?

1 Veldig nyttig	2 Nyttig	3 Verken nyttig eller unyttig	4 Unyttig	5 Veldig unyttig

Andre kommentarer:_____
_____
_____
_____

7. Ville du anbefalt dette produktet for andre?

___ Ja ___ Nei ___ Vet ikke

Forklar hvorfor du ville anbefalt eller ikke anbefalt dette produktet for andre:
_____
_____
_____

8. Andre kommentarer: