Master thesis outline

Identifying Usability Problems in relation to User Experience with System and Tests

Quang Minh Phung
Spring 2007 - Summer 2007
Group d611b
Department of Computer Science
Aalborg University

The Faculty of Engineering, Science and Medicine
Aalborg University

Department of Computer Science

SEMESTER TOPIC:

TITLE:
Identifying Usability Over Time in relation to User Experience with System and Tests

PROJECT PERIOD:
January 4^{th} 2007 - June 6^{th} 2007

PROJECT GROUP:
d611b

AUTHOR
Quang Minh Phung,
qmp@cs.aau.dk

SUPERVISOR:
Mikael B. Skov,
dubois@cs.aau.dk

NUMBER OF COPIES: 4

NUMBER OF PAGES: 60

ABSTRACT:

This study describes an experiment to discover difference between system novice and expert performance of a web-based email system, but in addition also the dimension of user experience with test evaluations as well. For this purpose a longitudinal laboratory based usability evaluation and a ”snapshot” evaluation of a web-based email system was conducted. Both the numbers of different types of identified problems and the severity of the problems are investigated. The study shows that after only a period of two months of system experience the number of usability problems decreased signifiantly. The study also shows that experience with tests did not caused users to identify significant different usability problems compared to inexperienced test subjects.

Preface

This report of master's thesis was written during my 8^{th} semester at Department of Computer Science, Aalborg University and the thesis is based on the research of Skov et al [1].

Reading Instructions

References in this report are marked by $[x]$ where x is a number which refers to an item found in the bibliography.

Acknowledgements

This master's thesis would not had been possible if not for the valuable comments and useful guidance through out the project of my supervisor, associate professor Mikael B. Skov, Aalborg University. I would also like to give my thanks to all the willing test participants that has made the experiment possible. Finally, a thank to Aalborg University for the technical support and provide of the Usability Laboratory.

Signature

Quang Minh Phung

Contents

List of Figures

Chapter 1

Introduction

Nowadays, information technologies are everywhere and have become an integrated element in our daily life. We are today surrounded by technologies, from the machines on our desks at work and at home, to our household appliances, from libraries and banking to supermarkets. These systems are developed for some specific or more general purposes. Some systems have been developed to support people solving tasks in work environments while others also appear in other contexts such as in private homes.

The interaction between humans and technologies has become quite comprehensive, that is why it is essential to have usable systems because then it will bring along convenience and efficiency for users accomplishing a task using the systems. However, if systems are unusable and the user's experience when interacting with system is poor considerable negative consequences might arise. For example in our daily life we might get annoyed that we are unable to record our favorite tv show from the VCR without looking up in the manual, that we hit the power off switch with the remote control instead of zapping between channels, that it requires great effort of learning and memorizing in order to make a simple phone call with our new mobile phones etc. These problems may not be considered as serious since no harm can be done to us physically, but they do however frustrate and annoy us as users of systems. In the context of work settings these problems tend to be more serious in terms of ineffective use of systems, loss of data, and wrong usage of systems. According to Jacob Nielsen usability is a necessary condition for the survival of a Web site [3]. If it turns out that a web site is difficult to use or users feel lost they will end up leaving. This will results in a loss of either customers or members. Moreover, for a company, usable systems provide benefits in terms of cost savings and product quality in which employees can improve productivity through efficient operations or fewer code revisions.

In the light of the comprehensive interaction between humans and computers lack of usability can result in cost of time and effort, frustration and in cases it can also determine the success or failure of a system [4].

Testing systems for usability depends on making assessments of how a given target group of users would interact with it. The dimensions of the target group can be break down into several subjects such as age, personal backgrounds and sex. However, one of the most important dimensions is user experience since different systems are developed for different purposes with different demand of user experience [5]. For instance some systems such as a VCR or vending machine are only used occasionally and therefore the requirement for user experience is limited. However, for more high-risk systems like air traffic control rooms, airplane cockpits etc the users must be experts in order to operate these systems. Hence, it is not of insignificant whether you involve novices or experts when evaluating systems. Several guidelines for usability evaluation rely on the assumption that there is a difference between evaluating with novice or experts [1]. For the novice users to interact with a system for the first time, they prefer simple actions and ease of learning [6]. But what will happen as their experience increase with the system? What will happen to the same user as he develops experience with the system?

This is an important point of view when evaluating usability over time, but to setup such study requires observation over time and a lot of resources in which not all company are able to apply.

Even though several experiments have inquired into examine usability over time through longitudinal studies [1, 7, 8]. Longitudinal evaluation method has turn out to be quite resource demanding since it requires observation over an extended period of time. An alternative and less resource demanding method to examine changes over time called cross-sectional that is often used in health and psychiatric research could be an interesting approach. Through cross-sectional studies relations between different subjects are examined at a point in time instead of lengthy studies as in longitudinal.

The results and consequences of involving either novice or expert users in a usability evaluation is still being debated. Depending of the experience of the individual user the interaction with interactive systems affects different kinds of problems. It is commonly that systems are tested with users who use the systems for the first time []dumas. However, it could be expected that there might be some differences in user's behavior when they have gained experience with the system and maybe also result in different sets of identified usability problems, since some problems may have been overcome while other problems might arise.

Addressing the questions of does the user's experience with the usability of a sys-

tem changes over time when evolving from being inexperience to experienced system users, and will usability problems disappear over time, this study aims to examine the difference between novice and expert performance. The study is derived from the experiment of Skov et al [1] and also aims to examine the difference between system novice and expert performance, but in addition also the dimension of experience with test evaluations since results from the study of Skov et al [1] and others [9, 10] also called attention to dimensions other than the experience with the system that could influence differences in the experience of usability problems when users got accustomed to the test evaluations as well. From this perspective it would also be interesting to examine whether other dimensions such as experience with the tests might influence the results of usability over time.

Research Questions

The purpose of this research is therefore focused on the following research questions:

RQ1: What characterize the identification of usability problems over time when comparing inexperienced with experienced system users?

RQ2: How does the experience with test evaluations influence the outcome of usability evaluations?

Part I

Background

Chapter 2

Theoretical Perspective

The purpose of this part is to provide the general theoretical background for usability and usability evaluation methods that will form basis for this project. Moreover, an overview of the general experiments with novices and expert in the field of usability is given.

2.1 Usability

Usability is, in its broadest definition, the ability or fitness to be used for a particular purpose [11]. Usability can be seen as an attribute of every product, just like functionality. Functionality refers to what the product can do, and testing functionality means ensuring that the product works according to specifications. Usability however refers to how users work with the product, and testing usability means ensuring that users are able to find and work with the functions to meet their needs. According to Dumas et al. [5] correctly working functionality is not sufficient for a product to be successful since a product by itself has no value only insofar it is used. And use implies users. Therefore, the way users interact with the product should be a goal for designers and developers.

Usability is sometime defined narrowly in terms of utility [11], while another measurement of usability can be described as:

> ”The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” ISO [12].

This definition presents three components in the interaction that should be considered in a usability evaluation [11]:

Effectiveness The accuracy and completeness with which users can achieve their goals such as number of tasks performed. The measurement of effectiveness involves quality of solution and error rates.

Efficiency The rate in which certain tasks are solved by the user and the system and also resources expended on achieving them. The measurement of effectiveness involve criteria such as task completion time and learning time.

Satisfaction The subjective satisfaction with the system that users express through questionnaires and interviews. Satisfaction is the user's experience and attitude toward the use of the system. Measurement of satisfaction can be conducted through questionnaires, interviews, or rating scales .

Usability can be specified and measured by means of these criteria. However, the two components - efficiency and effectiveness - are the ones easiest measured and documented through task based usability evaluations. In order to gain overall usability it is suggested that measurement of these criteria should be considered independently because there seem to be little correlation between effectiveness, efficiency, and satisfaction (Kasper et al [13]). It would be misleading to assume that overall usability can be achieved by only measuring a subset of usability.
From these three criteria of usability, it would however be difficult to determine the user's experience with the system along with the problems and difficulties during the interaction if you only measure how many tasks were solved and the rate in which tasks are performed. A different but often applied method for usability evaluation is to examine usability problems that are experienced by the users. The purpose is to evaluate the interaction of the users with the system, with the objective of identifying aspects of the interaction that can be improved to increase usability [14, 15].

There exist different evaluation methods in the literature to evaluate the usability problems of a system where some are theoretically and others involve users [15]. Jacob Nielsen and Rolf Molich are leading experts within Web usability and their research is to identify usability problems in order to to address the errors and poor usability of systems. This method enables system developers to be able to draw up a list of usability problems to reflect, evaluate, and design better user friendly systems.

2.2 Usability evaluation

Usability evaluation methods (UEMs) are used to evaluate the interaction of the human with the computer for the purpose of identifying aspects of this interaction so that these can be improved to increase usability (Gray & Salzman) [14]. Three types of UEMs have been identified: empirical methods, inspection methods, and inquiry methods.

Empirical methods includes Think-Aloud Test and User performance test

Inquiry methods includes User satisfaction questionnaires and Interviews

Inspection methods includes Expert reviews and Heuristic evaluation

The methods can be divided into two broad categories: (1) one that gather data from actual users and (2) one that can be conducted without the present of actual users. It is however recommended that empirical tests are a requisite for valid user experience and attitude measurement in comparison to informal methods such as reviews and heuristic evaluations since these methods do not evaluate with actual users [11].

In this report a think-aloud test from the empirical usability method was applied. The same method was also applied in the study by Skov et al [1] to discover usability problem over time that this report was inspired from. In this think-aloud test the metrics for the severity rating of identified usability problems will be based on the three problem types stated by Molich [15] namely: critical, serious, and cosmetic. The definitions of these three problem types are:

Critical A critical usability problem is classified one that causes the users to not be able to complete the desired task by themselves, but have to ask for assistance or completely giving up on solving the task.

Serious A serious usability problem is classified one that causes difficulties for the user to complete the task and cause irritation and frustration that will result in lengthening of solving the task.

Cosmetic A cosmetic usability problem is less severe and only causes the user small surprise and doubt for a moment.

The purpose of applying this usability evaluation method is to measure the quality of a user's experience while interacting with a system. This will enable an comparison of the severity and rate of identified problems from the test subjects in which important overview of how usability over time can be studied.

2.3 Empirical Methods: Evaluating with Users

As earlier mentioned usability can be obtained by assessing the effectiveness, efficiency, and satisfaction (ISO 1997). Moreover, usability evaluations must also involve representative users performing representative tasks of a particular system.

Empirical methods are based on users' experience with the system. The evaluations are conducted through user experiments and observation in real work environment or through usability laboratories. These include tests such as Think aloud test and user performance where participants in testing express their thoughts of the system while executing the set of tasks [16].

2.3.1 Users

The purpose of involving users is to assess the effect of the interface and functionality of the system on the users so that they can identify any specific problems with the system. However, the implication of involving representative users in the evaluation is not trivial.

According to Dumas et al [5] different studies have examined which factors matter most for the choice of users, and they have found that relevant experience and motivation are more relevant for understanding differences in how users interact with systems than demographic factors such as education, income level and age, since most systems are evaluated and designed for users with average physical and mental abilities. In case of systems designed for a population with special characteristic it would be obvious to consider relevant demographic characteristics an important part of the user profile. A part from specific systems targeted at a specific target group, the more important factors are user experience with computers, experience with systems or similar systems.

2.3.2 Novices and Experts

The important of involving test subjects with different user experience is still being debated because different systems are developed for different types of users. Some high risk systems are only developed for highly experienced users such as air plane cockpits whereas other more diversified systems like web-based applications should range for both novices and experts [17]. Therefore, it is preferable to have test subjects that reflect the expected profile of the end users.

In order to determine whether a system is sufficiently simple and intuitive for beginners, or its level of learnability, testing with novice users is essential. However,

usability measurement from all levels of users would be needed to gain a picture of its full range of usability [18].

The user's experience with a specific application domain differs in different ways. According to Nielsen [2] three main measurable dimensions of user experience with a system can be taken into consideration. Figure 2.1 illustrates these three dimensions of user experience; experience with the system, experience with computers in general and experience with the task domain.

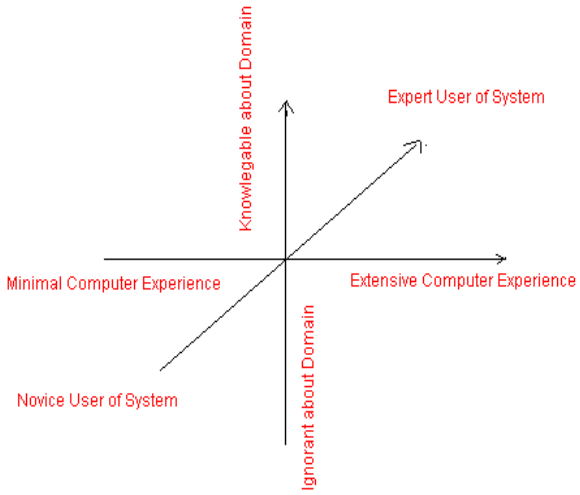

Figure 2.1: User experience model [2]

The acknowledgement of the need for usability evaluations of all levels of users has result in several empirical measurement between novice and expert users. However, different studies used different definitions of expert and novices. As example Dillon and Song [19] conducted a comparison study of textual and graphical interfaces for an art-resource database between user levels. The results from the two user groups were compared and it was found that expert performance was unchanged with the addition of graphical support, but novice performance was however improved. It was concluded that even though there were no significant differences in the accuracy with the solved tasks, the expert users performed faster than the novices. In another study, Jochen et al 1991 [20] examined errors by novices and experts when interacting with the computer in office work. In this study three criteria were used to determine the level of user expertise, namely total length of time the user has worked with computers, number of programs known and length of daily work time with computers. The result was however, in contrast to the widespread assumption that experts did not make fewer errors than novices, but instead spend less time handling the errors.

The studies mentioned above all showed that usability evaluations typically work with two levels of user experience: novice and experts. This is also in consistence with figure 2.1 in which the user's experience with a specific system is the dimension that is referred to when discussing user expertise. However, in this report another dimension that will be taking into consideration beside the user's experience with the system is user experience with the tests. The dimensions of user experience with the system and user experience with the test sessions are illustrated in 2.2. Results from several experiments [10, 9, 1] have indicated that test subjects often feel insecure and under pressure during usability evaluations because they feel like they were being assessed and not the system. This indication gives basis to this study's examination of whether user experience with test evaluation also has influence on the results of an evaluation.

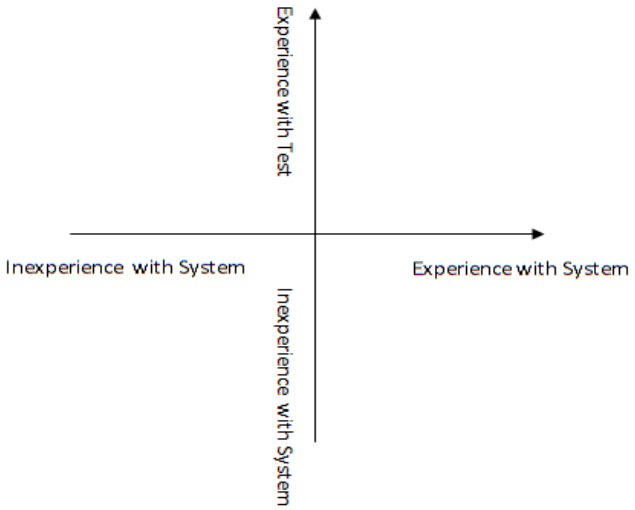

Figure 2.2: User experience with system and with test evaluations

As earlier mentioned in this report the examination of user experience will be based on the two dimensions:

1. Experience with system
2. Experience with test evaluations

The purpose of taking the dimension of user experience with test evaluations into account is to examine aspects concerning whether test subjects who are more experienced with test evaluations might be able to reveal more usability problems than if they were inexperienced with the feelings of insecure with the test environment.

Part II

Experiment

Chapter 3

Method

The following part describes how the usability evaluation experiment was setup and carried out. Furthermore, relevant details concerning participants, data collection, as well as the application domain are also described. The purpose of the usability evaluations was to examine possible usability problems and the workload that can be revealed over time.

3.1 Usability Evaluation Experiment

The dimensions of user experience that has been focused on in this report are between the system and test evaluations. An overview of the empirical study can be seen in figure 3.1.

The figure 3.1 gives overview of dimensions of user experience with both the system and the test. The grouping of the tests was set to the following names and these are described as follows:

IsIt Inexperience with System & Inexperience with Test

EsEt Experienced with System & Experienced with Test

IsEt Inexperience with System & Experience with Test

EsIt Experience with System & Inexperience with Test

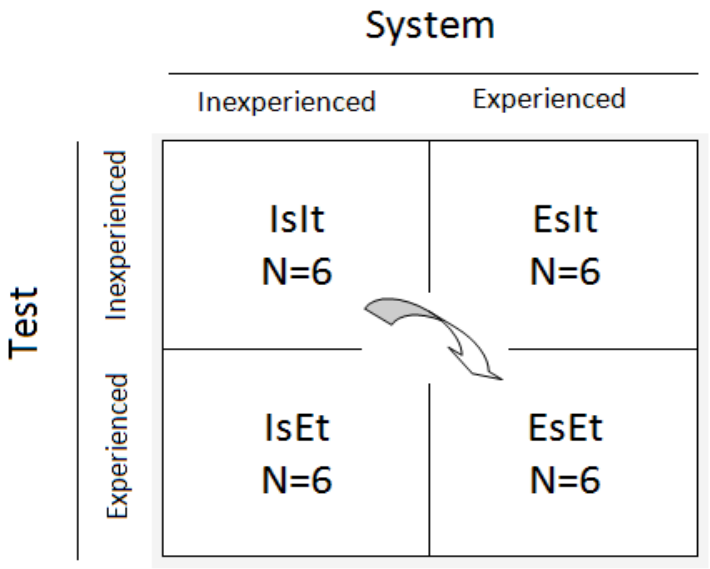

Figure 3.1: Overview of user experience and number of test subjects for each empirical study

Overall there were 18 test participants for this study. The arrow from IsIt to EsEt indicates that the same 6 test subjects were evaluated in both test sessions. The setup of this experiment was based on system use experience and test evaluation experience and these are described in the following.

3.1.1 System Use Experience

A longitudinal study for this experiment was based on two usability evaluations of the same system with the same users. The first evaluation (IsIt) was conducted in Marts 2007 where a sample of six test subjects were evaluated while using the system for the first time. The test subjects in (IsIt) were classified inexperienced with both the system and test evaluations. In the second evaluation (EsEt) the same six participants were again evaluated in May 2007 after a period of two months of using the system. The purpose was to measure possible changes of usability problems as the users gain more experience with the system and also the test evaluations.

In order to motivate the six test subjects to continuing using the system regularly so that experience with the system could be gained during the period of this experiment, several tasks were given throughout the period. The tasks included chain letters, news, jokes, calendar reminders etc. with the purpose of maintaining the activity of this group.

3.1.2 Test Participation Experience

The experiment of evaluating two user samples with different experience with both tests and system was carried out through two evaluations and these are outlined as following:

Experienced test users & Inexperienced system users (EtIs)

In this evaluation (EtIs) six test subjects who were experienced with test evaluations but inexperienced with the system were evaluated. The purpose was to compare the outcome of usability problems identified by this group of test subjects compared with those in the (IsIt) group since both were inexperience with the system but, had different experience with test evaluation

Experienced system users & Inexperienced test users (EsIt)

In this evaluation (EsIt) six experienced user of the system but inexperience with tests were recruited and evaluated. The reason why only experts were evaluated was due to the fact that six novices were already found in the EtIs group. Therefore, by merging the results of the evaluations from these two groups (EsIt & EtIs) a comparison of usability problems identified by different user experience can be observed.

3.1.3 Yahoo email System

The purpose of this study was mainly to examine usability over time for different user experience. However due to the fact that the study was limited to an extended period of time, a simple but yet extensive email system was chosen. In this way inexperienced users will be able to become quite familiar and gain considerable experience with the system within the evaluation period.

The interactive system used in this study was a web-based email system called Yahoo. A screenshot picture of the Yahoo email system can be seen in figure 3.2.

Yahoo provides free email to any person with access to a web browser and the Internet. This system provides basic email feature such as compose, send, attachments, filters and additional features such as notepad and calendar function for the users. The web-based email system Yahoo was chosen due to its accessibility and popularity worldwide, but at the same time also because it was not so well known and used like Hotmail. In this way it would be easier to recruit users who were inexperience with Yahoo compared to Hotmail.

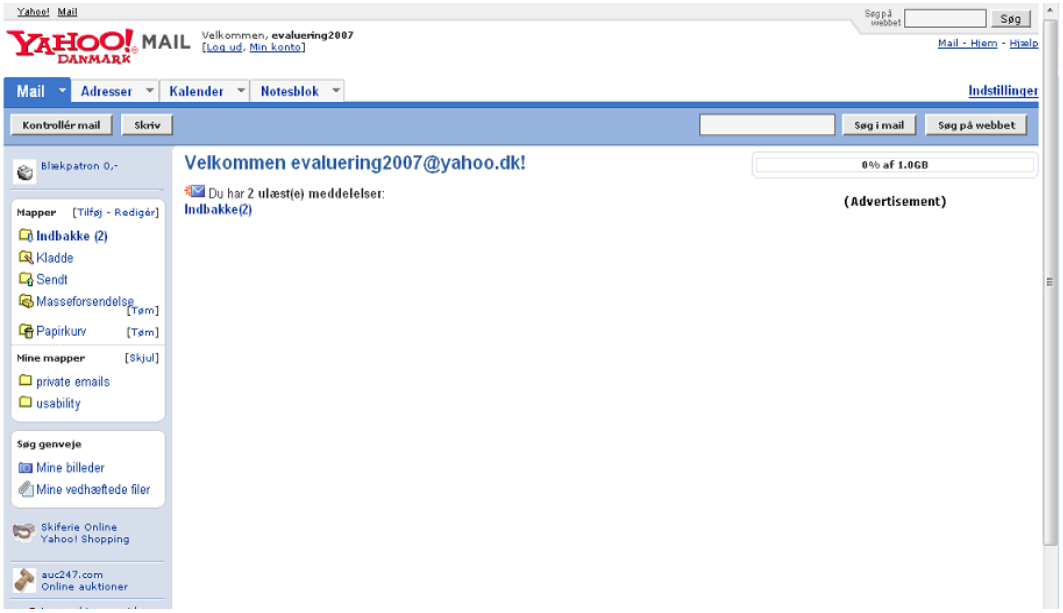

Figure 3.2: Illustration of Yahoo mail main page

3.1.4 Subjects

In this project 18 users participated in the experiment as test subjects. The distribution of these 18 test subjects can be seen in figure 3.3 .

From figure 3.3 it can be observed that 6 subjects (3 female & 3 male) in IsIt were the same that was evaluated in EsEt, whereas the 6 subjects in IsEt were different from IsIt and these 12 subjects were all male.

The selected number of test subjects was based on the assumption that it was not able to perform an unlimited number of evaluation sessions due to time and resource shortage.

The test subjects were all students at age 22-28 from either Informatics, Computer Science, and Humanities at Aalborg University. The test subjects were all

Group	IsIt	EsEt	IsEt	EsIt
Number of subjects	6	6	6	6
Sex	3F, 3M	3F, 3M	6M	6M

Figure 3.3: Overview of the distribution of test subjects in the groups

recruited either via email or oral invitation. The classification of the test subject's experience with both the system and test sessions was carried out through questionnaires. Different parametric were asked in the questionnaires such as personal characteristics the knowledge of using the computer, the amount and the frequency usage of Internet and their experiences for web-based email system. On this basis experience of the users could hence be classified into experienced or inexperienced users.

A user is classified experienced with the system if he or she had used the system for over a period of few months. Although real practice effects like piloting an air plane or driving a car cannot be managed in such a short time, the nature of learning how to use a email system is much simpler. Therefore, a user was considered experienced with the system if he or she had used it for few months. A user was experienced with test evaluations if he or she had participated in usability evaluations for several times. In this study a person who had no experience with the system in term of usage could still had experience with similar systems.

3.1.5 Tasks

The purpose of the usability evaluations was to examine the usability of a Web-based email system called Yahoo Mail. The tasks for this scenario were derived from another usability test report called Comparative Usability Evaluation (CUE) where professionals conducted a usability evaluation on a web-based email system called Hotmail. Since Hotmail and Yahoo Mail are similar types of web-based email systems, same types of usability tasks will be applied for this project. The tasks were intend to have the participants use a number of different functions in Yahoo Mail focusing on revealing potential usability problems related to each task. During the usability evaluation seven tasks were given to the participants to carry out using Yahoo Mail. The seven tasks will be enclosed in the appendix A and were based on following actions:

- Create account

- Send/Read mail with attachment
- Calender function
- Retrieve lost password
- Add folders
- Add email addresses

3.1.6 Usability evaluation procedure

In this study the test sessions were based on the think-aloud protocol as described by Molich [15]. The thinking-aloud technique is an important method for practical evaluation of user interfaces since it involves participants to think aloud as they perform a set of specified tasks. The thinking-aloud evaluation of the web-based Yahoo mail system was conducted for all test sessions. Throughout the sessions while participants used the system to solve seven tasks they were asked to say whatever they were doing, thinking and feeling. However, if the participants were not able to solve a task or had problems to continue on their own, the test monitor will provide them with helps and hints to find a solution. If they were still unable to complete the task even with help, they will be asked to go on to the next task.

3.1.7 Data Collection

The setting for the usability evaluation all took place at a dedicated usability laboratory at Aalborg University. During the four test sessions test subjects were all recorded on digital video while performing the tasks. The video recording included a PC screen along with a small image of the test subjects. From this all the actions of the subjects could be monitored along with their physical reactions. In figure 3.4 a screenshot of the recorded video screen is illustrated.

From the recordings all actions from the subjects, the time spent on solving each tasks, as well as their reactions while performing the tasks could be monitored since the data are relevant for the upcoming analysis.

3.1.8 Data Analysis

In order to gain information from the test sessions for data analysis, all test sessions were recorded on digital video. The video recordings were afterward examined thoroughly for identification of usability problems and also for the time spent

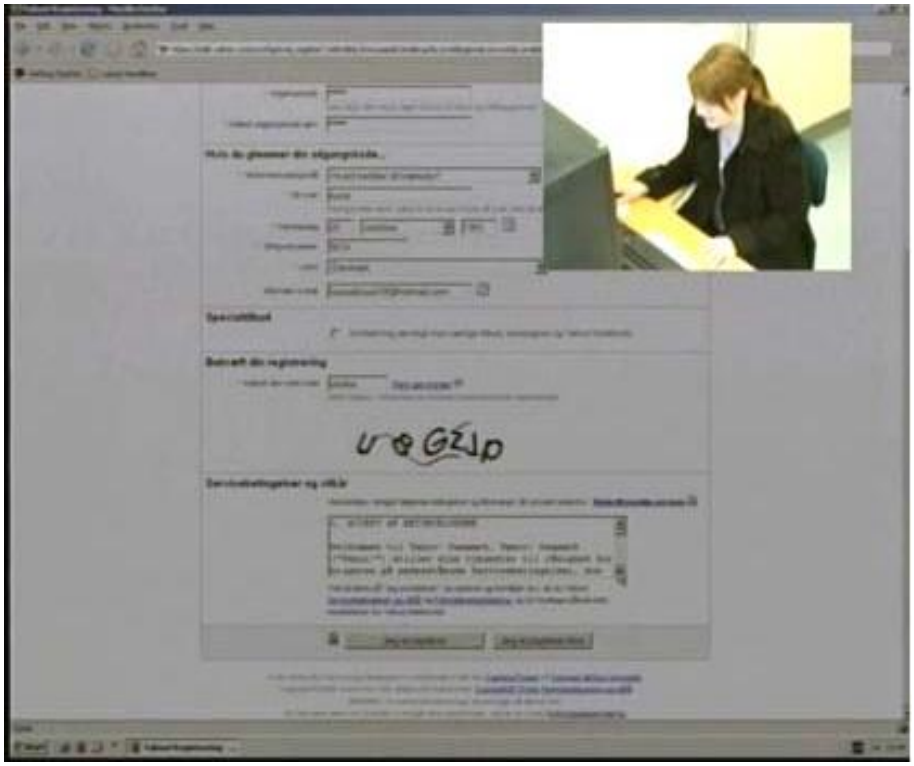

Figure 3.4: A screenshot picture of the recorded video from test evaluations

on solving each tasks. During the evaluation three types of data were collected for analysis:

Usability problems The usability problems were based on the guidelines stated by Molich [15] in which they were classified as critical, serious, and cosmetic. These usability problems were produced through analysis of the video recordings.

Performance By monitoring the subjects performing the tasks, collected data concerning the rate in which tasks were solved could be deduced. The measurement of the performance was also an indication of the effectiveness of the system.

Workload After the test subjects had completed the seven tasks, a workload measurement was made in terms of a task load index (TLX) technique from NASA [21]. During the TLX test, the test subjects were asked to rate their perception of the workload in a scale of 0 to 100 for each task. These subscales include mental demands, physical demands, temporal demands, own

performance, effort and frustration. The purpose of the TLX test was to measure the test subject's subjective experience with the given tasks and system.

Chapter 4

Results

In this chapter an overview of the identified usability problems from the test evaluations were analyzed and compared. First an analysis of the distribution of the identified usability problems following by the average number of identified usability problems, and finally an analysis of the subjective workload of the subjects while performing the tasks.

4.1 Distribution of Usability problems

To sum up on the categorization of the evaluation groupings for easier understanding of the following tables and figures, an explanation for the evaluation groups are described as follows:

IsIt Inexperience System & Inexperience Test

EsEt Experience System & Experience Test

IsEt Inexperience System & Experience Test

EsIt Experience System & Inexperience Test

From the analysis of the four test evaluations, a total number of 26 usability problems were identified where the total severity was assessed to 3 critical problems, 10 serious problems, and 13 cosmetic problems. The distribution of the total numbers of identified usability problems from the four test sessions can been seen in table 4.1.

	IsIt (N=6)	**EsEt** (N=6)	**IsEt** (N=6)	**EsIt** (N=6)
Critical (3)	3	1	2	0
Serious (10)	9	3	9	0
Cosmetic (13)	12	7	10	7
Total (26)	24	11	21	7

Table 4.1: Distribution of total numbers of identified usability problems, where N is the number of test subjects

As table 4.1 illustrates it could be stated that all test subjects who were inexperienced with the system (IsIt & IsEt) had identified mostly all of the experienced usability problems compared to those who had experience with the system (EsEt & EsIt). Moreover, it could also be observed that inexperienced system but experienced test users (IsEt) had identified almost the same total numbers of usability problems as the inexperienced test and system users (IsIt).

The test subjects that were inexperienced with both test evaluations and system (IsIs) experienced a total of 24 out of 26 usability problems in which 3 problems were assessed to be critical, 9 problems were assessed to be serious, and finally 12 problems were assessed to be cosmetic. The inexperienced system but experienced test users (IsEt) had identified a total of 21 out of 26 usability problems in which 2 problems were assessed to be critical, 9 problems were assessed to be serious, and 10 problems to be cosmetic. In contrast the experienced system users only identified under half of the total numbers of usability problems. Surprisingly the experienced system users users with no test experiences (EsIt) identified none of the critical nor serious problems and only 7 of 13 cosmetic problems. The result of this finding is debatable and give rise to some reflection that will be discussed in the discussion section.

To sum up we could conclude that by looking at the total numbers of usability problems the inexperienced system users identified mostly all usability problems in comparison to experienced system users. Furthermore, we could also state that experienced test users did not identify more usability problems compared to inexperienced test users.

By observing the distribution of the identified usability problems in figure4.1, it could be stated that overall a rather similar distribution of usability problems were identified by inexperienced system users (IsIt & IsEt). In contrast, a more dissimilar distribution of usability problems were experienced by the experienced system users (EsEt & EsIt) since the EsIt group identified none of the critical nor

serious problems whereas EsEt group identified 1 of 3 critical problems and 4 of 10 serious problems.

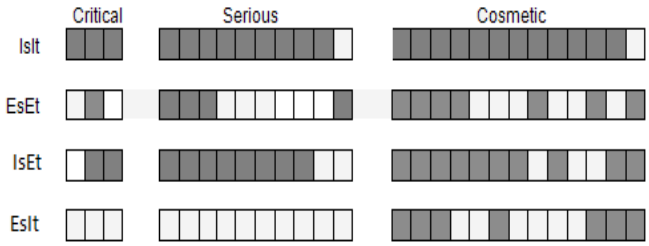

Figure 4.1: Distribution of unique usability problems, where each column represent a unique problem and a colored box indicated that the respective user group identified a usability problem. A white box indicates that a problem was not identified.

If we examine further into the individual level of identified usability problems we will be able to observe the distribution of individual user's experience with usability problems. The figure 4.2 gives overview of the number of usability problems experienced by the individual test subjects. Due to some technical problems during the data collection process, the results from two test subjects in EsEt could not be collected. This is also expressed in figure 4.2, where the test subjects nr 5-6 had not identified any usability problems.

Figure 4.2 summarizes the distribution of the identified usability problems where the numbers from 1 to 26 represent the identified 26 usability problem and each column represent one usability problem while a *X* indicates that the problem was experienced by the test subjects. A descriptioin of the 26 usability problems can be seen in appendix B

From figure 4.2 it was again clearly to observe a similar distribution of usability problems for inexperienced system users (IsIt & IsEt) whereas a more dissimilar distribution of the experienced system users were identified. In general many usability problems disappeared while few still remained. As example the critical usability problem nr 2 which was related to the logout function that enters a totally new random webpage instead of returning to the main page was identified by all inexperienced subjects. This problem was also identified afterwards by 3 subjects after a period of system usage. Surprisingly, there was one serious problem (nr 11) that was encountered only by one test subject in the EsEt group. This problem was related to the changing of user password, in which a user was trying to change her password in the option menu (Indstilling in danish). The user had difficulty of finding the option to change password because she was in the wrong page.

		IsIt						EsEt						IsEt						EsIt					
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Critical	1	x	x																						
	2	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x							
	3	x	x	x	x												x		x						
Serious	4	x	x	x				x									x	x							
	5	x	x	x	x									x	x		x	x							
	6	x	x	x	x									x											
	7	x	x	x				x						x											
	8	x	x	x										x											
	9	x	x	x	x									x	x										
	10	x	x	x														x							
	11									x															
	12	x	x	x	x		x							x	x	x		x	x						
	13	x	x	x										x	x		x								
Cosmetic	14	x	x	x	x									x	x										
	15	x	x																						
	16	x	x					x															x	x	
	17	x	x					x																	
	18	x	x	x	x			x	x					x	x	x	x	x	x		x	x			
	19	x	x	x	x	x								x	x		x		x						
	20	x	x	x	x	x	x							x	x	x	x	x	x	x			x	x	
	21	x	x	x	x	x	x							x	x	x	x	x	x	x		x	x	x	
	22	x	x	x	x	x	x	x	x	x	x			x	x	x	x	x	x	x	x	x	x	x	x
	23	x	x	x	x			x						x		x	x				x		x	x	
	24							x	x						x	x				x	x	x			
	25	x	x	x	x	x	x	x	x					x	x										
	26	x	x	x										x	x										
Total		24	24	20	14	6	6	10	5	3	1	0	0	17	14	8	8	9	8	4	4	4	5	5	1

Figure 4.2: Overview of the distribution of usability problems identified in the four test evaluations. Each number in the left side represent one usability problem where a *X* indicates that the respective user identified a usability problem.

Even though she was in the option menu, she was still only in the Mail option menu, and could therefore not find the proper function. It took the user several minutes before she realized that she has to change into the account information menu (kontoinformation) before she could change her password. This problem was quite unique since all other test subjects followed the task question that asked them first to logout of the account and then follow the guide of retrieving lost password.

A screenshot illustration of the problem nr 11 can be seen in figure 4.3. The highlighted function in the left side of the figure indicates that the user misunderstood the mail option from the account information.

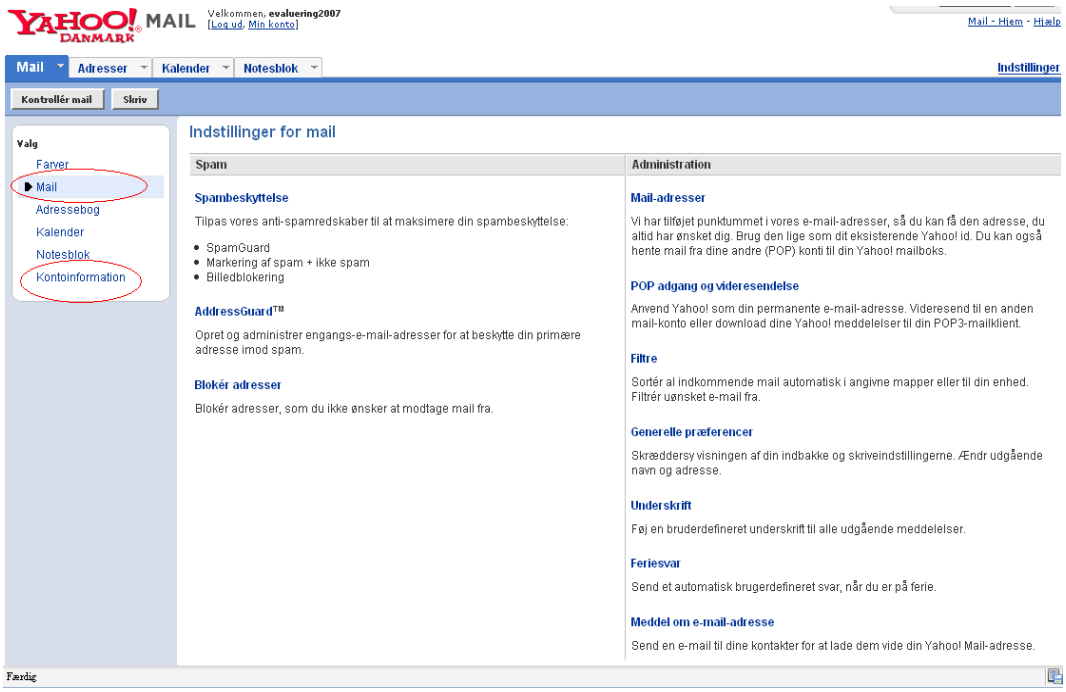

Figure 4.3: A screenshot of the serious problem that were only encountered by one user in the EsEt group

4.1.1 Average number of identified usability problems

By analyzing the average numbers of usability problems identified for each usability session showed in in table 4.2 , it could be stated that the inexperienced system users identified in average considerable more usability problems than the experienced system users. The six test subjects in the EsEt group identified on average only 3,2 usability problems after they gaining experience with the system compared to the first time when they identified 15,2 problems in IsIt. Furthermore, it could be observed that the experienced system but inexperienced test users (EsIt) identified on average even less usability problems.

In order to examine the variance in the number of identified usability problems from the four test evaluations, an analysis of variance (ANOVA) between groups was applied. The one-way ANOVA analysis showed that there was a significant difference between the groups in the number of identified problems where F_(6,4)

	IsIt (N=6)	**EsEt** (N=6)	**IsEt** (N=6)	**EsIt** (N=6)
Critical	2,0 (0,9)	0,8 (0,5)	1,3 (0,5)	0,0 (0,0)
Serious	4,8 (3,7)	0,7 (0,9)	3,3 (1,9)	0,0 (0,0)
Cosmetic	8,3 (3,3)	2,2 (2,8)	6,5 (2,3)	3,8 (1,4)
Total	15,1 (8,3)	4,8 (3,8)	10,6 (3,8)	3,8 (1,4)

Table 4.2: Overview of the average number of identified usability problems along with a derivation number ()

= 3,3 , p=0,004. And by using the post-hoc test it was possible to examine any significant differences across the groups. A post-hoc test showed significant difference at 0,1% level between inexperienced system users and experienced system users. These significant differences were between the IsIt and EsEt groups where p=0,004, and between the IsIt and EsIt groups where p=0,006. Furthermore, a significant difference at 0,5% level was between the IsEt and EsIt groups. However, the differences between experienced and inexperienced test users was not significant since $p > 0{,}05$.

To sum up it can be concluded that based on the average SD value of the groups there was in general considerable spread in the numbers of identified usability problems among individual users. This occur especially for the IsIt (SD=8,3),EsEt (SD=3,8), and IsEt (SD=3,8) while the EsIt (SD=1,4) had minor spreads. The analysis of variance between groups also showed that there was significant differences across the groups in the numbers and severity of identified usability problems and this was explainable through the significant difference between inexperienced and experienced system users. In contrast the experience with tests did not cause test subjects to identify significant different problems compared to inexperienced test subjects.

4.1.2 Task load Index

In this project a NASA-TLX test was used to measure how the subjects experienced the testing situation. The purpose was to measure the subjective workload of the test subjects in six factors: mental demand, physical demand, temporal demand, effort, performance, and frustration. For the six factors a value between 1 and 100 was attributed by the subjects depending on their assessment. Table 4.4 outlines the average level of all the test subjects' subjective workload in the six factors along with a SD value.

	IsIt (N=6)	**EsEt** (N=6)	**IsEt** (N=6)	**EsIt** (N=6)
Mental Demand	4,9 (2,5)	1,1 (0,6)	6,1 (2,6)	7,1 (2,6)
Physical Demand	0,1 (0,3)	0,0 (0,0)	0,0 (0,0)	0,0 (0,0)
Temporal Demand	6,3 (2,3)	2,7 (2,0)	16,6 (8,7)	6,2 (4,5)
Performance	4,3 (2,3)	3,2 (2,5)	8,5 (3,9)	4,6 (1,7)
Effort	4,3 (1,9)	2,1 (2,0)	3,5 (2,3)	5,5 (1,0)
Frustration	10,6 (5,5)	10,3 (8,8)	16,4 (9,9)	4,2 (3,7)
Total	21,4 (9,8)	19,3 (29,9)	51,1 (5,6)	23,4 (4,6)

Table 4.3: Average TLX-test value for the test evaluations

Table 4.4: Subjective workload where the average TLX-test values for the test evaluations are showed

The one-way ANOVA analysis showed that the differences in the subjective workload between the groups was not significant where F(9,3)= 0,36 , p=0,78. The minor differences could be explained from the insignificant differences between inexperienced system and test users (IsIt) and experienced system and test users (EsEt) as the total of subjective workload decreased from 21,42 in IsIt to 19,3 in EsEt. Even though the different was not significant, we can however observe that some test subjects still assessed the frustration factor to be more important while there were fall in all other factors. We can therefore state that the level of frustration had not changed even though the test subjects had gained experience with the system. However, the overall high SD number across the groups for the frustration factor showed that not all users assessed frustration as important as the other after gaining experience with the system.

We could from these results conclude that by gaining experience with the system the average workload of the users had minor fall although the level of frustration still remains.

Table 4.4 also showed that the subjective workload of the experienced test users (IsEt) were essential higher compared to the inexperienced test users (IsIt) from a total average of 51,11 in (IsEt) to 21,42 in (IsIt). This is mainly explainable through a higher average numbers in nearly all six workload factors for IsEt compared to IsIt.

We could from these results conclude that even though users had experience with other test evaluations, the average subjective workload while solving task are not smaller compared to those who are inexperience with tests. Overall it can be

stated although most factors reduced as the experience with the system increased, the level of frustration still remains.

In relation to the measurement of the time spent on solving the tasks, table 4.5 gives overview of the average time the test subjects spent on solving the seven tasks. Summarizing the time logs from the video recordings it could be stated that experienced system users spent lesser time solving the tasks. Moreover, the EsEt group was the fastest at completing the tasks in comparison to the EsIt group since they were already familiar with the tasks from earlier evaluation. From the observation of the time spent on solving the tasks we could conclude the combination of experience with system and experience with test evaluation had reduced the time spent on solving tasks.

In conclusion, a remarkably high number of usability problems disappeared together with reduce in the average amount of time spent on solving tasks as the test subjects gained experience with the system.

Group	**IsIt**	**EsEt**	**IsEt**	**EsIt**
Average Time	19 min (3,5)	14min (4,5)	20 min (4,1)	16 min (1,4)

Table 4.5: The average time test subjects spent on solving tasks along with SD values.

Part III

Reflections

Chapter 5

Discussion

The aim of this study was to examine usability over time by identify similarities and differences of usability problems experienced by inexperienced system users and experienced system users but also by inexperienced test users and experienced test users as well. Based on the results above; the numbers of identified problems, the nature of the identified problems and the lessons learned from conducting the evaluations, we present the following key findings:

Experience with the system

In relation to the identified usability problems, we found a significant difference between the number of problems by inexperienced system users and experienced system users. Not only did the numbers of identified usability problem decreased as users had gained experience with the system, so did the number of severe problems. The decreasing number of usability problems identified by experienced users may be due to users' development of workarounds. This especially account for the experienced system users in EsIt, who might had identified some of the critical or serious usability problems if they had not developed workarounds to compensate for these problems. As example when logging out of the system for relogin of another username, most of the experienced system users did not wait for the system to enter a default commercial web site before returning to the main menu, instead they just reenter the link of the main site and continue from there. Therefore, in relation to these workarounds many of the usability problems that were experienced by all other inexperienced system users were hence not experienced by several experienced system users.

Although the numbers of identified usability problem decreased dramatically as the experience with the system increased, we could still see that the workload had

however only a minor fall. Although there were fall in most factors in the TLX-test, the level of frustration still remained after a period of system usage. Mostly all test subjects had expressed that even though the interaction with the system became better after a period of usage, the small but many irritation of the layout, design and last but not least all the popups during the use of the system were unpleasant. All these small irritation had overshadowed the ease and satisfaction of use for these test subjects which was why the level frustration was not reduced.

In relation to the results of system experience in this study compared to other usability over time studies like Barendregt et al [8] and Skov et al [1]. Our study showed similar results as Barendregt et al [8] in which usability problems for children in a computer game decreased significantly as the children gained experience with the game after a short period of use. The efficiency, effectiveness, and satisfaction increased because the numbers of usability problems decreased. The similar outcome from our study and Barendregt et al [8] might be due to the relatively uncomplicated system and therefore easy to use and learn through a short period of use. However, in a study of a electronic patient record system of Skov et al [1] a remarkable number of usability problems still remained after one year of extensive use. Even though some usability problems disappeared over time, far from all still remains and some new appeared. In this perspective the difficulty level and type of different systems might influence experience of usability problems over time.

Overall we learned that the the numbers of usability problems decreased in this study as the users gained experience with the Yahoo web-based email system. Compared to other studies this study showed similar results in which usability problems decreased over time as users gained experience with the system. However, this might only apply for systems intended only for infrequently use by inexperience users to support quick and easy to learn, but for a more complicated system usability problems might still remain.

Experience with test

In relation to the usability problems that were identified by experienced test subjects and inexperienced test subjects, this study showed that the users' experience with the test did not result in significant differences of the identified numbers of usability problems nor the severity of them. During the test evaluations the experienced test subjects were better at thinking-aloud compared to inexperienced test subjects because they were more aware of the important of this process for the subsequent data analysis. Although the experienced test subjects were better at

thinking-aloud, they were less critical in their experience with the system which resulted in fewer identified usability problems compared to the inexperienced test subjects. This might be an expression of an inhomogeneous level of experience with computers and IT in general, since 3 of 6 subjects from the longitudinal study had less experience with computers and Internet in general and were those who identified most problems. However, this could also be a case of coincidence that was influenced by many other factors.

Results from the usability evaluation by Skov et al [1] called attention to the fact that even though inexperienced system users experienced significantly more mental workload and frustration than experienced system users, the cause may be due to dimensions other than the experience with the system, but maybe also the fact that they got accustomed to the test evaluations as well. Based on the results of this study the experience with tests did however not influence the results of usability evaluations or workload considerably.

Longitudinal vs Cross-sectional

In relation to the setup of this study's test evaluations. It can be said that the setup of (IsIt & EsEt) was based on the longitudinal study conducted by Skov et al [1] where two usability evaluations were conducted on the same system with same users. From this usability over time could be observed. However, in the other two test evaluations (IsEt and EsIt) different user experience samples were observed at one point in time in which this setup reminds of a cross-sectional study.

Cross sectional studies are often related to health and psychiatric research where observations concerning developmental processes or disease processes over time are observed. Cross-sectional studies are preferable for such experiments since it does not require large and lengthy studies as in longitudinal studies to give adequate statistical results [22]. A drawback concerning cross-sectional studies is related to previous experiences of the participants, since different participants are examined in a snapshot [23].

Several experiments have conducted cross-sectional studies to examine developmental process or disease processes over time. As example eight cross-sectional studies have investigated the relationship between indoor mold and respiratory, allergic or irritation symptoms [24]. Four out of the eight studies found significant association between mold exposure and asthma related symptoms such as cough, wheezing or breathlessness. In another example a cross-sectional study on male reproductive health from five European countries [25]. The participants were questioned about their health, current smoking habits. The result showed that men exposed to smoking had a reduction in sperm concentration, and also

that smoking may have longterm implications for the reproductive health of the offspring, which is why they also advise pregnant women to avoid smoking.

In this study the focus was to examine whether usability problems over time could be identified and are the results produced from a study similar to a cross-sectional study different from a longitudinal study. Based on the results identified from the IsEt & EsIt groups (cross-sectional) compared to IsIt & EsEt (longitudinal) groups, the longitudinal study identified more usability problems and a better user response in term of previous experience from the first test could be collected. Although the cross-sectional result from this study did not identified as many usability problems as the longitudinal, it might be an alternative method for usability evaluation over time if the purpose was to compensate the lengthy and resource demanding longitudinal study with a lesser intensive and comprehensive study. However, one should note that while it is relatively easy to find inexperienced system users, it can often be difficult to find experienced system users if the system is still under development.

Chapter 6

Conclusion

The purpose of this study was to inquire into the difference between inexperienced and experienced system users by studying users over time as they develop experience with the system. The motivation was to find interesting findings concerning how user's experience with the system changes and whether usability problems really disappear over time.

For this purpose a longitudinal laboratory based usability evaluation and a cross-sectional similar study evaluation of an web-based email system were conducted. The experiment described in this study examined similarities and differences of usability problems experienced over time respectively from experienced and inexperienced system users as well as experienced and inexperienced test users. Based on the results of this study we will now try to answer the research questions.

6.0.3 Research Question 1

What characterize the identification of usability problems over time when comparing inexperienced with experienced system users

The characterization of the identified usability problems over time as users gained experience with the system are outlined as follows:

The experiment showed that after only a period of two months of system experience the number of usability problems decreased significantly along with the number of severe problems. Experienced system users' developed workarounds to compensate for many problems and this was one of the main reasons why many usability problems disappeared over time and less became severe. In despite of a decrease in the number of usability problems as users gained experience with the

system, the level of irritation and frustration with the system still remains because of many irritating and annoying aspects such as poor design and usabilty of the system that could not be avoided even over time. Based on the results of this study we have concluded that usability problems disappeared over time and experience with the system influenced greatly the outcome of the experiment.

6.0.4 Research Question 2

How does the experience with test evaluations influence the outcome of usability evaluations?

In relation to answering this research question, we have concluded that experience with tests did not caused test subjects to identify significant different problems compared to inexperienced test subjects. Both experienced and inexperienced test subjects experienced more or less the same numbers of usability problems, and the differences were not significant. Although experienced test subjects did not identified significant different usability problems they were however better at thinking-aloud compared to inexperienced test subjects because they were more aware of the important of this process for the subsequent data analysis.

On basis of our experiment we have also gained insight in evaluating usability through a cross-sectional study compared to a longitudinal study. We found that even though the cross-sectional result from this study did not identified as many usability problems as the longitudinal, this could be an alternative method for usability evaluation over time to compensate for the lengthy and resource demanding longitudinal study. However, the comparison was only applied in one case, and in order to answer whether cross-sectional studies could be beneficial for analysis of usability changes over time, further researches will be needed.

6.0.5 Limitations

In the light of the process and the results of this study we have outlined three limitations that will be discussed in detail.

Data Analysis

First of all we will discuss the analysis procedure of the test sessions and its impact on the results. There is always a risk of evaluator effect when analyzing video materials according to Jacobsen et al [26] which means that different evaluators

identify different sets of usability problems and the severity assessment of these problems. In this relation it can be discussed whether the identified usability problems for this experiment were reliable or not, since only one person analyzed the results and also the same was present during the test evaluations. On one hand the result of this project was limited to only one evaluator which result in no disagreements. On the other hand the classification of the usability problems was also limited to one perspective with no correlations and this might be too biased. It can therefore not be excluded that the results was greatly influenced by one evaluator.

User selection

In relation to the implication of involving participants for test subjects we can state that the inhomogeneous level of experience with computer and IT in general of the test subjects in this study had greatly influenced the test results. The users with a high computer and IT experience identified relatively less usability problems in comparison to the inexperienced computer and IT users.

Time period

As the focus of this study was to examine usability over time the aspect of the chosen time period can be discussed. The inexperienced users in this project was limited to a period of two months to gain experience with the system, and it can be discussed whether this was too long or too little to be for these users to be considered experienced users of this system. Moreover, in relation to the realism aspects it can be discussed whether the users were well motivated to gain experience with the system. In the study of Skov et al [1] the test subjects were studied over a period of one year and these users were maybe more motivated in using their system in comparison to the system in this study, since their electronic patient record was highly related to their daily work. From this perspective proper motivation for using the system might influence the rate of user experience and the time it takes to become experienced system users.

6.0.6 Future work

As for future work, we have just outlined the limitations of this study in the aspects of selection of time period and consequences of realism on the identification of usability problems over time. From this perspective it would be interesting to

research further into the consequences and influences these aspects will have on usability problems over time.

Part IV

Appendix

Appendix A

Taks

1. Start en browser og find Yahoo Mail - Derefter registrer dig selv som en ny bruger.
2. Send en mail til qmp@cs.aau.dk hvori du skriver at du vil invitere mig til din fødselsdag. Du skal også have vedhæftet et billede med i din mail som findes på pc'ens skrivebord under navnet fest.jpg.
3. Du ønsker at oprette en kalender påmindelse for din fødselsdags fest på din fødselsdag.
4. Du bedes gå ud af din email profil og prøv at logge på igen. Denne gang antag at du har glemt det password og nu vil du gerne prøve at genfinde det.
5. Nu skal du prøve at logge på en anden profil der hedder evaluaring2007@yahoo.dk med kodeordet " testing"'. Du har nu fået en mail fra QMP med overskriften Hejsa. Læs mailen og gem den i en privat mappe hvor du skal kalde den for usability.
6. De er nu stadig logget på evaluaring2007@yahoo.dk og bedes tilføje 3 ny email adresser qmp82@hotmail.com ,qmp@cs.aau.dk og din nye yahoo email.
7. Du har fået en del emails fra Minh, men du har ikke tid til at læse dem nu, så du vil flytte dem til en privat mappe ved navn private emals

Appendix B

Identification of Usability Problems

Critical	1	Brugerne misforstod betydningen af "huske spørgsmål" med et ledende spørgsmål der skal minde dem om hvad deres kodeord var når de opretter deres bruger for første gang. Dette skyldes at de er vant til den slags struktur udfra andre email programmer. Men huske spørgsmål i Yahoo var mere om at vælge et prædefineret spørgsmål og skriv ens svar på det.
	2	Når brugeren logger ud af systemet bliver hun henvist til en enten en tilfældig yahoo side, eller yahoo's forside og ikke yahoo mail's forside. Dette medførte at mange mistede overblik og kunne ikke komme tilbage til mail forsiden for relogin.
	3	Når brugeren får et nyt (randomized) password, er der tvivl om hvordan man kan ændre det til det ønskede password. Mange havde besvær med finde ud af hvordan man ændrede det tildelte password til det man ønskede.
Serious	4	Brugerne ønskede at kunne definere deres eget huske spørgsmål, men pga at det var prædefineret og ingen af dem var brugbar for nogle af brugerne, gav det anledning til irritation og frustration der forlængede tiden med at fuldføre opgaven
	5	Når man logger ind for første gang kommer man ind i en helt tilfældig side med masser af reklame og andre funktioner og emner der slet ikke er relateret til yahoo mail. Dette gav anledning til forvirring og irritation.
	6	Formuleringen på funktionen send email var svært at forstå hvilket betød at det tog brugerne en del tid at finde det når man var vant til en anden formulering fra eks Hotmail. Dette skyldes oversættelses funktioner fra engelsk til dansk. Brugerne var ik vant til "skriv" men mere create mail/send mail
	7	Test personerne var ikke klar over hvormange felter der skulle udfyles når der skal oprettes en kalender note, nogle felter var markeret med gult, men dette betød ikke at de var nødvendigt at udfylde dem.
	8	Når brugerne har glemt deres password og ønsker at få den fremvist igen, bliver de overrasket over at den funktion ikke findes, men at der var kun et menu med at få en ny password. Dette førte til irritation idet de brugte en del tid på at finde funktionen der vil give dem deres gamle password igen.
	9	Gul markering af tilfældige felter der til forveksling ligner (*) der plejer at indikerer at disse felter skal udfyldes, gav ofte brugerne besvær og irritation, når de blev misforstået
	10	Sikkerhedsspørgsmål såsom kæledyrs navn gav anledning til forvirring og huske problemer idet brugerne havde ingen anden alternativer end at vælge en af dem.

Figure B.1: Overview of identified usability problems part1

Serious	11	For at ændre password under Indstillinger, er der forvirringer mellem indstilling for mail og konto. Ændring af password kan kun gøres under konto indstilling, men yahoo vil altid som default åbne indstilling for mail og ik konto.
	12	At oprette en ny mappe er ikke overskueligt, da knappen er alt for lille og overses ofte af brugeren.
	13	Under tilføjelse af en ny email adresse bliver nogle felter markeret med gul og igen skaber det forvirring og irritation.
Cosmetic	14	Feltet der står at kodeord skal være på mindst 6 tegn er alt for lille og overses af mange
	15	Der er sikkerhedstjeck på password kun for .com brugere og ikke .dk brugere. Dette skaber irritation idet mange brugere ønskede at få deres password sikkerhedstjekket.
	16	Brugernan ønskes undersøgt on the fly, men dette var også kun muligt på .com account.
	17	Bekræft kode felt med de svære og underlige tegn bliver et irritations faktor hvis man laver fejlskrivninger pga utydelighed
	18	For mange menu felter under kalender funktionen forvirrer mere end det er til gavn. Mange burde være skjut som default
	19	Logud knappen er alt for lille
	20	Placeringen af logud knappen er svært at genkende fra andre programmer der ligger i hjørnerne.
	21	Logud funktion der skal logges ud 2 gange før man er helt ude medførte irritation
	22	Efter login med det nye random password som brugerne fik tildelt kommer de ind i en ny menu der ser uoverskueligt ud. Skift password knappen er lille og ikke til at finde
	23	Mange mellem led og spring gennem flere sider for at skifte ens password skaber irritation.
	24	Spørgsmål om hvorvidt man har brugt ens kreditkort hos yahho, som skal besvares gav anledning til irritation og forvirring
	25	For mange menu felter under tilføj addresser gav også anledning til irritation og forvirring for brugerne
	26	Brugerne ønsker at der gøres opmærksom på et felt som (*) så de ved hvilke felter er nødvendige og hvilke er unødvendige

Figure B.2: Overview of identified usability problems part2

Appendix C

Summary

The aim of this study was to inquire into the difference between inexperienced and experienced system users by studying users over time as they develop experience with the system. The motivation was to find interesting findings concerning how user's experience with the system changes and whether usability problems really disappear over time.

For this purpose a longitudinal laboratory based usability evaluation and a "snapshot" evaluation of an web-based email system was conducted. The experiment described in this study examined similarities and differences of usability problems experienced over time respectively from experienced and inexperienced system users as well as experienced and inexperienced test users. The usability evaluations investigated both the numbers of different types of identified problems and the severity of the problems.

The study showed that after only a period of two months of system experience the number of usability problems decreased signifiantly. The study also showed that experience with tests did not caused users to identify significant different usability problems compared to inexperienced test subjects.

Bibliography

[1] Jesper Kjeldskov, Mikael B. Skov, and Jan Stage. Does time heal? a longitudinal study of usability. *Proceedings of OZCHI*, 2005.

[2] Jing Wu. Accommodating both experts and novices in one interface.

[3] Jacob Nielsen. Usability 101: Introduction to usability.

[4] Foraker design. Introduction to usability.

[5] Joseph S. Dumas and Janice C. Redish. *A Practical guide to usability testing*. Ablex, 1993.

[6] Ben Shneiderman. *Designing the user interface*. Addison-Wesley Longman, 1998.

[7] Valerie Mendoza and David G. Novick. Usability over time. 2005.

[8] W. Barendregt, M. M. Mekker, D. G. Bouwhuis, and E. Baauw. Identifying usability and fun problems in a computer game during first use and after some practice. *International Journal of Human COmputer Studies*, 2006.

[9] J. Preece, Y Rogers, and Sharp H. *Interaction Design: Beyond Human-Computer Interaction*. New York: John Wiley & Sons Inc., 2002.

[10] Rubin J. *Handbook of Usability Testing: How to plan, design and conduct effective tests*. New York: John Wiley & Sons, Inc, 1994.

[11] Brian Shackel and Simon Richardson. *Human Factors for Informatics Usability*. Cambridge University Press, 1991.

[12] Anjoo Navalkar. Usability engineering – quality approach (iso 13407).

[13] Kasper Hornbæk, Morten Hertzum, and Fjøkjær. Measuring usability: Are effectiveness, efficiency, and satisfaction really corelated? *CHI*, 2, 2000.

[14] Wayne D. Gray and Marilyn C. Salzman. Damaged merchandise? a. review of experiments that compare usability evaluation. methods. *Human Factors, Applied Cognitio*, 1998.

[15] Rolf Molich. *Brugervenlige edb systemer*. Teknisk Forlag, 1999.

[16] Pauline Genise. Usability evaluation: Methods and techniques.

[17] Jacob Nielsen. Novice vs. expert users.

[18] Laura Faulkner and David Wick. Cross-user analysis: Benefits of skill level comparison in usability testing. *Interacting with COmputers*, 17, 2005.

[19] Andrew Dillon and Min Song. An empirical comparison of the usability for novice and expert searchers of a textual and a graphic interface to an art-resource database. *Digital Information*, 1997.

[20] PRUMPER OCHEN, FRESE MICHAEL, ZAPF DIETER, and C. BRODBECK FELIX. Errors in computerized office work : Difference between novice and expert users. *ACM SIGCHI Bulletin*, 1991.

[21] NASA. Task load index.

[22] Helena Chmura Kraemer, Jerome A Yesavage, Joy L. Taylor, and David Kupfer. How can we learn about developmental processes from cross-sectional studies, or can we? *American Psychiatric Association*, 2000.

[23] Changyu Chen. Longitudinal studies.

[24] David J. Miller, Nicolas L. Gilbert, and Robert E. Dales. Fungal contamination in public buildings: Health effects and investigation methods. 2004.

[25] Tina Kold Jensen, Niels Jorgensen, Margus Punab, Trine B. Haugen, Jyrki Suominen, Birute Zilaitiene, Antero Horte, Anne-Grethe Andersen, Elisabeth Carlsen, Oystein Magnus, Valentinas Matulevicius, Ingrid Nermoen, Matti Vierula, Niels Keiding, Jorma Toppari, and Niels Skakkebaek. Association of in utero exposure to maternal smoking with reduced semen quality and testis size in adulthood: A cross-sectional study of 1,770 young men from the general population in five european countries. *American Journal of Epidemiology*, 2004.

[26] N. E. Jacobsen. Usability evaluation methods, the reliability and usage of cognitive walkthrough and usability test. 1999.