A Case Study In Methods For
Usability Evaluation
- of mobile phone applications in a controlled environment

Oscar Cosmo - Catia Hansen

Master Thesis

Department of Design Sciences
Lund University
ISRN: LUTMDN/TMAT-5079-SE

2005

Abstract

The purpose of this thesis was to examine the benefits for Goobi AB to use a controlled environment for usability tests of their products in the future. In addition flaws found should be reported along with a proposal for test method that can be used to test in future projects. (Due to disclosure circumstances Goobi AB is a fictive name for the company where the thesis was conducted).

Goobi AB is a world-leading company in the field of developing and selling information-related applications for mobile phones and other mobile devices.

The product suite is based on open standards and contains applications that include messaging, content management, browsing and more. Traditionally the attention has been mostly on protocol and standard aspects of similar applications but as operators and market require more focus on usability aspects Goobi AB wishes to be prepared for usability requirement that will show up in the future.

To meet the company's request and address the objectives for this thesis, two usability tests were performed in the test laboratory at Ingvar Kamprads Design Centrum in Lund. The tests were conducted according to written test plans, which were composed before each test. All together 37 tests were carried out; test 1 focused on identifying usability flaws and include 20 test persons plus 2 pilot tests while test 2 spotlighted the method for usability testing. Test 2 used 12 test persons and 3 pilot tests. Each test were videotaped and analyzed. Method and results for each test is presented as well as an overall result concerning the method for testing.

During the usability test it was found that the most obvious problem in this application suite is that there is very poor consistency in and between the applications. It appeared during the test that there are several ways to carry out the same action. The same problem apply to the soft keys; sometimes the left soft key is used to confirm and select items and some times it is the centre soft key and in a few occasions both keys operate the same way. Another major flaw is the feedback that varies from function to function. It is important for Goobi to point out that the identified flaws will be or has been attended to.

Results from the tests leads to a recommendation for Goobi AB how usability evaluation can be conducted. The recommendation presents two separate ways of conduct. These differ in that the first aims to identify primarily quantitative data and is best performed early in the development process

while the other is carried out late in the process and has a more validating approach. The later method focuses on identifying qualitative data to compare against usability levels or other products.

This report contains a theory part which is supposed to provide a background and work as an orientation within the subject. Moreover this part put the discipline of usability engineering into context as well as reflects the knowledge basis of the authors. Readers familiar with the area may disregard this chapter.

Sammanfattning

Syftet med detta examensarbete var att undersöka fördelarna för Goobi AB att utnyttja en kontrollerad miljö för att utföra användbarhetstester i framtiden. Dessutom skulle funna användbarhetsdefekter rapporteras tillsammans med ett förslag på testmetod som kan användas för att testa i framtida projekt. (Goobi AB är ett fiktivt namn då det egentliga namnet på företaget inte kan nämnas på grund av sekretesskäl används).

Goobi AB är ett världsledande företag inom utveckling och försäljning av informationsrelaterade applikationer för mobila enheter, exempelvis mobiltelefoner.

Produktsviten är baserad på öppna standarder och innefattar applikationer med funktioner för att hämta och skicka meddelanden, innehållshantering, och Internetåtkomst, bland många andra. Tidigare har det under utvecklingen fokuserats mest på tekniska aspekter som protokoll och standarder men allteftersom marknaden, operatörer och slutkund, intresserar sig mer på bättre användbarhet vill Goobi AB vara väl förberedda för framtida krav.

För att syftet med examensarbetet gentemot Goobi AB skulle bli uppfyllt utfördes två användartestomgångar i användbarhetslaboratoriet på Ingvar Kamprads Design Centrum i Lund. Varje test följde en utvecklad testplan. Allt som allt genomfördes 37 enskilda test i de två omgångarna. Test 1 fokuserade på att identifiera existerande användbarhetsproblem i produkterna medan test 2 koncentrerades på metoden för användbarhetstest. I test 1 ingick 20 individuella test och utöver detta två pilottest. Test 2 använde 12 testpersoner för skarpa test plus 3 pilottester. Varje test videofilmades och analyserades. Testmetod samt resultat för varje test presenteras i rapporten, likaså övergripande metod och resultat.

Under användbarhetstesterna kunde det observeras att det mest uppenbara problemet med applikationssviten är dålig konsekvens i och mellan applikationerna. Det visade sig under testerna att det fanns flera sätt att utföra samma funktion. Även knapparnas funktion led av detta problem. Ibland användes vänster *softkey* för att konfirmera och välja, ibland användes höger *softkey*, och ibland kunde båda användas för samma funktion. Ett annat stort problem som uppdagades är variationen på eller avsaknaden av återkoppling, *feedback.* Goobi AB anser det viktigt att påpeka att de problem som identifierats kommer att eller har åtgärdats.

Resultatet av testerna leder till en rekommendation till Goobi AB hur användbarhets-utvärdering kan utföras. Rekommendationen presenterar två olika tillvägagångssätt. Dessa skiljer sig i utformandet på så sätt att den

första avser att ta fram främst kvalitativ data och utförs med fördel tidigt i utvecklingsprocessen. Den andra utförs sent i processen och är av validerande karaktär. Metoden fokuserar på att ta fram kvantitativ data som kan användas för att jämföra mot användbarhetsnivåer eller andra produkter.

Följande rapport innehåller ett teoriavsnitt avsett att ge en bakgrund samt fungera som en orientering inom området. Dessutom sätter detta avsnitt användbarhets-centrerad utveckling i sammanhang och ger en bild av författarnas kunskapsbas inom området. Läsare redan förtrogna med området kan ignorera nämnt kapitel.

Preface

This thesis is a part of the final examination for master of engineering at the institution of technology in Lund. The research work presented in this report has been conducted in the mobile telephones research and development apartment at Goobi AB in Lund and the tests has been performed in the usability test laboratory at Ingvar Kamprad Design Centrum (IKDC). This is not a traditional academic research setting since the goals of the master thesis were mostly business-driven and had an experimental rather than a classic literary approach. It was required from the tutor that the thesis was performed "by the book", and the guiding book was *Handbook of usability testing, How to plan, design, and conduct effective tests* by Jeffrey Rubin [2].

Both authors of this master thesis are educated within computer science majoring towards human computer interaction and prior this thesis they both took a class in usability evaluation (MAM120) at the institution of technology in Lund. In this class Oscar Cosmo evaluated IKEA kitchen planer. An online program where users can enter the size of their own kitchen, drag and drop furniture from IKEA and view it in a 3-D view. Catia Hansen evaluated BlueBellMouse, a graphical user interface, written to facilitate laboratory lessons in combustion technology at the department of physics in Lund. (Laboratory lessons were before BlueBellMouse performed on a command based system.) Totally both tests comprised eleven test persons.

Since the needs and demands from school board and Goobi AB are slight different this report is an attempt to work as a basis for forthcoming usability tests at Goobi AB but also to stimulate the interested student. It is desirable that this report can be reed free-standing from the test reports of test 1 and 2 therefore some information is duplicated. The aim of the content in theory is to present an overview of the context in which usability and usability tests appear.

Acknowledgements

We would like to thank the people who helped us throughout this master thesis.

Thanks to:

- Our academic tutor **Joakim Eriksson** at the institution of design science for arranging this thesis, letting us figure out our own answers and make our own mistakes, and for not giving us all the answers ahead.
- **Anders Borg** and **Simon Thorsander** our contacts at Goobi AB, for providing us with necessary equipment, feedback and support.
- **Arne Svensk**, **Magnus Haake**, and **Jonas Borell** all active at the department of design science, for feedback and important comments concerning the questionnaire.
- All those who participated in the testing; we could not have done this without you.

I, Catia Hansen, would like to take the opportunity to thank the people who supported me through my education in Lund, so I add my warmest gratitude to**:**

- **Oscar Cosmo** for accomplish this thesis together with me and for all knowledge you shared. Thank you for five entertaining months, four years of friendship and hopefully more to come.
- My parents; mother **Kirsten Hansen** for inexhaustible love and support and father **Marko Kavio-Oja** for letting me know he always loves me.
- My sisters; **Caisa Hansen** for encourage and always being there for me, **Heidi Asperheim** for good advice in life and inspiration, **May-Lill Cervin** for laughs and positive perspectives.
- My true friends; **Isabel Zavalia**, **Marleen Svensson, Johanna Carlsten, Jacob Larsson**, and **Josefin Lindqvist**, for support, laughter and being true friends. And my new won friends **Erik Lindberg** and **Hannafia Lasson** .You all know how much I love you.
- All the unmentioned persons, such as teachers, classmates and old boyfriends and other familiar faces. :-)

I, Oscar Cosmo would like to personally thank:

- **Catia Hansen** for telling me I'm wrong.
- **Josefin Lindqvist** for letting me drive her everywhere.
- **Bantorget crew** and everyone else who feel they deserve it.

Table of contents

1. Introduction

This section is an introduction to the background, purpose and limitations of this thesis.

1.1 Background

In September 2004 one of Teleca's companies, Teleca Mobile Technologies changed name into Obigo AB to closer correspond to the product with the same name. Obigo AB is still a subsidiary to the Teleca group and a world-leading company in the field of developing and selling information-related applications for mobile phones and other mobile devices. The product suite is based on open standards and contains applications that include messaging, content management and browsing and more. Traditionally the attention has been mostly on protocol and standard aspects of similar applications but as operators and market require more focus on usability aspects Obigo wishes to be prepared for usability requirement that will show up in the future [15][16]. As the demand of usability is increasing Obigo AB wanted to examine the possibilities test laboratories at Ingvar Kamprads Design Centrum (IKDC) had to offer.

1.2 Purpose

The purpose of this thesis was to examine the possibilities for Goobi AB to use IKDC's test laboratory for usability tests in the future. The objectives of the thesis were led by the objectives for each test and could in the end be summarized under the headline for this report; *A Case Study in Methods for Usability Evaluation*
-of mobile phone applications in a controlled environment.

It was requested to conduct a usability test on three of Goobi ABs products. Partly to make one example on how to carry out a usability test in that environment, but also to report the logic flaws in these products found during the test. It was desired that the following questions were answered or addressed in some way.

1. How many test persons:
    a) needs to perform the test to identify trends?
    b) needs to perform the test to receive reliable results?
    c) shall perform the same test at the same time?
2. How to set up test environment?
3. Which device is suitable to perform the test on?
4. Where to place test monitor and technical equipment such as cameras and microphones?
5. How much shall the test monitor interact with the test person?
6. Should the "think aloud" method be applied or not?

This thesis seemed to be important from two different angles. On one hand it is important for the institution of design science to promote their facilities and maintain a healthy relationship with the industry. On the other hand Goobi AB wanted to examine what the test laboratory could offer them as a company.

1.3 Limitations and problems

During this thesis some problems occurred and some limitations were defined. These problems and limitations are presented by activity.

1.3.1 Presented theory

There are running meters about usability, HCI, software engineering and other relevant literature in this field. To prevent absorption into new literature, one limitation in this thesis has been only to present theory from already studied sources. Several surveys have been done in this area and still there are plenty of confusing definitions and explanations. This is not an attempt to distinguish conflicting descriptions or to present a complete sum up in this area. It is just some lines to give the reader a grasp on the field.

1.3.2 Purpose for thesis

Originally the question at issue for this thesis was; which way is the best way to usability test Goobi AB´s mobile telephone products? With this problem formulation the thesis was supposed to be divided into two tests. The first test were to be conducted on an already released product and was to be some kind of pilot test for the second test, which aimed to do a sharp usability test on a new version of the product before it was released. Then the two tests should be compared in terms of usability. However, since the new version was not ready within the time for the thesis, the purpose and question at issue was reformulated. As a consequence parameters as which environment is the best to use during usability tests was not treated.

1.3.3 Writing test plans and reports

As a consequence to the vague formulation of purpose and question at issue, it was difficult to know which information to include in the main report, describing the thesis itself, and which information to include in the two test plans. Another confusing fact was to whom the report should be directed, Goobi AB or a student[1]. The needs and demands from school board and Goobi AB were different, which affected contents, language and layout in the report.

[1] Schools guidelines for master thesis can be found at www.it.lth.se.

1.3.4 Conducting the test

Some minor problems and limitations arose during the thesis. For example an important device was not ready or in its place when needed and the access to the test laboratory was limited because of an ongoing class in usability testing. Because of the insecurity in when it was possible to conduct the test no extensive search for test persons was made. Instead friends were asked to participate just before the tests began. This however led to a shortage of female novices. Another factor that contributed to the lack of novices was that the asked novices felt insecure about their knowledge in the English language. As a consequence not all data in the first test were analysed[2].

[2] Test with remote control, two persons, and one male expert were discarded.

2. Theory

Following chapters aim to present an overview of the history and evolution of mobile telephones as well as provide a background for usability evaluations. Why it is needed and when in the development process it can be carried out. To clarify this some background to the software development process itself is required and hence given.

2.1 Background

Several inventors and technologies lies behind the development of what today is referred to as a mobile telephone.

2.1.1 From radio to mobile telephone

It is not obvious where or when to start a description on the evolution of the mobile phone. In some literature the history of the mobile phone began in 1842 with the fax machine, patented by the Scottish inventor Alexander Bain[3] while other prefers to highlight James Maxwell who predicted in 1864 that radio transmission was possible [5]. Predictions that were supported through experiments in 1888 by German physician Heinrich Hertz' experiment [18].

Thirty years after the prediction of radio transmission (1894) the Italian physicist and inventor Guglielmo Marconi was able to transmit a Morse signal over a distance of 2 km, another two years later he patented radio transmission [18].[4] In the early 1900s the Canadian engineer and inventor Reginald Aubrey Fessenden improved Guglielmo Marconi's work by making it possible to transmit speech through a radio channel. Fessenden presented radio's first program on Christmas Eve 1906 where he played Holy Night on his violin [19].

Almost parallel with the devising of radio transmission another important technology was developed; the wired telephone. It was invented by Alexander Graham Bell (1876) and constructed by Thomas Watson [5]. A wired telephone was reconstructed in 1910 by Swede Lars Magnus Ericsson who used it to make calls from his car. His wife Hilda threw two sticks over the telegraph lines and in this way they managed to connect with an operator who then could put through their call [22]. Nearly ten years later, in the

[3] The fax machine read text written in raised metal letters and transmitted it through a telegraph line. The telegraph relies on Samuel Morse idea from 1832 [27].

[4] For this he won the Nobel Prize in 1909 [25]. This was a prize he had to share with German Karl Ferdinand Braun.

early 1920s, the first police car radios and walkie-talkies were used in New York[5] [23].

Since a mobile phone converts the sound of a person's voice into radio frequency energy (radio waves), knowledge about radio transmission was necessary in the process of inventing the mobile phone [24].

2.1.2 Changing parameters

All the essential technologies to develop mobile phones, such as wireless telephony, cellular network design and frequency re-use, existed already in the late 1940s, all it needed was the software and hardware to make it work [26]. The transistor, invented in 1948[6], made it possible to build smaller, cheaper and lighter radios, which increased the number of radio sets in use.

In December 3, 1950 Sture Lauhrén made the world first cell phone call using a prototype system developed by Ericsson and The Swedish Telecom [23]. Six years later (1956) Ericsson introduced the first Swedish mobile phone; it was as large as a suite case and weighed 40 kg. During this time the telephone net only stretched out over an area of 25-30 km. The number of subscribers was few, about one hundred, since the equipment was very expensive[7] and outside the city demands was limited since only one fourth of the population owned a stationary phone [21].

In the 1980s the commercial mobile telephone took to the air and 1981 the first automatic mobile telephone net were inaugurated which stretched over a wider area and the numbers of subscribers soon reached 35000. Mobile phone units also evolved and became smaller. The new generation were the hand portable phones which weighed around 9 kg. Five years later, for a price at 30 000 SEK the first pocket able phone entered the scene. It weighed-in around 1 kg presenting a standby time of 6 hours and a talk time of 30 minutes.

During the last ten years development has really taken off, the mobile industry has gone from analogue to digital.[8] In 1994 Nokia 2110 appeared containing new functions such as SMS. At first SMS was just a way of sending text messages in complement to recording voice messages. It was not

[5] During this period the television was also demonstrated and the inventor Zworykin applied for a patent.
[6] This was the same year Claude Shannon published the Shannon-Hartley equation, which was important to understand the conditions for error-free communication [23].
[7] One phone cost as much as a car (ca. 7500 SEK) [21].
[8] 1992 the first digital mobile phone came [21].

predicted to become the success that it did. One reason for the success might be the slight paradox that the technique was difficult to use. For young people this meant that they had a way to communicate their parents and teachers were unlikely and unwilling to adopt. SMS gave birth to an entire new alphabet where abbreviations[9] and smileys were used to reduce the number of the keystrokes and convey feelings [20]. By 1997 the number of mobile telephone subscribers exceeded the stationary phone subscribers.

Both mobile telephony and computer systems appeared on the market in the 1950s and at that time they were difficult to interact with [30]. Computer systems were designed by specialists for specialists and not intended for the public. Pioneer users of the mobile telephony were wealthy businesspeople. Within both mobile telephone and computer development the target group has changed along with needs and technology.

Today most of the mobile phones fit in to the palm of a hand, palm phones, and weigh less than 60 grams [17]. Mobile phones are relatively cheap and the net is widespread but most important for this report; the incorporated features and functions has grown significantly in numbers. Today music devices, camera functions and various options to collect, send, and manipulate these are integrated in the mobile telephone unit. It also offers clock, timer, games, and the possibility to "surf" the net.

Today users face a world of technology they in some way are forced to use; and some which they are not forced to use but that surrounds our modern way of working and living. Technology such as ticket machines, library information systems, computers, stereo, video, mobile phones, digital cameras; the list could be very long [3]. Together the new functionality requires a new way of designing the telephone. In *Interaction Design beyond human-computer interaction* the authors' states how all these devices need interaction but very few of them are designed with a user in mind [3].

> Typically, they have been engineered as systems to perform set functions. While they may work effectively from an engineering perspective, it is often at the expense of how the system will be used by real people. The aim of interaction design is to redress this concern by bringing usability into the design process.

[9] See you later could be written "C U l8er" [20].

2.2 Development processes

For developers to be able to deliver within a predicted time and to produce a product with predetermined qualities, guidelines are crucial. There are several different abstract methods which all try to describe a certain approach to develop a product in a well structured way. In traditional software engineering the methods focus on delivering a product or system while Human Computer Interaction (HCI) models in some sense are developed to focus on the user of the product or system throughout the development.

2.2.1 Process models of software engineering

Software engineering concerns all aspects of software production. The term software engineering was first mentioned during a conference in 1968 according to Natt och Dag et al who quote Pressman with the following sentence [29].

> 'Software Engineering is the establishment and use of sound engineering principles in order to obtain economically software that is reliable and works efficiently on real machines (Pressman, 1992)'

During product development several facts influence the results. According to Ian Sommerville four important attributes characterizes a well-designed software system. These are [14]:

Maintainability
Software shall be produced in a way that it is possible to evolve to meet the changing needs of customers and business environment.

Dependability
Is divided into four sub dimensions, which all aims to prevent physical or economical damage if system fails.

Availability
Ability of the system to deliver services when requested.

Reliability
Ability of the system to deliver services as specified.

Safety
The ability of the system to operate without catastrophic failure.

Security
The ability of the system to protect itself against accidental or deliberate intrusion.

Efficiency
Software systems should not waste system resources. E.g. memory, processor cycles and responsiveness.
Usability
Software shall support the type of user for whom it is designed, it shall be usable without unjustified effort.

To obtain a product containing these factors the software engineering process should follow a certain set of activities, such as *software specification*, *software development*, *software validation* and finally *software evolution*. There are a number of different general models for software development, which describes when, and to some extent, how these activities should be executed [14] and the relationship between them [3]. In most literature[10] the *waterfall model*, *evolutionary development*, and the *spiral development* are mentioned, sometimes the term *lifecycle model* is used to describe the development process [3].

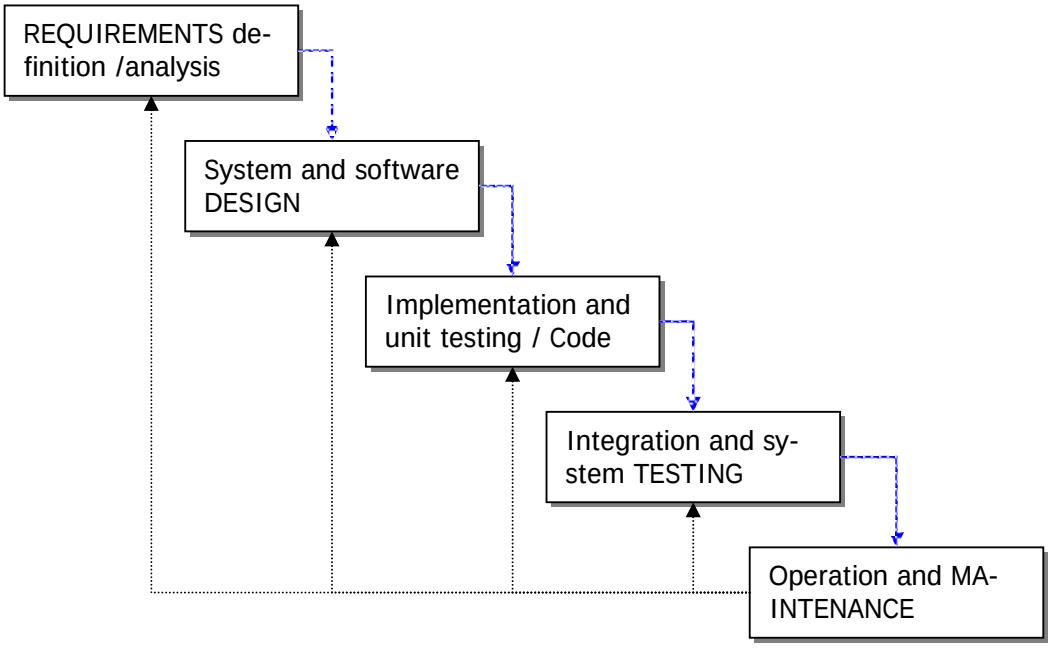

Figure 1. *The waterfall model* [14]. *Authors' interpretation.*

[10] Among others: Lauesen [4], Sommerville [14], and Preece-Rogers-Sharp [3].

2.2.1.1 The waterfall model

Depending on which literature studied, this model has small variations in layout and main heading but the basic principle is always the same. A project starts in one phase which has to be completed before next phase is entered. In theory it can be viewed like a linear model with a clear beginning and a clear ending in each phase [3].

Unfortunately reality is barely ever linear. Requirement change over time and it is necessary to be able to *iterate* through different stages. To limit costs and avoid entering a never ending loop, the development is frozen after a predefined number of iterations. It means that even though errors might be found during iteration they will have to be ignored [11].

2.2.1.2 Evolutionary development

This development process is sometimes called prototyping. The process starts with an initial implementation which is then exposed to users and continuously refined through several versions until an acceptable system has been developed. Compared to the waterfall model it has a rapid feedback across the activities, since specification, development and validation is carried out concurrently.

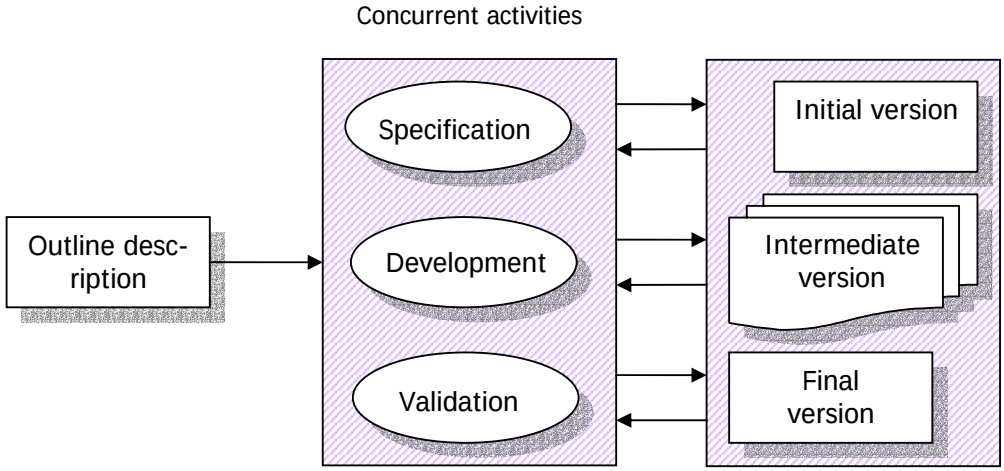

Figure 2. *Evolutionary development* [14]. *Authors' interpretation.*

2.2.1.3 Spiral development

This model was suggested in 1988 by Barry Boehm [3]. It differs from the other two models by its graphical spiral representation of the phases. Also, the spiral development model focuses on the risks involved in developing the system rather than the intended functionality. Development starts in the core of the spiral and than proceeds clock-wise. Each new loop in the spiral describes a new phase in the project and it may have different activities. For example, the innermost orbit could represent system feasibility followed by system requirements and then system design, etc. [14]. Each loop is divided into four sections; planning, risk analysis, development and customer evaluation.

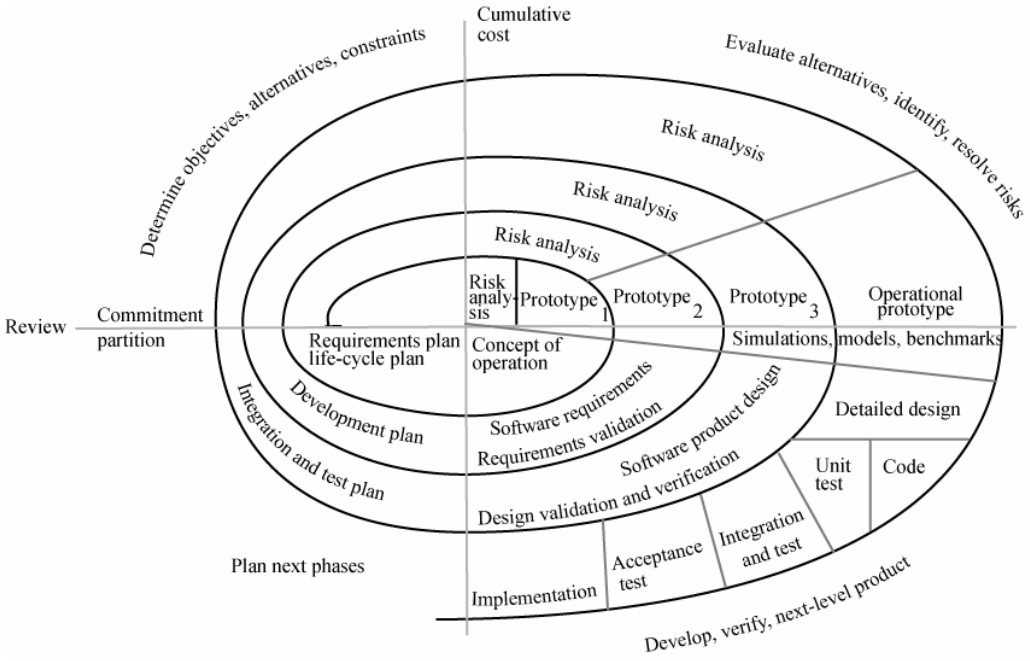

Figure 3. *Spiral development* [14]*. Authors' interpretation.*

Johan Natt och Dag and Ofelia S. Madsen have in their master thesis *An Industrial Case Study of Usability Evaluation* [29] put together an easy to grasp table with advantages and warnings for several software development process models. Following facts are mentioned concerning the waterfall model, evolutionary development and spiral development.

	Advantages	Warnings
The Waterfall Model	• Easy to understand • Easy to adapt to • Makes the development process more visible • Fertile • Perfect for individuals	• Does not support parallel activities • Does not support reuse well • No user involvement throughout the process • Does not support developers need for quick results • Heavy documentation preparation burden the developer
Evolutionary Development	• Allows rapid development and delivery • Reduced development cost • User involvement endorsed • Both users and developers more positive to the resulting product	• Documentation might be disregarded since it is not cost-effective in the quickly evolving prototypes • Can give a corruptive system structure • Might require special skills and well motivated developers • Does not make the product easily maintainable • Quick-and-dirty methods can make its way to the final system • Common to compromise on aspects or features of the product • Poor process visibility
Spiral Development	• Highly flexible • Minimises risks • Repeated normalized activities • Cost effective	• Hard to determine objectives, constraints and alternatives • Demands and relies on high risk expertise

Table 1. Advantages and warnings for waterfall model, evolutionary development, and spiral development. Copied with permission from Natt och Dag et al [29].

Of course there are several other models that can be used to describe the set of activities in the software engineering process. The reason to present these selected few is to supply the reader with an overview of the context

and show that requirement, design and validation are often closely related to each other.

2.2.2 HCI and user-centered models

Human Computer Interaction or HCI as it abbreviates, derive from several fields such as computer graphics, operating systems, human factors, ergonomics, industrial engineering, cognitive psychology, and the systems part of computer science [13]. Just as in the field of software engineering, lifecycle/design models has arosed from the field of HCI. They are fewer and unlike software engineering models they have a clearer user focus.

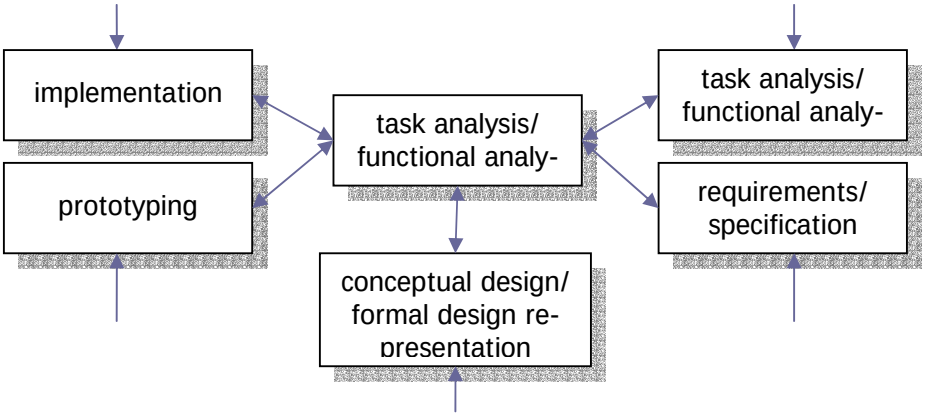

Figure 4. *The star lifecycle model. Authors' interpretation from Precce et al* [3].

2.2.2.1 Star model

The star model was proposed by Hartson and Hix (1989) and was the first HCI model to apear and it was an alternative way to support the design of interfaces. The model was the result of empirical observation of how interface designers works. Mainly the work could be devided into two different activitys; analytic mode and synthetic mode. Peerce et al describes the modes with the words [3]:

> The former is caracterized by such notions as top-down, organizing, judical, and formal, working from the systems view thowards the user´s view; the latter is characterized by such notions as bottom-up, free-thinkting, creative and *ad hoc*, working from the user´s view towards the systems view.

All together the star model represents a very flexible process with evaluation as a centrum focus. Without any specification about what order the activites shall be performed.

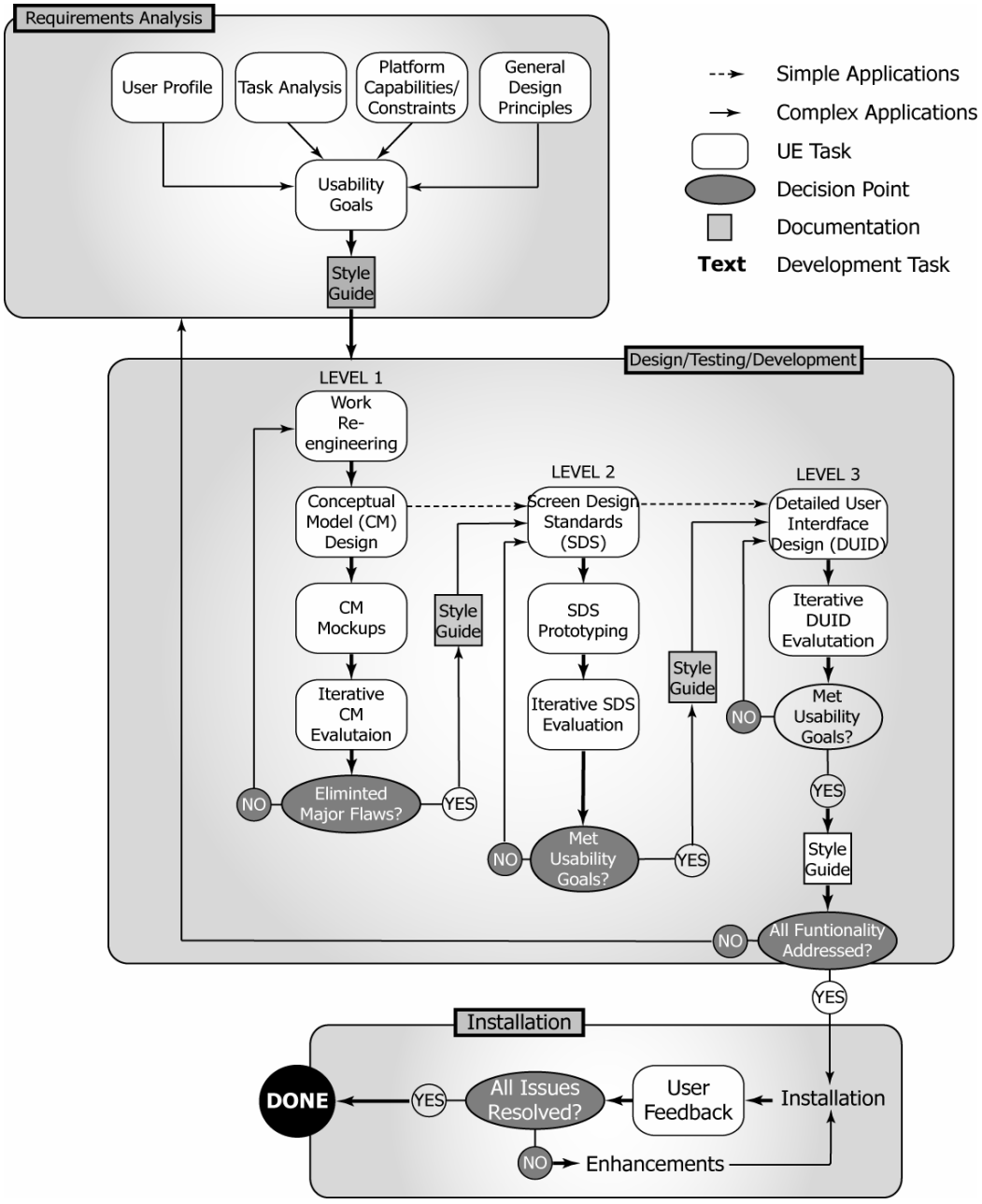

Figure 5. *Usability engineering lifecycle model* [2]. *Authors' interpretation.*

2.2.2.2 Usability engineering lifecycle model

Another model called usability engineering lifecycle model, has a more straight-forward structure. It was proposed by Deborah Mayhew in 1999 who clamed that the usability engineering tasks included was not a reinvention, according to Mayhew she just gave a holistic view over usability engineering and a description of how to perform usability tasks. The model describes how traditional software development can be integrated with usability tasks and contains three main sections; requirements analysis, design/testing/development and installation [3].

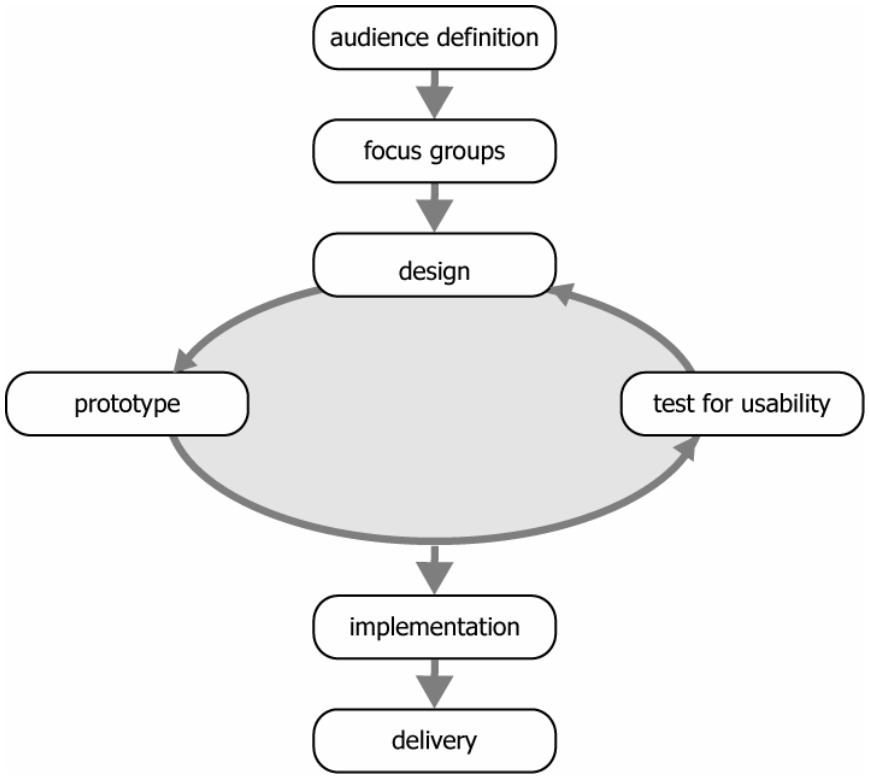

Figure 6. *Design cycle for handheld usability* [5]**.** *Authors' interpretation.*

2.2.2.3 Design cycle for handheld usability

There exists several different variants of design models, one is the design cycle for handheld usability presented by Scott Weiss in the book *Handheld usability* [5]. It describes a process for design and development beginning with audience definition. Iteration is entered in the design phase and sonsists of designing, prototyping, and usability testing. Once the design is

finished it is implemented and delivered. This model is mentioned in this thesis for the reason that it is one of the design models focusing on mobile, or at least, handheld usability and development [5].

2.3 Requirement

One of the phases in software processing is to define and analyze requirements. In *Software Requirements Styles and Techniques* [4], Søren Lauesen portray the circumstances in the software industry in 1962. He explains that software requirement were relatively unimportant because software at this time was cheap and hardware comparatively expensive.[11] Following quotation from Lauesen shows how different the approach was at that time [4].

> Software development was carried out either on a time and materials basis, or as a small part of the really important job – making better hardware. The customer paid until he had a program that printed results he could use with some effort. Nobody thought of usability. Everything to do with computers was a specialist's job.

Lauesen claims that the converse relation is in effect on the market today with cheap hardware and expensive software. He also adds one important factor; it is hard to keep within a budget. Lauesen explains that to effectively meet a customer's demands, software requirements are necessary.

One of the phases in software processing is to define and analyze requirements. There are different types of requirements to consider.

2.3.1 Functional requirements

Explains how the system should respond and react to different input and in different situations. It is simply what the system should do or not do [14]. Functional requirements clarify what data should be used and how to, for example, compute, update, store, transmit and transfer these data [4].

There exist a number of different techniques for identifying requirements. To understand the domain it is helpful to use a context diagram. A context diagram is a diagram of the product and its surroundings and it shows the possibilities of the product. Another way to identify requirements is to write event or function lists. They explain what functions the product respective the human and the computer shall handle or can simply just be a list of

[11] "Renting a computer for an hour cost the same as paying someone to work for 30 hours and computers were 5000 times slower than they are today." [4].

user tasks. Other ways could be to use data flows, or scenarios, the enumeration can be long.

Software requirement is not only a matter of distinguishing which requirement that are necessary for a system, it is also about present, validate, and verify requirements. There are several different styles to present requirements, for example with diagrams, plain text or structured text [4]. To obtain a satisfying product functional requirement can not cover all demands therefore another type of classification is needed, non- functional requirements.

2.3.2 Non-functional or qualitative requirements

Non-functional requirements has a sound of being unimportant and it is easy to catch the impression that it treat requirement that don't function therefore term *qualitative* requirements is introduced and used in *Software requirement styles and techniques* by Søren Lausen [4]. Qualitative requirements explain how well systems perform and they can be categorised in quality factors. Lausen presents three standardizations, McCall and Matsumoto, ISO 9126[12], and IEEE 830[13]. They are recommended to be used as checklists when quality factors are to be distinguee. According to the standardization by McCall and Matsumoto following separation is useful[14].

[12] Standard defined by International Organization for Standardization.
[13] Standard defined by the IEEE Standards Association.
[14] The definitions of the quality factors are direct quotation from Lauesen [4].

Quality factors

Operation Daily use by end users	Revision Maintenance and extension of the software	Transition Use of the software in new technical situations
Integrity (security) How well the system handles physical disturbances?	**Maintainability** How easy it is to locate and repair errors?	**Portability** How easy it is to move the system to new software after a change?
Correctness How many errors are there in the system?	**Testability** How easy it is to test the system after	**Interoperability** How easy it is for the system to cooperate with other systems?
Reliability How frequently the system malfunctions and the percentage of time it is available?	**Flexibility** How easy it is to expand the system with new features?	**Reusability** How easy it is to reuse parts of the software in other systems?
Usability How easy it is to learn the system? How efficient it is for carrying out day to day tasks?		
Efficiency (Performance) How fast the system responds, how many resources it uses, how accurately it computes values		

*Table **2.** Overview of quality factors according to McCall and Matsumoto* [4].

2.4 Usability

Usability can be an abstract concept and several figures are trying to provide the adopter a conceptual understanding. One of them is the usability triangle. It shows how user, task and artifact must interact with each other to provide for a good system.

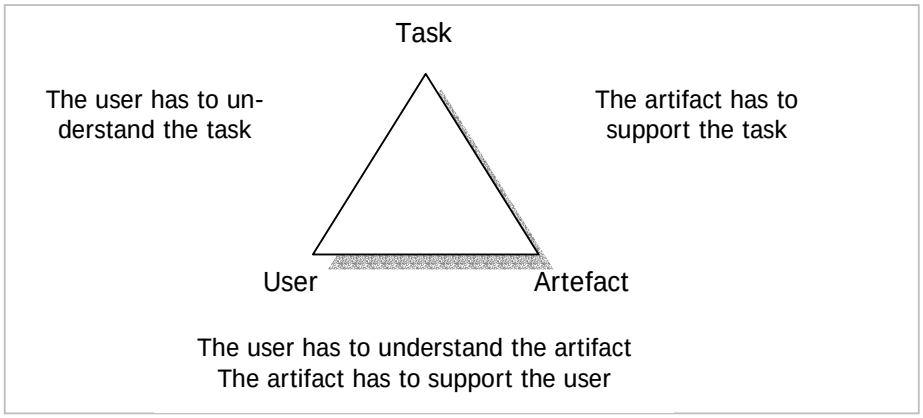

Figure 7. The usability triangle [31]*. Authors' interpretation with inspiration from Caroline Olsson* [32]*.*

Today[15] there is still no accepted standard for the concept of usability and the definitions vary from source to source. One way is presented below.

A multidimensional attribute that relates to the extent to which a product or service facilitates the goals of end users. In general usability refers to the efficiency with which customers can accomplish their tasks with the product/service and the overall satisfaction of users [12].

According to *User-Centered Design in the industry – a survey within usability* [11], a thesis written on this subject, Brian Shackel is regarded as the one who introduced those concepts associated with the attributes of usability. Izdebski and Johnsson present a table with an overview of definitions of usability attributes from four important sources. The first is Shackels own definition, second and third two other eminent characters within usability; Löwgren and Nielsen and finally ISO:s definition.

[15] Spring 2005

	Shakel	Löwgren	Nielsen	ISO
Usability attributes	Efficiency	Efficiency	Efficiency	Efficiency
	Attitude	Attitude	Satisfaction	Satisfaction
	Learnability	Learnability	Learnability	
			Memorability	
		Relevance		Appropriated
			Extent of error	

Table 3. Free interpretation and translation of definitions of usability attributes [11].

Efficiency

The attribute occur within all four definitions. The term efficiency is also used as a quality parameter, but the difference is that efficiency as an usability attribute refer to how efficiently the users can carry out their tasks using the system, while the quality parameter refer to how efficient the system can carry out its tasks.

Attitude / Satisfaction

Both attributes describes the same thing and occur within all four definitions. It comprises the user subjective feelings towards the system [11].

Learnability

Is common for three of the definitions, ISO does not have anything corresponding. It deals with how easy the system is to learn for initial use and how well users will remember the skills over time. Nielsen has an attribute, memorability, separately for how well the user remembers the skills over time. Nielsen separates them since it exist things that take time to learn, but once you have learned them you don't forget them easily. (Take for example to learn how to ride a bike.)

Relevance / Appropriated

Deals with how well the system serves uses needs.

Extent of error

This attribute is unique for Nielsen and measure if catastrophic errors occur often or not it also deals with how easy it is to make errors [11].

In some literature[16] the usability attributes is called usability goals and is broken down into following goals:

- Effective to use (effectiveness)
- Efficient to use (efficiency)
- Safe to use (safety)
- Have good utility (utility)
- Easy to learn (learnability)
- Easy to remember (memorability)

However, with the increasing complexity of new technologies and new functions within already existing technologies Preece et al explains that usability goals is not enough. It is also important to concern for the users experience – what the system feels like to the users. Usability goals and experience goals can be represented by following figure.

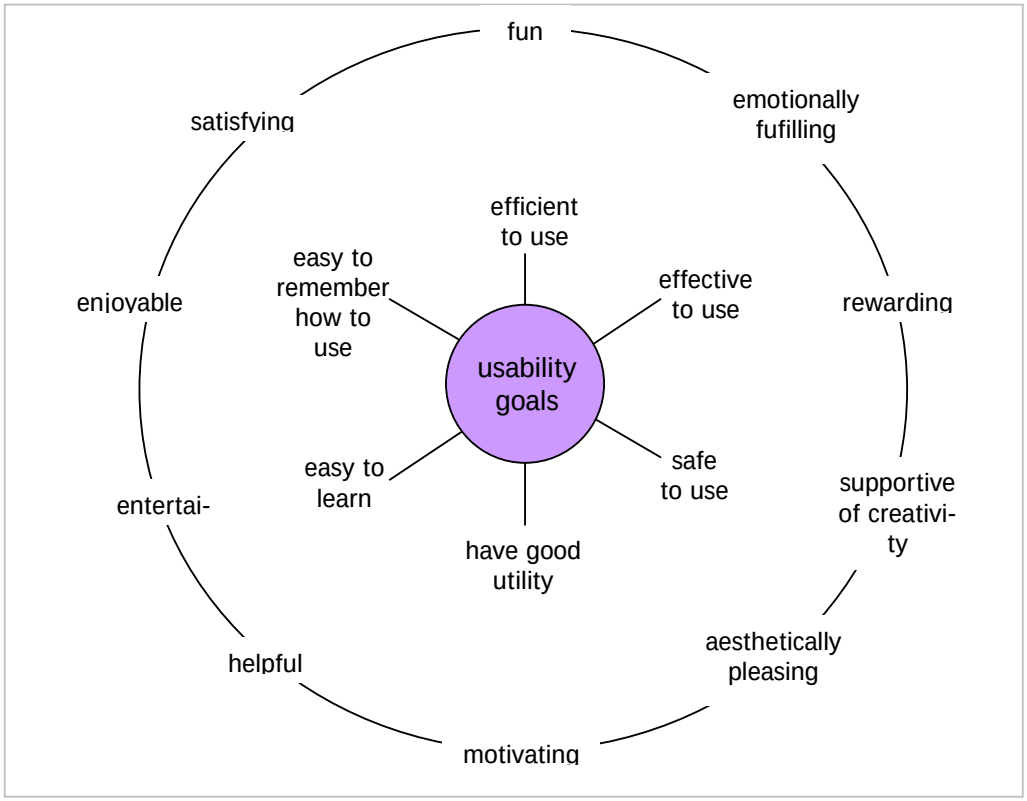

Figure 8. *Usability and user experience goals* [3]. *Authors' interpretation*.

[16] Preece-Rogers-Sharp [3].

2.5 Design

The discipline of user interface (UI) design has a very wide span; from interdisciplinary fields as information systems, cognitive engineering and human factors to graphic design, information architecture and ergonomics. Closest to the design of the final product are the information architect from the field of information architecture and the graphic designer, working in the field of graphic design.

The role of the information architect is to plan and structure information, this includes defining the content of the interface and defining the task structure. Specifically content in menus and their hierarchical structure, which menu contains what choices, are important design matters [5].

To make the information on the display visible and appealing to the user a graphic interface is applied to the framework of the information architect. Designing this is the work of a graphic designer. Studies imply that appealing interfaces increases usability to some degree [10]. Positive affect of an interface stimulates creative thinking and can make difficult tasks more easily solved. However, a fancy skin does not cover for bad information architecture or usability flaws, all these disciplines must work together to achieve a positive result.

Most studies and literature in the field focus on traditional desktop computer user interfaces. However, designing for desktop computers and designing for handheld devices, such as the mobile phone, differ in many aspects. Most obvious difference is the size of the screen [6]; a mobile phone screen is significantly smaller than a desktop screen and is therefore more easily cluttered up by layers and multiple type windows, hence become stressful and chaotic to look at [1]. Techniques for displaying data in new ways better suited for the small area are being developed[17] but the evolution of the mobile phone screens has gone incredibly fast, just three years ago the future of colour displays was doubtful [7], and the new techniques take time to develop and perfect. Another apparent difference to desktop computers is that a vast majority of today's mobile phones does not support free navigation style like using the mouse with a desktop computer. Instead a few keys on the phone's hardware interface are used to navigate step-by-step through and interact with the menus and elements on the display [6].

Limitations like the ones mentioned above and like restricted memory and storage capabilities force interface designers to design the user interface to

[17] See an example at buddybuzz. http://www.buddybuzz.net/rel/Web/index.html.

negotiate these limitations and provide a certain degree of usability [5]. In contrast to computers which have extensive usability research and years of experience to support development, the mobile phone usability is still a new concept struggling with standardisation issues and an abundance of competing developers, all with economic interests. Furthermore software developers are trying to make their product hardware independent which basically means developing for a wide range of hardware platforms, varying key sets, screen sizes, and memory capabilities [9].

There is also a slight difference to how the software is used on a mobile unit such as the mobile phone compared to the desktop computer. An illustrative example is the Internet access provided by both modern mobile phones and desktop computers. Usage on a desktop computer is best described as browsing; several pages can be open and compared, scanning web pages for interesting content is common behaviour, and much time can be spent reading articles or, for that matter, reading e-mail. When utilizing the Internet service provided in a mobile phone, information is hunted for and websites are visited with a purpose and usually not just for browsing and entertainment. This behaviour is most likely a result of the not so rewarding user experience along with per-minute charges [5].

Understanding for and definition of the problem space[18] supports interaction designers when drawing the outlines of the design or solving a known problem of an existing design. Understanding the problem space involves conceptualizing what to create and why.

Developers of mobile phone software, like any software developer, are under heavy pressure distinguish a company's products by producing applications with more features and flashier graphics in decreasingly shorter time [9]. This creeping featurism [1], more features and applications "under the hood", makes the work of the informational architect even harder since more functionality needs to be categorised and stowed away in some menu. More features cause more things to be invisible inside the application.

Decreasing time intervals between releases do not only potentially affect the quality of the software but also interfere with the natural design process which relies on evolution, improvements to the original design based on feedback from users [1]. Short time intervals between releases means that new versions of the product are already in development when a release is

[18] Problem space concerns the problems the design are to solve including limitations such as physical and cultural [5].

made and no time for feedback is allowed, why designers need to rely on usability tests during development.

2.6 Guiding principles

When designing for mobile phone platforms some of the principles developed during the years devoted to desktop computers can be adapted but in some cases need modification to fit the specific platform. Many of the existing principles are based on cognitive science and address how users perceive information and react to different environment variables. Both guidelines for software design and usability evaluation may be considered during the design process since the software ultimately must conform to the usability evaluation principles when an evaluation is performed and conversely design principles may be considered during usability evaluation [3].

2.6.1 Eight golden rules of interface design

Ben Shneiderman proposes a set of principles derived heuristically from years of experience in the field. The principles focus on the dialogue between user and system and increasing user control and comprehension of the system. Each principle has to be interpreted and extended to apply to the context of the mobile phone interface[19].

1. *Strive for consistency.*
   Consistent sequences of actions should be required in similar situations; identical terminology should be used in prompts, menus, and help screens; and consistent commands should be employed throughout.

2. *Enable frequent users to use shortcuts*
   As the frequency of use increases, so do the user's desires to reduce the number of interactions and to increase the pace of interaction. Abbreviations, function keys, hidden commands, and macro facilities are very helpful to an expert user.

3. *Offer informative feedback*
   For every operator action, there should be some system feedback. For frequent and minor actions, the response can be modest, while for infrequent and major actions, the response should be more substantial.

[19] Direct quotation from http://www.cs.utexas.edu/users/almstrum/cs370/elvisino/rules.html

4. Design dialogue to yield closure

Sequences of actions should be organized into groups with a clear beginning, middle, and end. The informative feedback at the completion of a group of actions gives the operators the satisfaction of accomplishment, a sense of relief, the signal to drop contingency plans and options from their minds, and an indication that the way is clear to prepare for the next group of actions.

5. Offer error prevention and simple error handling

As much as possible, design the system so the user cannot make a serious error. If an error is made, the system should be able to detect the error and offer simple, comprehensible mechanisms for handling the error.

6. *Permit easy reversal of actions*

This feature relieves anxiety, since the user knows that errors can be undone; it thus encourages exploration of unfamiliar options. The units of reversibility may be a single action, a data entry, or a complete group of actions.

7. Support internal locus of control

Experienced operators strongly desire the sense that they are in charge of the system and that the system responds to their actions. Design the system to make users the initiators of actions rather than the responders.

8. Reduce short-term memory load

The limitation of human information processing in short-term memory requires that displays be kept simple, multiple page displays be consolidated, window-motion frequency be reduced, and sufficient training time be allotted for codes, mnemonics, and sequences of actions.

2.6.2 Heuristic guidelines

Usability principles are mainly aimed for evaluating prototypes and existing systems and provide a basis for heuristic evaluation and usability testing. Nielsen's guidelines are considered more as a rule of thumb than actual design guidelines. However, some of the guidelines overlap and combine some of the eight golden rules presented above.

1. *Visibility of system status.*
   Always keep users informed about what is going on, through providing appropriate feedback within reasonable time.

2. *Match between system and real world.*
   Speak the users' language, using word, phrases and concepts familiar to the user, rather than system-oriented terms.

3. *User control and freedom.*
   Provide ways of allowing users to easily escape from places they unexpectedly find themselves, by using clearly marked 'emergency exits'.

4. *Consistency and standards.*
   Avoid making the users wonder whether different words, situations, or actions mean the same thing.

5. *Help users recognize, diagnose and recover from errors.*
   Use plain language to describe the nature of the problem and suggest a way of solving it.

6. *Error prevention.*
   Where possible prevent error occurring in the first place.

7. *Recognition rather than recall.*
   Make objects, actions, and options visible.

8. *Flexibility and efficiency of use.*
   Provide accelerators that are invisible to novice users, bur allow more experienced users to carry out tasks more quickly.

9. *Aesthetic and minimalist design.*
   Avoid using information that is irrelevant or rarely needed.

10. *Help and documentation.*
    Provide information that can be easily searched and provides help in a set of concrete steps that can easily be followed.

2.6.3 Cognitive interpretation

A more elaborate description of the most common design principles is provided in the book *Design of everyday things* by Donald A. Norman [1].

Mapping

Mapping is a technical term in this case referring to the relationship between two things; the controls and their movements and the result on the display. In all design, extensive utilization of natural mappings, consisting of physical analogies and cultural standards, is recommended.

Visibility

Good visibility is attained by making things visible. This means keeping functionality and controls evident. The consequence of good visibility is meaningful and sensible relationships among the user's intentions, actions, and results. Visibility is closely related to feedback and mapping. Good mapping tends to support visibility.

Feedback

Feedback is a well known concept in the science of control and information theory. Supplying the user with appropriate and instant information about what action has been done and what result has been accomplished is crucial when it comes to usability. There are different kinds of feedback such as tactile, audible and visual, all of which may be combined in a suitable manner. Today's advanced displays enable usage of high-quality visual feedback.

Constraints

Constraints are used to guide the user to what to do and what can not be done. A common design practice in graphical user interfaces is to deactivate certain menu options by shading them. This makes the menu option still visible to the user but not selectable.

Affordance

The term affordance refers to the perceived and actual properties of an object. What an object looks like determines or gives clues to how it can be manipulated. Affordances in user interfaces best utilizes perceived affordances which are essentially learned conventions of virtual nature.

Consistency

Similar actions and elements should be used when achieving similar tasks. The concept of consistency applies both to appearance of the interface and interaction such as input and navigation. Consistent interfaces are easier to learn and use since only a single mode of operation has to be learned.

2.7 Evaluation

Evaluation of usability is an important step for developers to get feedback and to measure how well the project meets the set usability goals or requirements. There are several techniques and paradigms to choose from, all with some benefits and drawbacks. In *Interaction design - beyond human computer interaction* [3] the authors categorises the techniques into four core paradigms.

2.7.1 Quick and dirty

Emphases on quick input rather than formally correct and documented findings. Developers receive informal feedback from users to confirm that their ideas are in line with the users' needs. This kind of feedback is considered essential to achieve a successful design.

2.7.2 Field studies

The distinguishing feature of field studies is that they are done in natural settings. Mainly aimed to increase the knowledge of how users act naturally and how technology impacts them. Qualitative techniques for recording data, such as interviews and observation, can be used.

2.7.3 Predictive evaluation

Commonly also referred to as *heuristic evaluation*. Key feature is that no users need to be present. Instead evaluation is entirely carried out by experts, utilizing heuristics and their knowledge of the typical user. The technique is rather common, probably due to its quick and inexpensive nature, but has limitations. Typically many problems are identified but certainly not all and sometimes even ones that are not a real problem.

2.7.4 Usability test

During a usability test typical users' are employed to carry out tasks typical for the software tested or tasks of specific interest. Their performance is measured qualitatively, by questionnaires, and quantitatively, by number of errors. The evaluator has strict control of the test. Usability testing is commonly considered to be the same thing as usability evaluation. There is however a distinction between them; testing is more accurately a technique, among others, of usability evaluation.

When comparing heuristic evaluation to testing; usability testing generally attains better results or a better hit ration of real problems. Combining the two techniques should prove the best results.

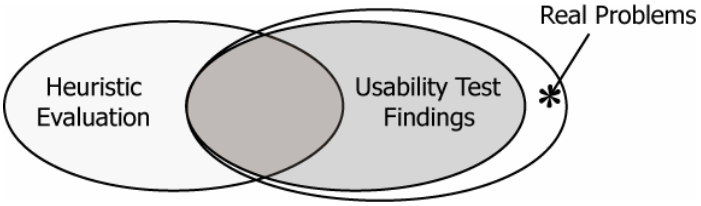

Figure 9. Heuristic evaluation compared to usability testing for finding the real usability problems [4]. *Authors' interpretation.*

2.8 Usability test

To perform an usability test has proven to be the most effective way to identify usability problems and flaws. There are several various methods to design a test to fit a specific purpose, often using different terms to describe identical techniques. *Handbook of Usability Testing* [2] focuses on four types of tests and associates them to phases in the development life-cycle. The types of test described are exploratory test, assessment test, validation test, and comparison test. Each of these tests has slightly different goals.

2.8.1 Exploratory test

Exploratory tests focuses on examining the effectiveness of preliminary design concepts. Developers and designers can gain a lot from understanding whether the user grasps the general design concept and the presented conceptual model. Typically the exploratory test is carried out on a mock-up or prototype early in the development process.

Since very much of the desired results are of cognitive nature an exploration of the user's thought process is necessary. In an exploratory test this is achieved by having the user and test monitor exploring the design together, emphasising on discussion.

2.8.2 Assessment test

An assessment test can be described as evaluating how effectively the concept of choice from the exploratory test has been implemented. Users perform realistic tasks and their performance is measured.

Assessment tests are typically carried out when the fundamental or high-level design has been established.

2.8.3 Validation test

A validation test aims to prior to release determine how the developed product compares to set usability requirements or benchmarks. It also

evaluates how well all parts of a product work together, including documentation and help. The test procedure is similar to that of the assessment test but require more rigor and consistency since performance is measured against a standard.

The validation test is carried out quite late in the development lifecycle, close to release of the product.

2.8.4 Comparison test

As the name suggests comparison tests are used to compare alternative designs. Comparison can be done to a competitor's product, to an alternative interface design, or a proposed design to another.

Comparison tests are not associated with a specific point in the development lifecycle it can be done whenever it is considered necessary.

2.8.5 Usability test according to Jeffery Rubin

Rubin states that an usability test is "a process that employs participants who are representative of the target population to evaluate the degree to which a product meets specific usability criteria." [2]. The criterion[20] he writes about is similar to the ones mentioned in section 2.4, usability.

Rubin also explains that testing is always an artificial situation; the sterile laboratory environment and knowledge about that it is just a test can affect the results. Further Rubin express some other issues with testing. He claims that the test results can not *prove* that a product work. Participants are a representation of the end users and it is only a small selection of the target population. The real or actual users can be difficult to describe and identify. Finally testing is not always the best technique to use, sometimes other methods can be more effective, but at the same time Rubin assert that "it is better to test than not to test" [2]. Jeffery Rubin believes it is important to remember that an usability test is not the same as a classical experiment. The following table describes issues with a classical experiment and Rubin's argument against such a strict approach.

[20] These factors are to insure products that: are satisfying to use, are easy to learn and to use, provide utility and functionality that are highly valued by the target population [2].

	Classical experiment	Rubin's argument
Problem statements:	A hypothesis must be formulated	Usability is rather about improving a product than formulate and test specific hypothesis.
Collected data:	Is used to collect quantitative data to obtain a proof of a research hypothesis. For example when comparing two designs.	Is used rather to obtain qualitative information on how to correct problems and redesign products.
Target group:	Choose participants randomly using some systematic method.	It is difficult to select participants that satisfy the demands on randomly statistical correct chosen test persons since this factor is hard to control. Sometimes the product is classified as secret and can not be exposed to outsiders. Perhaps the system to be tested is only used by a few users.
Dimensions of test:	The sample of users must be of sufficient size to measure statistically significance between groups.	To obtain generalized results the test needs 10 to 12 participants per conduction. This will require 40 or more test participants to ensure statistical significant results, which is a reason for a less formal approach[21].

Table 4. Classical experiment versus Rubin's argument aganst such a strict approach.

[21] The number of test persons is recommended by Rubin to be at least four while Weiss recommend six with the motivation that it is possible to conduct six one-hours test during the same day and six tests is enough to identify trends.

Rubin rather recommend following approach, summarised in six items [2].

1. Development of problem statements of test objectives rather than hypothesis.
2. Use a representative sample of end users which may or may not be randomly chosen.
3. Representation of the actual work environment.
4. Observation of end users who either use or review a representation of the product. Controlled and sometimes extensive interrogation and probing of the participants by the test monitor.
5. Collection of quantitative and qualitative performance and preference measures.
6. Recommendation of improvement of the design of the product.

2.9 Software prototyping

Software prototyping is a way to visualise an unfinished or concept of a product. By using a prototype it is easier for a user to explain expectations for the new product and problems of the existing product. A prototype can be a paper-based outline of an interface or an interactive. The prototype provides a way to communicate an envisioned product to stakeholders and allow them to gain experience of realistic usage and explore imagined features. Furthermore, prototyping support designers in choosing between alternatives [3].

A simple and cheap way to do prototyping is by *low-fidelity* (lo-fi) prototyping. These prototypes are quickly produced and hence easily modified. The lo-fi prototype does not have to look very much like the final product and it can be built from very different materials. Sketches and storyboards are forms of lo-fi prototypes.

On the other end of the prototyping scale is the *high-fidelity prototype* (hi-fi prototype). In these prototypes materials are used that can be expected to be found in the finished product; hence it looks more like the final product. Hi-fi prototypes take longer to build and are not easily modified. Furthermore, its finished look may give too high expectations on its functionality why users testing it could feel disappointed.

2.10 Data collection

Data collection is about gathering information. Sometimes used to determine a choice of stakeholders' opinions in a project, or bring together various test persons background and skills. There are different techniques suit-

able for different situations, depending on what data needed to be collected and what phase of development presently in [4].

Interviewing
A set of questions is asked orally either face to face or on a medium like phone to elicit, among many things, present work and problems.

Observation
Observe users in their daily work to improve knowledge about current work and work around the problem with users not always aware of what they really do.

Task Demonstration
Combining interviewing with observation by having a user carrying out a specific task. This is usually easier than explaining how it is done.

Document studies
Study existing documents such as forms, letter files, computer logs and documentation of existing system to cross-check interview information and get information of how things use to work.

Questionnaires
Distribute questionnaires to collect opinions and suggestions from a large amount of users or to get statistical evidence for an assumption.

Brainstorming
Gather a group of people in a stimulating environment to come up with ideas for goals and requirements for the new system. Ideas are later prioritised and some discarded.

Focus groups
Resemble brainstorming but with more structure. Several groups of stakeholders participate.

Domain workshops
Map the business process and produce some sort of task description or activity diagram that describes what goes on in the domain and can later be turned into requirements.

Design workshops
Users and developers co-operate to design typically the user interface. Make sure the design is regularly checked against the task description.

Prototyping
Developers experiment with the prototype to get an idea of how it would work in real life. This technique may result in product-level requirements as well as design-level requirements.

Pilot experiments
A small part of the organization tries the new system on a trial basis to enable the project team to evaluate the cost and benefit of it.

Study similar companies
Look at what other companies do to get realistic ideas of how to handle similar problems. Internal auditing and consultancy companies can provide performance measurements in many fields.

Ask suppliers
Suppliers of considered products can be an important source of ideas for new solutions. Can supply features lists and refer to other customers.

Negotiation
Resolve conflicts between stakeholders or between customer and supplier. Goals for each party may conflict and solutions to satisfy both parties results in better requirements.

Risk analysis
To identify risky areas of the project and find ways to reduce the risks possible consequences are considered. This can be done by working with stakeholders in interviews or workshops.

Cost / benefit analysis
Examine factors as changed revenue, changed costs, product costs, employee satisfaction customer satisfaction and decision quality to compare the cost of the project and the benefit from it.

Goal-domain analysis
An analysis that address the relation between goals and tasks so that important goals are not forgotten and features do not lack a goal.

Domain-requirement analysis
Similar to goal-domain analysis but at a lower level.

Choosing the appropriate technique for a specific development project is a question of which serves the purpose best as well as which is most cost efficient. Several techniques can be used in parallel to collect both qualitative and quantitative data. Typically defined data is called quantitative if it is in numerical form and qualitative if it is not [8]. More specifically qualitative data claims to be contextual, nuanced, and sensitive while quantitative is rigorous, credible, and scientific.

3. Method

To describe the approach for this thesis the work is divided into analysis, design, test and report. One of the first things done in this master thesis was an initial time plan for the five months work ahead. It contained the information known at the moment and time buffers where coming tasks were added. Naturally new mile-stones were added as they appeared and the time plan was continually updated each week.

3.1 Analysis

It was important to understand the demands from different stakeholders and the purpose for this thesis. Once it was clear that a test plan was to be composed, old experience from writing a test plan was combined with a review of other students' work in the same field. During the analysis process objectives were stated according to Goobi ABs requests and the software to be used was explored.

Furthermore a representative group of users was selected for the test. Since Goobi AB consider all mobile telephone users as potential end-users of their product, test users could be selected among known contacts. Finally, questions for briefing and debriefing were chosen. For the first test they were determined by help of other students' work in the same field and for the second test with inspiration gathered from the book *Handheld usability* [5] and an article by Harri Kiljander [6].

3.2 Design

In this phase all necessary information was collected and was summed up in a test plan closely following the structure suggested in *Handbook in usability testing* [2]. The work also contained design of questionnaires and formulation of scenarios for the test. For the first test this was done in an arbitrary fashion, but since the outcome was not the desired it was extensively redesigned for the second test, with help from Arne Svensk, Magnus Haake, and Jonas Borell, all active at the department of design science.

3.3 Test

Two tests were carried out on the software suite. The software is developed for deployment in mobile phones based on the Symbian OS. Its look and feel can best be described as similar to the software found in many of today's mobile phones. However the software is presented on a computer using a simulator of an imaginary mobile phone which is interacted with through mouse and keyboard.

The tests were performed in IKDC's test laboratory in Lund according to the written test plans, see Test report 1 and Test report 2. The two tests had different goals and variations in how they were conducted according to *Table 5*.

	Test 1	**Test 2**
Variations	Test participants were divided into groups depending on gender and experience in modern technology.	Test participants were divided into groups depending on test order, and interaction with test monitor.
Goals	To find usability flaws as well as test persons' attitude towards the method. Evaluate test method.	Examine which device most preferable among the test persons, test monitors position, and test persons' attitude towards method and environment.

Table 5. Comparison of goals and variations in tests 1 and 2. For further details see Test report 1 and Test report 2.

All together 37 tests were carried out; test 1 used 20 test persons and 2 pilot tests while test 2 used 12 test persons and 3 pilot tests. Each test went on for about one hour. Test one was followed up by a summation of the main flaws observed during the test. Seven tapes were studied task by task and observations were gathered and presented in an intermediate report. The second test followed closely in time with the first and therefore a full review of tapes from test one was done only after the review of tapes from test two.

3.4 Report

Mainly this thesis had three deliverables; test plan for test one, test plan for test two and final report, but along the way an intermediate report and an oral presentation of intermediate results were requested. To put the tests into context a literary study of already known sources was performed in the end of the thesis, possible drawbacks from this are discussed in the chapter Sources of error, in parallel with writing results and finishing the review of test one.

3.5 Time estimation

All together 37 tests were conducted and each test took about one hour. Seven tapes from test one was reviewed for the intermediate report and had to be studied over again for the second review which included all tapes, each review took two hours. To complete missing information all tapes from test one was reviewed a second time and some even three times, which took about one hour per tape. Approximately it took 37 hours to conduct the tests and between 86 and 110 hours to review the tapes
To facilitate the description of the method, *Figure 10* explains the disposition of this thesis.

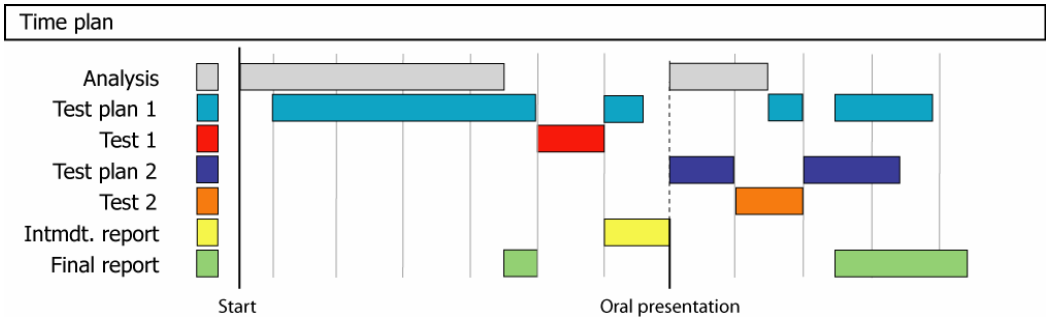

Figure 10. Time plan disposition.

4. Results and conclusions

The purpose of this thesis was to examine the possibilities for Goobi AB to use IKDC's test laboratory for usability tests in the future. The purpose can be divided into following main points:

1. Conduct a usability test on three of Goobi AB's products and report flaws in the logical flow.

Address the following questions

2. How many test persons:
    a) needs to perform the test to identify trends?
    b) needs to perform the test to receive reliable results?
    c) shall perform the same test at the same time?
3. How to set up test environment?
4. Which device is suitable to perform the test on?
5. Where to place/put test monitor and technical equipment such as cameras and microphones?
6. How much shall the test monitor interact with the test person?
7. Should the "think aloud" method be applied or not?
8. Make one example on how to carry out a usability test in a controlled test environment.

1. Conduct a usability test on three of Goobi ABs products and report the logic flaws.

Halfway through the thesis, once all tests from test 1 was completed, Goobi AB asked for a short report on the main flaws discovered during the test. These are presented in *Intermediate report.*

It was found that the most obvious problem in this application suite is that there is very poor consistency in and between the applications. The main consistency flaw is in how the various menus differ; or rather not differ, in functionality. There are several ways to carry out the same action. If each of the menus had a distinct and understandable connection with the elements it affects it would be easier for the user to guess where to find certain functions. The same yields for the soft keys; sometimes the left soft key is used to confirm and select items and some times it is the centre soft key and in a few occasions both keys operate the same way.

Another important flaw is the feedback that varies from function to function. Sometimes the user receives feedback several times, which can upset the user since they don't expect feedback at al or at least not that much feedback. Other times the absence of feedback, or just the fact that the user gets use to much feedback, confuses the user.

It was requested from the school tutor that a complete review of the tapes were made. At the beginning this felt like unnecessary work, since the main results was already presented. However, analysing the tapes from test one further gave more quantitative data and other observations was made. It was found that not every test person was treated in the same way. It appeared on the tape that the test monitors attitude towards the test person can affect for example how much help the test person receive. Other divergences as how much time the test person is allowed to continue with a task or how much the test monitor interact with the test person were also found on the tapes. One way to prevent this is to interact less with the test persons, especially if the test monitor do not have much previous experience in usability testing.

One procedure recommended for future test is to observe the tapes from the pilot tests carefully before the live tests are conducted. Important issues can be found which can be added as issues to look for in the observation sheet and directly logged during the real test. Also, the test monitor can learn more about how to interact with the test person.

2. How many test persons:

a) needs to perform the test to identify trends?

For the intermediate report trends were found within the seven observed tapes. Scott Weiss recommend six test and Jeffery Rubin four. However there is one problem with this recommendation. The number of available test persons depends on the product domain and this factor can be hard to control. For example sometimes the product is classified as secret and can not be exposed to outsiders. It can also occur that the system to be tested is only used by a few users which restrict the number representative end users.

b) needs to perform the test to receive reliable results?

If the purpose with the test is to compare different products and select the best one or if it is to measure if specific usability levels[22] is reached then a more strict approach is recommended. To obtain generalized results the test needs 10 to 12 participants per conduction according to classic experimental test recommendations.

Usability tests are rather used to obtain qualitative information about problems and information on how to correct problems when redesigning. When it comes to conduct usability tests significant result is not important in the

[22] An usability level may state that a certain percentage of the users should be able to carry out a certain task within a set time.

same way. Usability tests are not that strict and are rather used to observe trends.

c) shall perform the same test at the same time?

Our tests did only apply two test persons at a single test why this question is not entirely investigated trough the test sessions. However, at this single test the setup did not work very well why we recommend testing no more than one test person at a time. This is probably the most common way a user interacts with the mobile phone and it also force an exploratory behaviour from the user.

It may though be possible to carry out simultaneous test where several test persons are tested in parallel. This would require considerably more resources and is not recommended due to the fact that it makes the test more difficult to control.

3. How to set up the test environment?

When testing in a controlled environment such as a test laboratory, it is necessary to camouflage it to look like an environment familiar to the user or an environment where the tested product is likely to be used. This aims to make the user more relaxed and at ease and hopefully approach the scenario in a natural fashion.

Using little resources our test environment was set up to replicate a school environment. Since a majority of the participants were students, this environment would not be unfamiliar and it is a likely place to use mobile phone in.

A frequent comment was that the one-way mirror that covers a whole wall made the participants feel a little uneasy and it is recommended to disguise this as much as possible or place the participant so that the mirror can not be seen during the test.

4. Which device is suitable to perform the test on?

The choice of device depends on what goals are defined for the test. Different devices present slightly different results. Using a real mobile phone unit is of course the most natural way to go but it has its drawbacks. Recording an image that is good enough to review later is a bit more difficult using the mobile phone unit than it is using the simulated mobile phone on a computer. The mobile phone has to be fixed or the camera has to be very maneuverable. Also the mobile phone used, brand or type, may influence how the test participants conceive the software due to previous experience with it.

To achieve an authentic feel and to increase the users' curiosity the mobile phone is to recommend. Choose to use the simulator if it is too much of a problem making the software work properly on a mobile phone unit or if the software is only partly ready. Also consider using the simulator if the recorded image must be perfect.

5. Where to place test monitor and technical equipment such as cameras and microphones?
Participants during our tests claimed not to be aware of the cameras or microphones while testing. Perhaps perception of the surroundings became secondary once the test began and the tasks required concentration and focus.

Equipment such as cameras and microphones are positioned out of sight as much as possible. Even though participants claimed to not notice them it should not be obvious that they are being observed.

When it comes to the test monitors placement if in the test room a position beside and just slightly behind the test person is recommended to keep out of direct sight of the test person but so that it is still comfortably close.

6. How much shall the test monitor interact with the test person?
This is a question that depends on what is to be achieved by the test. If the test is mainly an exploratory test, interaction should be extensive. If the test is of a validating nature, interaction should be restricted and close to none.

It has proven during our tests that some interaction from test monitor is regularly needed to provide feedback and make the test participant feel relaxed and secure. Also, sometimes a little guidance about the scenario is required, not everyone understands what to do from just reading a scenario.

More interaction with the test monitor provides more correct and elaborate results concerning attitude, qualitative results, while it makes it difficult to attain reliable quantitative results.

7. Should the "think aloud" method be applied or not?
To have the participants thinking aloud is a really good way to understand what is going on and way certain actions are carried out. The difficulty with applying it though is that not all participants are equally keen on talking to themselves and results can vary very much. Also, not every comment pro-

nounced by the participants is true, it is common to say one thing and do another.

Think aloud adds to the basket of qualitative results while again the quantitative results may suffer from the extra time and effort applied to the test participants.

8. Make one example on how to carry out a usability test in a controlled test environment.

This thesis can be regarded as a basis for usability tests in the future. Test report 1 and 2 can in some extent be reused to conduct other usability tests and test plans. When the product is more complete a good idea is to start measure more quantitative data instead of mainly qualitative data which will evolve the assessment test in to a validation test and enable validation against usability requirements. Requirements provided by a customer or by a set of usability goals. It is recommended that Goobi AB hammer out usability goals that can be used long-term.

One conclusion drawn from this thesis is that there are several different ways of conducting the test. It is possible to change several parameters and thereby receive different results. Depending on what the objectives of the tests are, different approaches are recommended. Below a short description of two separate test methods follows that can be used to capture mainly qualitative or quantitative data.

	Exploratory test set up	**Validation test set up**
When	Used when mostly qualitative data is requested.	Used when mostly quantitative data is requested.
Objectives	It can be used to find user obstacles, conceptual understanding and their opinions about the product or the design.	Measure data to control whether a product manage predefined usability goals.
Platform	To explore design concepts it is not necessary to use a test device similar to the device intended. Even paper prototypes can be utilized.	To validate usability goals it is recommended that the environment the product is intended for is replicated and tested on.
Test persons	Depending on target group. The number of participants is at least four (if the context allows that many) and not more than eight.	To get reliable results 10 – 12 test persons should be used.

Dimensions of test	From this thesis it is recommended to do several smaller test rounds to test variations of different dimensions, instead of combining them in the same test, all to simplify the analysis of data.	Data collected are of mainly qualitative nature and hence more straightforward to collect. Test rounds can be carried out in slightly larger scale.
Positions	The test monitor is preferably in the room with the test person to generate discussion and hence attain qualitative results essential for the exploratory result.	Since natural behaviour and quantitative results are preferred the test monitor should not be present in the test room during the test.
Length of test	Around one hour altogether (briefing, test, debriefing). The test persons should not feel that the test is too long and reviewing the tapes should not be too exhausting.	See left.
Number of test per day	It was found during the test that it was hard to keep focus if to many tests were conducted at the same time. Preferable is 4-6 tests which also was the recommended number to observe trends.	Since more quantitative results are desired and logging of these data is less exhausting and possible to complete by reviewing the tapes a few more tests can be carried out per day than 4 - 6. Perhaps 6-8 is a good number.

Table 6. Comparison of exploratory and validation test setups.

4.1 Other recommendations

It is not suitable to test as many test persons as is in our tests; the amount of data becomes difficult to manage. It is better to conduct smaller tests, testing only one parameter at the time. A fruitful tip is to have at extra test persons who can be called in with short notice if some ordinary test person does not show up. Of course the break between the tests must be long enough to allow this action.

Test persons do not always say what they really mean, for example most of the test persons said that the environment did not affect them, but then they spontaneously said that the one mirror wall was intimidating. Therefore it is recommended to be careful when analysing the test persons' opinions. Another relevant recommendation is that it is better to carry out the

briefing in a separate room and there inform the test person about cameras, mirror wall etc.

Future work can contain an analysis of the end users and if the result is the same as the trends in this thesis, that novice feel insecure using a foreign language, it might be a good strategy to provide the products with support for the Swedish language.

In the theory chapter it is clear that there exist numerous of different models all invented to achieve better products. None of them can guarantee a better result but they are important when trying to understand the process and when planning the work to be done. It seems that all models work as guidelines which have to be modified and extended to each application's conditions and needs.

It is possible that various definitions and the lack of unambiguous standards bring confusion to the field of usability. Different companies use their own interpretation (if focusing on usability at all) and maybe this is a reason for why the focus on usability has been so weak until now.

The final words for this section is that usability test is not always the best method to collect information, but it is better to test than not to test.

5. Source of error

Even though usability tests do not strive to be statistically significant there is some information that needs to be presented to clarify which events occurring during the process to be considered sources of error. First of all the goal with the thesis was not completely formulated when the thesis began and during the work the purpose and question at issue was reformulated since the new version of the program was not ready within the time for this thesis. A possible consequence of this might be that the red line in the main report is sometimes not that obvious. If the purpose and question at issue had been stated from the beginning maybe a different approach would have been chosen.

Another possible source of error was the questionnaire for briefing used throughout test one. It did not provide the correct information needed about the test persons. The questionnaire did not show if the test persons used were suitable to use in this test or which predetermined group each participant would fit in to. In our tests previous knowledge about the participants allowed us to conclude this.

The malfunction of the questionnaire together with the fact that most of the male novice and experienced test persons were from a technical education while the spread among the female participants were larger, led to a lack of female novice test persons and an overflow of experienced male test persons. This was one of the reasons to rearrange the members for each group, another reason were the variations made during the test.

One of the variations was to use the remote control. The intention was to apply this variation to a number of the following test persons but since the remote did not function in a satisfying way, results from this test were discarded. Another variation was made due to the concern for two of the participants' knowledge of the English language. All text in the entire system is presented in English and therefore these two test persons were allowed to perform the test together.

Several smaller mistakes probably occurred during the test process which have not been recognised. An example of processes in which mistakes might have been made and not recognised is the actual test and the reviewing of the recorded data. Mistakes are probable to have occurred since the restricted time to carry out the test respectively the review forced many tests to be carried out each session and a large amount of data to be processed relatively quickly.

It can be discussed whether a literature study should have been conducted before the test or not. On one hand, if literature studies were performed before the test more information about what data to collect would have been available, how to design questionnaires would have been known, and other important issues could have been addressed in a more scientific way. On the other hand the risk for absorption into new knowledge would have been much larger and the thesis could have ended up like lots of other thesis with a summation of the studied field instead of a practical/experimental study. However, of the literature that was used, only one book addressed usability for handheld devices. The others addressed usability for computers, software or other technical devices.

6. Test report 1

Usability test plan for Master thesis
at the Department of Design Sciences, LTH
22/02/2005

Evaluation of Goobi project suite

-version XXXX

Authors:
Oscar Cosmo
Catia Hansen

6.1 Background

In this evaluation the client is a leading company in the field of developing and selling information-related applications for mobile phones (e.g.: browsers, messaging clients, file/content management, digital rights management, etc). Traditionally the client has been focusing mostly on the protocol and standards aspects of such applications. Since operators and the market require much more focus on the user interface and usability aspects now and in the future the client requests an evaluation of one of their recently released software application suite.

6.1.1 Definition of usability

The word usability has various definitions. In this report usability comes from Löwgren's definition of REAL [28]. Below is a short description of the meaning of each letter.

- **Relevance**: how well the system serves the users needs.
- **Effectiveness**: how efficiently the user can carry out their task using the system.
- **Attitude**: the users' subjective feelings towards the system
- **Learnability**: how easy it is to learn for initial use and how well the users remember the skills over time.

6.2 Test method

To evaluate this product a usability test will be done. Although the current status of the tested system version is released, the product test can be labelled as an assessment test since the objectives, as presented below, are to examine and evaluate the product in realistic tasks and identify specific usability deficiencies.

An assessment test is usually done during development and is characterised by the task based testing where the test person has little or no interaction with the test monitor. The tasks have to be carefully developed to provide just enough information to the test person. If too much information is incorporated in the task description it may affect the test person's manner of carrying out the tasks. If too little information is provided the test person may experience unnecessary difficulties understanding and solving the tasks at hand.

A pilot test is carried out to identify any problems with the test plan itself. If there is an obvious problem with it, it will be revealed and the test plan can be revised. The pilot test follows the test plan but observers and test monitor focus on how the test itself works and how tasks and questionnaires work instead of focusing on usability aspects of the product tested. These results are *not* included in the evaluation.

To ensure reliable results and to provide big enough groups for each variation to the test, twenty-four test participants are utilized. The twenty-four participants are divided into two main groups, experienced users and novice users, in which half of the members are female and the other half male. Experienced users are users that consider themselves familiar with this kind of applications and are used to handling similar products. These users will get very little, if any, help from the test monitor and are expected to complete all tasks. Novice users are not so familiar with application of this complexity. A novice user may use the mobile phone to make occasional calls and perhaps send SMS. Also the novice user is not so used to other technical artefact like a PC or PDA (Personal Digital Assistant). These users will have access to some help from the test monitor. However, help will consist of hints and encouragement rather than detailed and specific help. The test monitor will have to judge how much help is equivalent from one participant to the next.

The testing is not only an assessment test of the product but also an exploratory test of the test procedure why variation to the test is made. One of the variations is to test the tasks in different order, sequence variation.

This can be done to expose any transfer between applications; it may be easier to complete certain tasks in an application due to resemblance to another task in another application that is carried out first. To test this, the two main groups, experienced and novice, are each divided into two subgroups of three female and three male participants. One of the subgroups will carry out the tasks in a specific order, function by function (FbF), and the other subgroup will carry out the tasks in a random order.

	Advanced (FbF)	Advanced (Random)	Novice (FbF)	Novice (Random)
Male	*3*	*3*	*3*	*3*
Female	*3*	*3*	*3*	*3*

Table 7. Number of participants in each group.

6.2.1 User profile

A typical user of this system is difficult to identify since it has a very large target group. The tested version of the software is official, which allows the test to be performed on persons whom have no connection to the developing company, without the need for a disclosure agreement to be signed. Since the test environment is located within a technical university, it is easy to engage students in the testing procedure.

6.2.2 Test environment and equipment

Testing is carried out in the usability test laboratory at Ingvar Kamprad Design Centrum (IKDC) and facilitates a classic testing laboratory setup. Separated rooms for testing and monitoring characterize a classic testing laboratory setup. The separated rooms are divided by a one-way mirror, which allows the test person to be observed without feeling the direct presence of observers during the test.

Further measures to disguise the test performed are by creating an office setting instead of the sterile testing environment. This type of setting is typically static and this helps to ensure that the different tests will not be influenced by environmental disturbances. The test person is supposed to recognise the surroundings and feel comfortable with the situation. This aims to prevent the test person from acting like a test is performed and trying to fulfil expectations connected to the test, instead of using the device as intended.

The tested product is used in various situations, but since the test is performed on a PC, an office environment is the most natural and easiest to recreate. An office is often characterised by equipment like a desk, a PC, screen walls, telephone.

Evaluating mobile applications is a bit different from evaluating PC applications. To avoid operations normally connected to the conceptual use of a PC, for example using the right mouse button, which is not supported on a mobile phone, the mouse is removed and the keyboard is used only for entering text. Interaction is instead handled through a remote control. The remote control is slightly larger than a regular mobile phone handset and consists mainly of a touch sensitive LCD screen where the key set of a mobile phone is simulated. The screen contains, apart from the key set, an area which is not to be used in the test; therefore it should be covered during the test. On the face of the remote are some physical buttons not to be used in the interaction; these buttons will also be covered to prevent the test person from using them. Since the remote communicates with the

computer through an infrared receiver it has to be fixed in a position that allows direct line of sight between the remote and the receiver and this position should also simulate the position relative to the screen of a regular handset. To augment the connection between the remote control and the screen, the computer screen is covered, not including the part where the mobile phone screen is presented, and the remote is placed as close beneath the mobile phone screen as possible. A checklist of the equipment can be found in Appendix D.

6.2.3 Roles

During the test there will be different roles;

- **Test monitor:** The test monitor has the ultimate responsibility for the test in conduct.
- **Timer:** Incorporated in the test monitor role. Times actions specified in the test objectives.
- **Test assistant:** Interacts with test person. Introduces the test person to the test, objectives and goals. Aim to make the test person feel relaxed. Responses to help calls
- **Data logger/Recordings operator:** Responsible for collecting data and operate data collection instruments; video- sound recorders.
- **Observer:** Invited guests with special interest in the testing procedure. Observes the testing without interacting.
- **Product expert:** Responsible for technical aspects of the product being tested.

In this test there will be two personnel sharing these roles, observers not included. The roles are divided so that interference between them is minimized; for example can the responsibilities of the recordings operator role interfere with the responsibilities of the test monitor role.

6.2.4 Performing the test

The test is divided into three parts, before, during, and after the test. Below each part is described by it's specific activities.

Before: Once the test person arrives the test assistant greets personally and by giving a calm impression the assistant should make the test person to feel comfortable and relaxed. The test person is asked to fill out a simple questionnaire (found in Appendix A) gathering information about which mobile phone functionality the test person regularly uses, how frequent, and in what situations. At this time the test person is informed of the video/sound recording, observation wall, confidentiality of participation and

the golden rule for every test; it is the product being tested, not the test person.

When the questionnaire is filled out the assistant escorts the participant to the test room, explains the background and objective of the test according to a prewritten script. Also the test person is informed about some of the equipment used and the restrictions, such as not to use the buttons on the remote that are covered. The last thing the test assistant does before the test begins is to hand the test person the task list, explaining that the tasks are to be solved without assistance and that the test person is welcome to explain or comment every step taken to solve a task according to the “think aloud” principle.

During: Once the test assistant and the test monitor have left the test room and observe from behind the mirrored wall and the video/sound recording has commenced, the test begins. The test person is instructed through the intercom to begin solving the tasks. Timing of each task is carried out by the timer and the test monitor observes the test person and denotes any actions or behavior that could be of interest on a spreadsheet, along with a notation of time.

Depending on which group the test person belongs to different levels of help is provided. If the test person belongs to the experts group, no help is provided. However the test monitor is allowed to inform the test person when maximum time of completion is reached and instruct the test person to continue with the next task. The novice users are allowed some help and are told to ask for it when desired. When desired the test monitor will enter the room but only to supply the test person with hints and encouragement and no real hands-on help.

After: When the actual test has ended, because of maximum time of test is reached or all tasks fulfilled, the test person is handed another short questionnaire (found in Appendix C) aiming to capture the attitude towards the test and the applications. As a token of appreciation the test person is handed a small gift.

6.2.5 Data collection

During testing various data will be collected. Collection will be done automatically by video/sound recording and manually through questionnaires and logging during the test. Logging is supported by a spreadsheet where timestamps of interesting events is recorded (spreadsheet can be found in Appendix E). Data is either qualitative or quantitative and the way it is collected depends on what type it corresponds to and present phase of test.

Information about the test person's background is collected in the initial briefing questionnaire. During the test both types of data is gathered. Quantitative data by recording time of completion for specific tasks, time to recover from errors, time to realize and recognize an error, number and percentage of tasks completed correctly with and without assistance, number of errors, and counting incorrect selections. Qualitative data is gathered through study of the video recordings, observing the test person's behaviour and expressions; like hesitations, irritated behaviour, and quotes.

6.3 Goal of usability testing

Request from the client is specified in a confidential document. A primary interest is a general evaluation of usability aspects in their products and feedback that can be applied to coming releases. During the evaluation it is requested to obtain results about how well IKDC's usability lab fits evaluation of this kind of product, mobile phone applications. Furthermore the client desires a method to efficiently test mobile applications and guarantee a certain level of usability to be incorporated in each new release. It is desirable that the designed procedure is re-usable in future projects.

The client also prefers:

- Qualitative rather than quantitative data collection during usability testing.
- Task-based testing ("achieve this") rather than step-by-step instructions (to find non-intuitive UI flows etc)
- Focus on main products such as Browser, Messenger and Content Manager

6.3.1 Problem statements

Application related statements

- Is the line of action logical to perform operations?
    - Is it possible to make them more efficient and intuitive?
- Is the response time a cause of user frustration or errors?
- Does the user feel control over the application?
- Is the user provided with understandable and detailed feedback?
- Is the product easy to use?
- Is the product easy to learn how to use?
    - What areas can be improved?
- Are all the terms of menus and functions intuitive? If not, are terms learned through performing typical tasks?
- Does performed task order affect understanding for application?

Test related statements

- How well does this test method work for testing mobile applications?
- What is the best way to test mobile application?
- How many individual tests need to be done to capture the most important flaws? (One test per person or more)
- Which composition of test roles is preferred during the test?
- Which test persons are the representative end users?
- How many test persons are needed to ensure reliable results?

- How does help from test assistant influence the test results?
- Does performed task order affect test results?

6.3.2 Goals attained

Due to a dynamic test method one of the application-related statements could not be achieved. Variation of task order was not performed. Hence no results of how the task order affects understanding for application could be attained.

The test related statements could not all be answered in this one test set. However some of the statements were answered and tendencies for most of the statements were given.

6.4 Summary of test results

Test 1 was not only analyzed for the intermediate report, it was also a request from the tutor at LTH to review all tapes and log events properly according to the test plan. This section is divided into two sub sections where the first one presents overall results and the second presents results gathered from specific tasks.
Overall results: maximum time of test, total time for test, time to complete task

What task that corresponds to each number can be found in Appendix B. Scenario.

Throughout presentation of the results if nothing else is mentioned the x scale on the diagrams either shows each task (when numbers run from 1 to 15) or each test person (numbers run from 1 to 16).

	Female	**Male**
Novice	1-4	9-13
Expert	5-8	14-16

Table 8. The number of the test persons belonging to a selected group.

6.4.1 Maximum time of test MTT

Obviously no strict maximum time of test (MTT) limit was set. Some participants were allowed to continue the test several minutes after the informal MTT was reached. Two of the test persons did not solve task fourteen, which constitutes a basis of the last task, even though extra time was given.

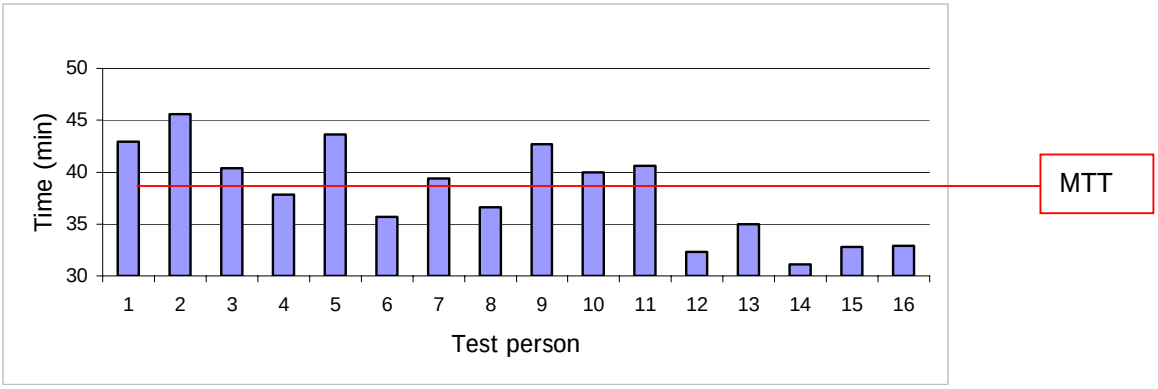

Diagram 1. Total time for each test person to complete the test.

6.4.2 Total time for test

In general it took between thirty and forty-five minutes to complete the test, see *Diagram 1*. Female novices required more time than the other groups while, as expected, male experts proved to be the fastest group. The remaining two groups displayed about the same mean time.

There are two noticeable deviations in completion time. Looking at the group of female experts, participant five seems to deviates most from the average time in the group by having a slower completion time. The second noticeable deviation is found in the group of male novices where test person twelve displays a significantly shorter completion time than the other three. If these two participants' results where to be excluded, a more evident difference between novice and expert users' completion times could be observed in *Table 9*.

	Female	Male
Novice	41,7	38,9
Expert	38,8	33

	Female	Male
Novice	41,7	41,1
Expert	37,2	33

Table 9. Mean time (in minutes) needed to complete the test for each group. Left: All participants are included. Right: Some participants are excluded.

Since the total completion time depends on how many tasks attempted and the completion time of each task along with the test person's application of the think aloud principle, which is time-consuming, these data are not reliable and relevant to measure.

6.4.3 Time to complete task

Diagram 2 shows the mean time it took for all test persons to complete each task. It is evident that task four, nine and fourteen required more time to complete.

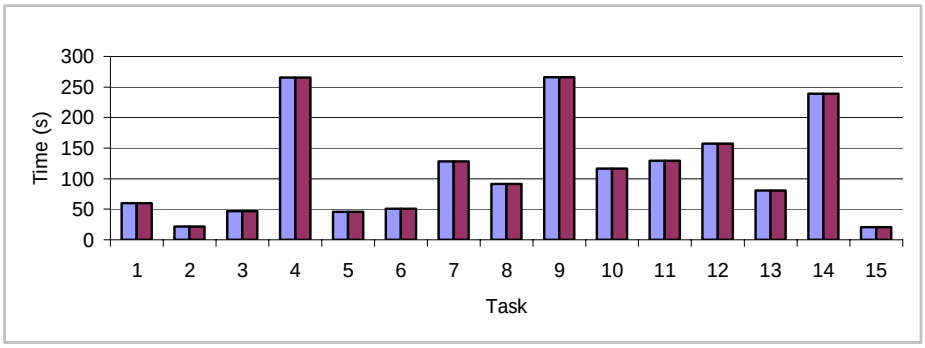

Diagram 2. Mean time to complete each task.

Studying diagrams 3-6, all tasks are sorted out that the majority of group members in each group spent more than three minutes trying to solve. This displays by group which task that consumed the most time. The result corresponds well to what can be observed in *Diagram 2* concerning task overall time.

	Female	**Male**
Novice	4, 9, 14, 12	4, 9, 14
Expert	4, 9	4, 9

Table 10. Tasks which took more than 3 minutes to solve for 3 / 4 test persons in each group.

Possible reasons for the amount of time spent at each task and in total could be that the logic in the simulated phone was difficult to understand, the scenario was complicated, and/or thinking aloud restrained test persons to act in a way they would do if they did not use this method. Other influencing parameters could be nervousness caused by the test environment, or that the test person tried to complete quickly due to awareness that a test was performed and therefore neglected important information.

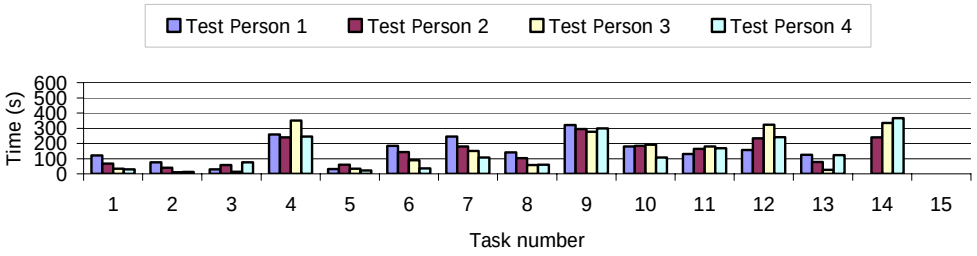

Diagram 3. Time to complete each task for female novices.

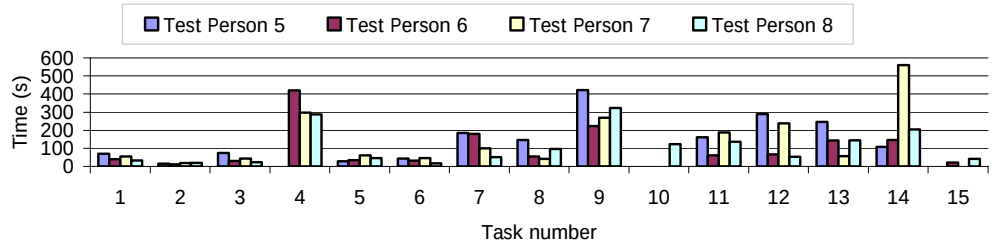

Diagram 4. Time to complete each task for female experts.

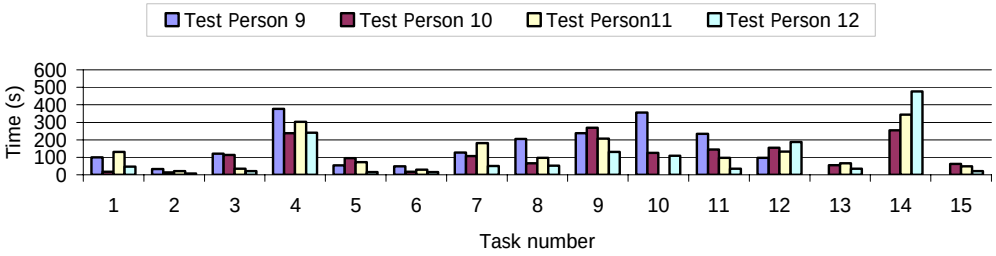

Diagram 5. Time to complete each task for male novices.

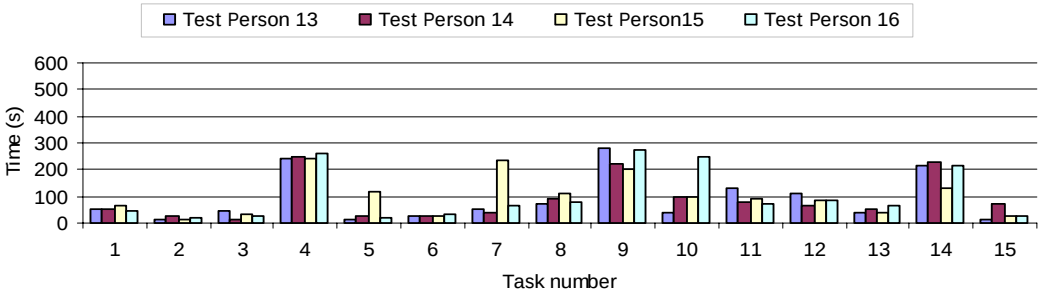

Diagram 6. Time to complete each task for male experts.

6.4.4 Help and m-button

In test method it was stated that the novice groups would be allowed to receive more help than the expert groups, but there were no strict definitions of what kind of and how much help should be allowed. During the test, it became obvious both groups needed to be informed, or clued-up, about the existence of the m-button. It was interesting to observe the manner in which the test persons solved the remaining tasks when they had been allotted another key. Help was given when the test monitors decided that the test persons had tried enough. A guideline for what was enough was in this case set at about 28 minutes of struggle. The results would probably be fairer if the guideline were set at a number of tasks instead of how long a time the test had progressed, since each test person required various time for each task. Now it is impossible to tell whether a test person who failed a task actually could have solved it if they had knowledge of the m-button. From *Diagram 7* it seems as a proper number of tasks to set the guideline at would be seven since seven out of the eight who found the m-button by themselves did so within the first seven tasks.

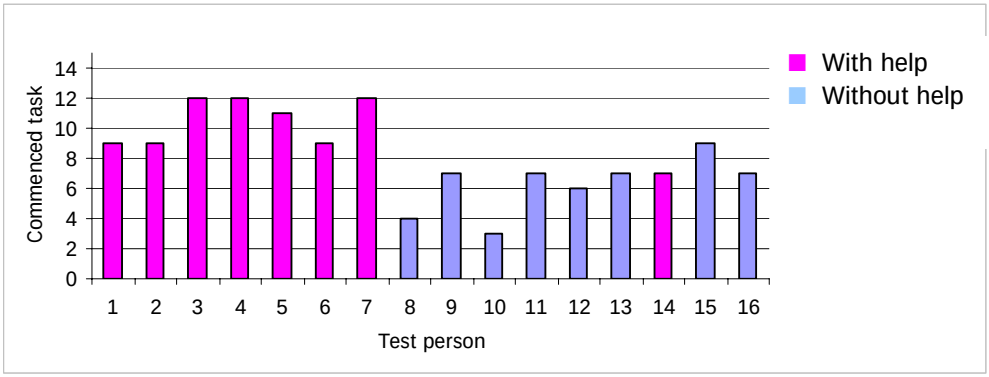

Diagram 7. Task commenced in which the m-button was found.

6.4.5 Passed tasks

Diagram 8 shows how many test persons in each group that failed respectively passed each task supplemented by *Table 11* showing which tasks the majority of participants in each group failed to complete.

	Female	Male
Novice	4, 7, 9, 10, 11, 12, 13, 14 (15)	4, 9, (10), (14)
Expert	(4), 7, 9, (10)	4, 9, 10

Table 11. Task which the majority in the group failed with.

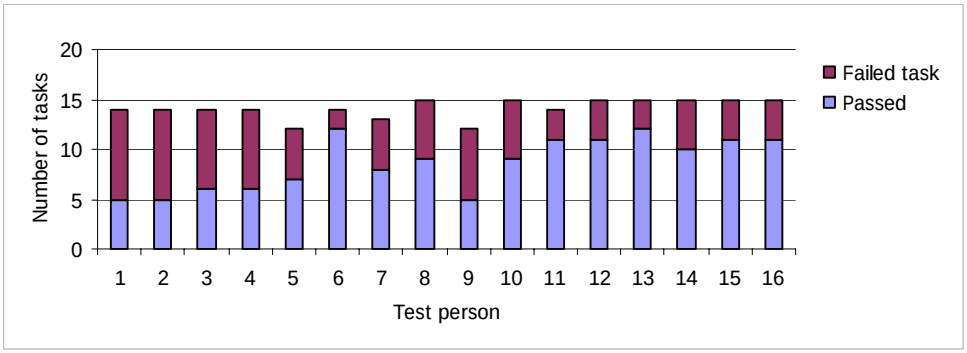

Diagram 8. Shows how many tasks each participant finished and passed respectively failed.

6.4.6 Graphical User Interface GUI –slips

Since the test was performed on a PC it was interesting how many times the test persons by mistake used the mouse to click in the area representing the GUI on the phone. From *Diagram 9* it can be seen that most slips were done in the beginning of the test and *Diagram 10* shows that it was very different who made the slips. It is remarkable that two persons in the female novice group only did two slips and the other group members did twelve or more.

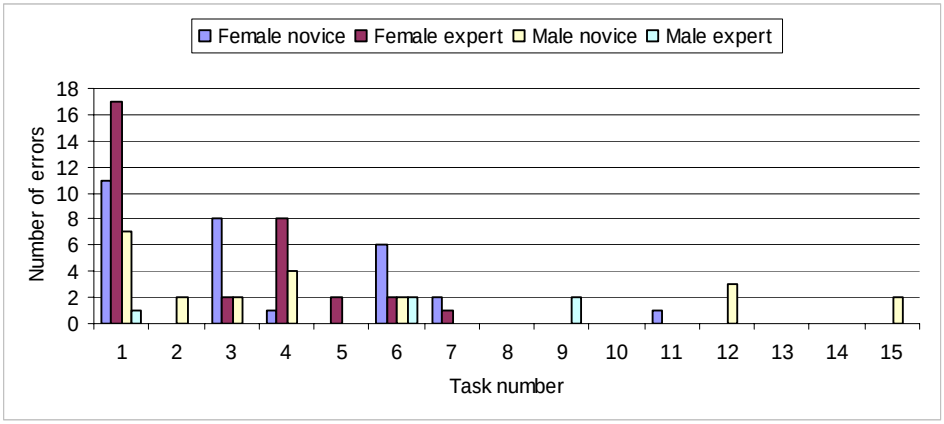

Diagram 9. Number of GUI-slips by task number.

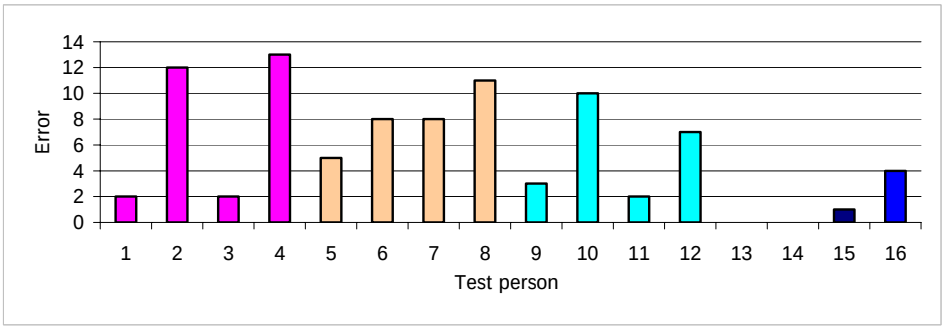

Diagram 10. Number of GUI-slips by test participant. Each colour represents a group presented in Table 8.

Another factor measured, inevitably reminding the test person about the test situation, was how many times the program broke down. Break downs, when the simulator presented an error message and then inevitably shut down, occurred regularly. A pattern of when, certain menus and options selected, the break downs occurred could be identified but was not recorded, hence is not presented in this report.

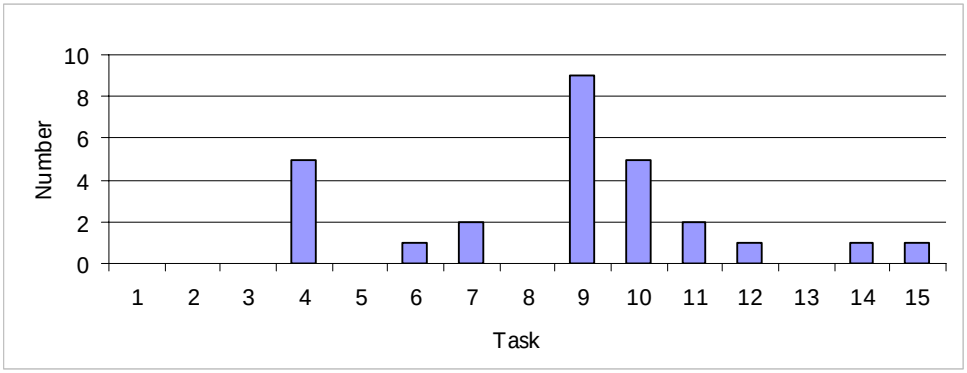

Diagram 11. Number of breakdowns during test.

6.4.7 Specific task study

Tasks are found in the scenario description, Appendix B.

1. **Scenario content:** Watch a movie.
   Observed action: Which key, to activate the highlighted icon, does the test person use?
   Result: Six test persons chose left soft key (LSK). Ten test persons chose centre soft key (CSK), which is not implemented to select highlighted icons in this view.
   Interpretation: Test persons use CSK by force of habit and because the button looks inviting by position and design.
   Several of the test persons claimed in the debriefing that they were used to be able to select menu items with CSK.

2. **Scenario content:** Open a web browser and view a homepage.
   Observed action: Which menu option does the test person choose to enter?
   Result: All test persons (except for one who didn't do this task) chose the same option. *Load homepage.*
   Interpretation: The scenario is clear and contains the same word as the name of the option.

3. **Scenario content:** Find a web address to a TV-tableau that has been used in the phone before.
   Observed action: Which menu option does the test person choose to enter?

Result: In this task the test persons chose four different ways to start the scenario, when they searched for a TV-tableau. One test person chose *enter address*, four chose *load history*, five *bookmarks* and six chose *history.*

Interpretation: Most participants understood the concept of saved URL addresses. However, the scenario was a little bit weak and not all understood that the address would be available on this phone.

4. **Scenario content:** Read and respond to an MMS with a MMS containing text and a picture.

Observed actions: **a)** Does the test person replay the received MMS?
b) How many times does the test person select LSK, believing this would confirm sending?
c) How many of the test persons tried to save a message in the drafts folder?
d) How many of the test person tried to send a message from the drafts folder?
e) Does the test person find the option *send* through the menu alternatives, or the menu which can be selected once an object is selected, or not at all?

Result: **a)** Four test persons did, ten did not.
b) Three test persons tried one time and two test persons tried three respective four times.
c) Eight test persons tried to save a message in a folder for unfinished messages. Seven did not try and one did not do this task.
d) Seven test persons tried to save a message in a folder for unfinished messages. Eight did not try and one did not do this task.
e) Eleven test persons found it once an object was selected, four did not find send and one did not do this task.

Interpretation: **a)** It seems like the test persons did not understand that they were supposed to respond to a message. The test persons probably thought they were only to create a new message and send it. Another possible explana-

tion is that they did not know where the option for this action was.

b) Analysis of the tapes shows that only five out of sixteen test persons make this mistake. The test monitors was under the impression that LSK was selected more often than these results shows. A possible explanation is that this action of error was made during other scenarios containing sending MMS but was not measured during these tasks.

c and d) When the test persons could not find the option for sending a message, they tried another way. All except one test person of those who saved a message also tried to send it from the folder for unfinished messages.

e) None of the test persons used the intended way to send a message. Many of those who found a send option were surprised to find it when an object was selected. They knew there was a way to send the message and either tried until they found a way or gave up.

5. **Scenario content:** Create a folder.
   Observed action: None since this scenario was used to increase the test persons' comfort in the test. All participants (except one) accomplished the task. The test person who did not pass gave up on this task.

6. **Scenario content:** Check tomorrows weather from a web site.
   Observed action: How much time does the test person spend on error messages?
   Result: See *Diagram 12*.
   Interpretation: It seems as the beginners had more difficulties dealing with error messages, but since the test persons received a different number of error messages this conclusion might be hasty.

7. **Scenario content:** Save a page offline.
   Observed action: **a)** How much time does the test person spend on error messages?
   b) How many test persons chose LSK to find a menu when they are online on a web page?
   c) Are the test persons online or offline the first time they try to save the page or do they not save the page at all?
   Result: **a)** See *Diagram 12*.
   b) Nine test persons chose LSK to find a menu when they are online on a web page and seven test persons do not.
   c) Three test persons tries to save the page when they are online, seven test persons tries to save the page when they are offline and five test persons don't save the page at all.
   Interpretation: **a)** See interpretation of specific task study number 6.
   b) It seems that the test persons have learned that LSK can be selected to find a menu.
   c) The test persons don't know that they have to be online on the page when they want to save it, or they can't find the options to save a page when they are online.

8. **Scenario content:** Add a personal picture in a MMS template.
   Observed action: Where does the test person save the new template?
   Result: Seven test persons saved the new template in a folder for templates. Six test persons saved the new template in a folder for unfinished messages (draft). Three test persons did not perform this task.
   Interpretation: The test persons were either not familiar with the concept of templates or tried to complete the task quickly and there for didn't pay attention to where the message was saved. The fact that many test persons had saved a copy of the templates in draft probably contributed too.

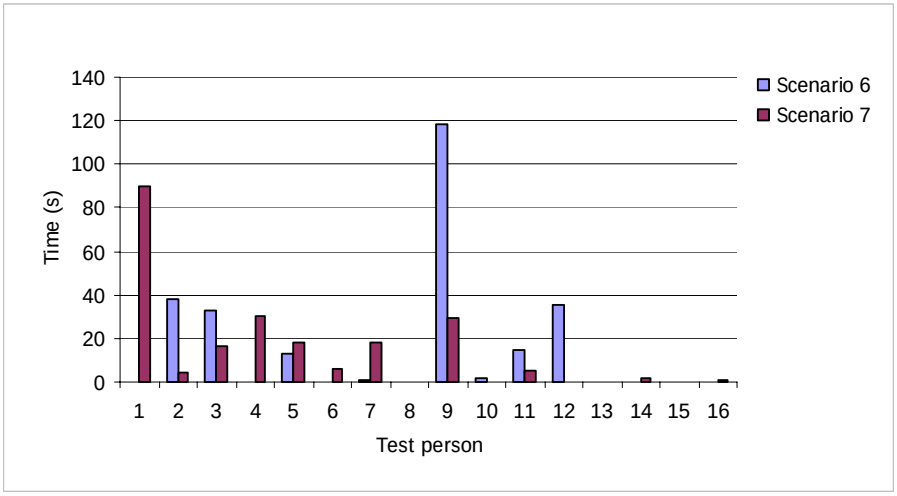

Diagram 12. Time spent on error messages.

9. **Scenario content:** Send an MMS to an address given on an internet page.

   Observed action: **a)** Does the test person start by creating a MMS or by surfing to the given address?
   b) How many times did the test person choose the link with the address before they send a MMS?
   c) How many of the test persons had problem with sending a MMS?

   Result: **a)** Fourteen of the test persons chose to surf to the given address. Two test persons started with creating a MMS.
   b) See *Diagram 13*.
   c) All of the test persons had problem with sending MMS.
   c) All test persons had problem with sending a MMS.

   Interpretation: **a)** Since the majority chose the right way they probably understood the scenario.
   b) All test persons chose the link at least one time before they send the message. It seems like they did not understand how to use this function.

c) Even though the test persons had tried to send an MMS before it was difficult for them to do it again.

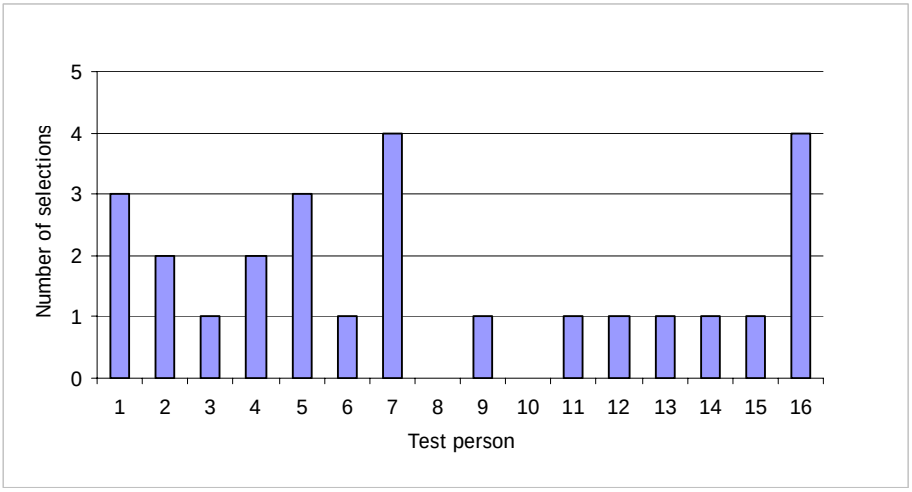

Diagram 13. Number of times the link was selected.

10. **Scenario content:** Use an already sent MMS (scenario 4) and add a sound.
    Observed action: None, since only one person was able to send the MMS in scenario 4 and some of the test persons added sound in scenario 4 and could therefore skip this scenario.

11. **Scenario content:** Delete all except for one message in the folder where unfinished messages saves.
    Observed action: **a)** Does the test person use the function to select and manipulate several messages at once, or are they deleted one by one?
    b) How many times does the test person enter a MMS (into edit mode) while trying to erase the message?
    Result: **a)** Eight test persons found the operation and six test persons deleted the messages one by one. Two test persons did not do this task.
    b) See *Diagram 14*.
    Interpretation: **a)** The intended operation was used by only half of the participants which imply that it was not easy to find or expected to exist.

b) The participants had different amount of drafts saved when they started to delete them, so if they enter each message, the number of drafts affects this result.

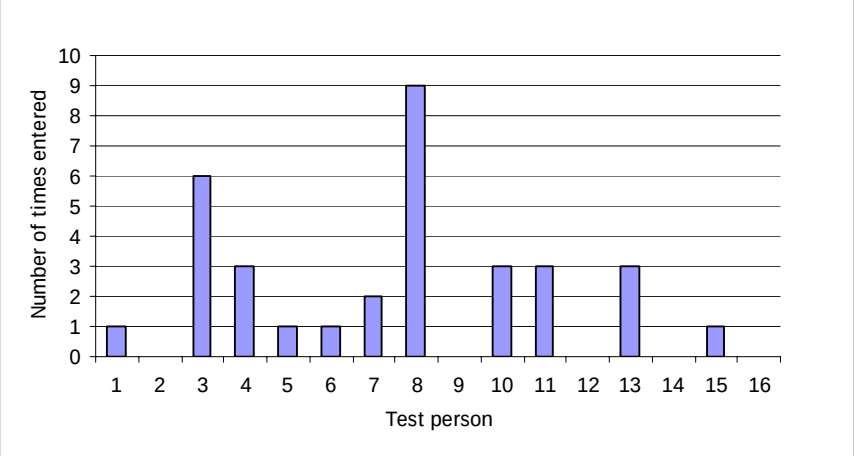

Diagram 14. Number of times each test person entered edit mode while trying to delete a message.

12. **Scenario content:** Save a picture from a given web page.

Observed action:
a) Does the test person use the menu key (m-key)?
b) Can the test person find the menu option for selecting a picture online?
c) Does the test person use LSK to select the picture once it is highlighted?
d) Can the test persons find the menu option for deactivate highlighted objects?

Result:
a) Fifteen of the test persons use the menu key one do not.
b) Eleven test persons finds the option for selecting a picture online, five do not.
c) Nine test persons do use LSK to select the picture once it is highlighted, four do not and three don't do this task.
d) Ten test persons find the menu option for deactivate a highlighted object, four do not and two do not do this task.

Interpretation:
a) At this time all test persons knew about the menu key and probably used it because

they had learn that it was a key with interesting menu options.

b) It is still many who found the option, but not as many who entered the menu. It is possible that the name of the option was unexpected and therefore a number of test persons missed it.

c) This probably shows that the test persons are confused whether LSK contains menu options or not.

d) Almost all the persons who were able to find the option for selecting the picture also found the option to deselect, but to most of them it was not clear that they should use an option to be able to exit the page. They found the option to deselect the picture when they were trying to find a way out from the page.

13. **Scenario content:** Do a setting that prevents pictures to be shown automatically on a web page.

Observed action:

a) How many times does the test person choose LSK respectively CSK to select the option?

b) How many times does the test person enter the option for changing picture settings, to confirm that the selected option is chosen?

c) Does the test person surf to a web page to confirm the selected option?

Result:

a) See *Diagram 15*

b) See *Diagram 16*

c) See *Diagram 16*

Interpretation:

a) Most participants that used LSK became insecure of whether the option was selected or not and went back to check sometimes trying LSK again.

b) Most participants do not enter again to check once the CSK is used.

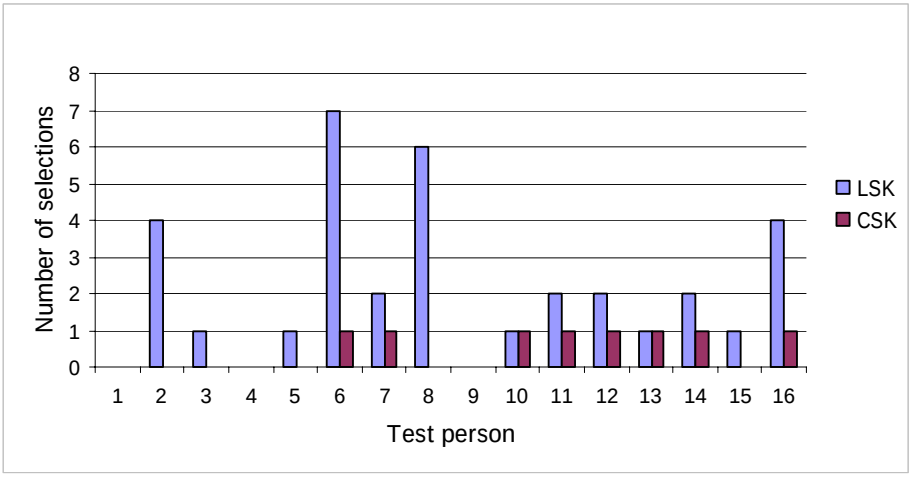

Diagram 15. Number of times LSK respectively CSK was selected.

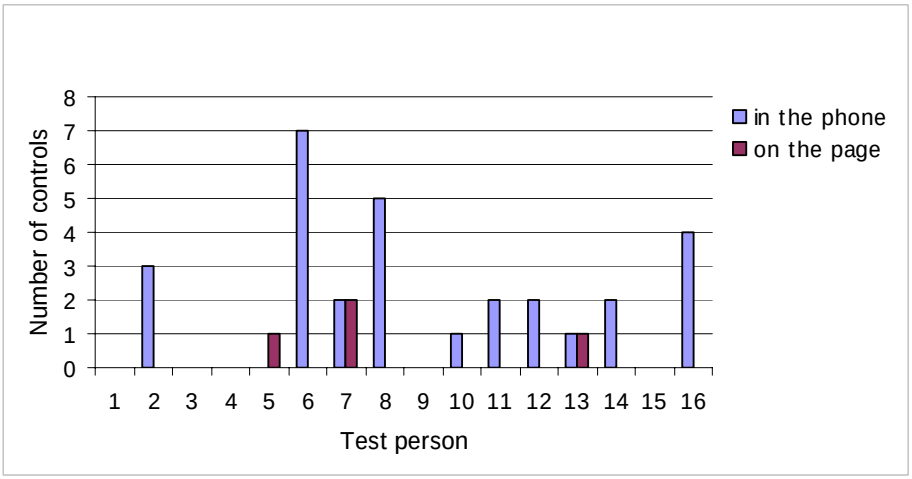

Diagram 16. Number of times the setting was checked.

14. **Scenario content:** Save a picture from a received MMS together with an already saved picture in a new folder.
    Observed action: **a)** Does the test person save the message using an option that saves templates directly, manipulates objects or is instead the picture saved using an option that manipulates objects?
    b) Does the test person use any soft keys while the MMS is playing?

c) Does the test person enter a folder with a name that reminds of MMS (MMBox)?

Result:

a) Four test persons saved the message using an option that saves templates directly, seven were using an option that manipulates objects, and four saved the picture.

b) Thirteen of the test persons used some soft key while the MMS was playing.

c) Four test persons did enter MMBox, and eleven did not.

Interpretation:

a) The option to save a template directly was not obvious. The option was not expected or the distinction between this option and the save option is not recognised.

b) The test persons did probably not understand that the MMS were playing.

c) When looking for an option the closest to MMS in this menu is MMBox. This is probably why this is selected. Most do not know what it means.

15. **Scenario content:** Rename a folder (which was created in scenario 14)

Observed action: None since this scenario was used to increase the test person's comfort in the test.

6.5 Source of error

Test method, test environment, and other unexpected factors influence the test results. It is recommended to keep these influences as low as possible to avoid misleading results. However due to the nature of the unexpected factors it is hard, if not impossible, to prevent occurrence. To minimise the impact of unexpected incidents to the test result, all known occurrences should be noted and kept in mind while analysing the results. Any problems with the test plan, questionnaires, or other test specific documents observed during the set should also be noted and analysed to elucidate possible influence and allow improvement.

During this test set the test method was not rigorously fixed. Changes during the test were allowed to a certain degree. The major change made was to remove the variation of task order and to decrease the number of participants. This dissolved the original classification of the participants and produced new classification of advanced- and less advanced male and advanced- and less advanced female. Each group consisting of four participants. In addition to these sixteen classified participants four extra participants were tested, some with variations in the test method. The main reason for abandoning the variation of task order and consequently dissolving the original classification was the final formulation of the scenario. The original idea of task order variation was to ascertain how learnability between applications is transferred. However, when the tasks were added to the scenario their internal order managed to establish the desired result without the need of reordering.

To broaden the variation of test set-ups some of the last participants, that could not be categorised, carried out the test in different manners. One test was carried out using the original idea of a remote control for navigation and another was carried out with two participants working together with the tasks. A third variation was made with the test monitor present in the test room but avoiding interaction with the participant. The results from these tests are included in the analysis with reservation.

Many of the participants had trouble understanding the idea of the scenario and in some cases the scenario itself. Participants became confused when a task did not describe exactly what to do and exactly when the task is completed. The idea of the scenario is to give the participants a background to why a task needs to be performed and then provide some liberty of action to explore ways of solving the task. Consider explaining this to the participants and also consider including a clear description on when a task is completed in each task description.

A problem that came to attention after a few test sessions was that the questionnaire used to collect data about the participant's background was flawed, Appendix A. A distinction between advanced and less advanced participants was difficult to do merely from the questionnaire data why it is not presented in the results.

Testing was carried out on a purely English-language simulator. This constituted a problem to some of the participants since their native language is Swedish. Although some of the technical terms and abbreviations are the same too many of the function names were hard to translate.

Utilisation of the think aloud principle affects the time of completion for each task as well as entire test. Participants, who tend to speak very much, explain and perhaps stop to make a comment, of course spend more time at each task than a participant that is quiet and focuses on solving them. Both types are valuable to the test but the fact has to be considered when evaluating timing data.

If test is performed on test persons known to the test monitors, this might inflect the attitude towards the test person. If the test monitor already has a good feeling for the test person, the test person might receive more help and allowed to carry on longer with the scenarios, on the other hand if the test monitor has a bad feeling then the test person might receive less help and be forced to proceed even though the current task is not completed.

6.6 Conclusions

This test used sixteen test participants, plus four that are not included in the results, and three pilot test participants. It became clear that this was too many participants since the amount of data collected was immense and time spent on carrying out the tests can not be justified compared to the result they presented. No specific results were found after the seven tapes initially observed of the collected data. Statistics of how many participants that made a specific error is however more correct when using many participants.

Results presented in the

Summary of test results along with the intermediate report will constitute the foundation for an analysis, below, of the tested products using the REAL definition presented in *Definition of usability* in the beginning of Test report 1.

Relevance
It is very hard to rate the relevance of this kind of product since it is very restricted in its functionality and has no specific problem to solve other than supporting mobile telephony, text messaging services, mobile and connectivity. All of which are well supported in functionality.

Efficiency
Using these products individually they support the user quite well. All functionality expected is included and complemented by some additional features aiming to make any task easier to carry out. Also, interaction between products is well supported. However, the terminology used makes it harder for the user to realise all functionality included and expectations are lowered. Some tasks even become more cumbersome having to figure out where to find the function if it even exists.

There are small, if any, possibilities for the experienced user to speed up the work, no shortcuts exists.

Attitude
As declared above, expectations on functionality were not very high. Partly due to the mentioned terminology but also as a consequence of the unfinished look of the interface. Also, participants in this test expressed that they were often confused and clueless of what to do next. Despite this most participants claimed that the overall attitude was positive. Possibly this is an unconscious lie observed to occur [2] to satisfy the test monitor.

The attitude toward the products was also influenced by the fact that the applications were tested on a simulation of a mobile phone and not a real unit. Not only did this impact on the graphic design and layout of the interface, some parts of the applications did not look very appealing, but also on the conception of the product as finished. A majority of the test participants have some experience in programming which may have caused them to consider the simulation as a not fully developed program.

Learnability
Findings concerning inconsistency harm learnability of the tested applications. To achieve good learnability there must be consistency in the way

similar actions are carried out and where similar information can be found and how it is presented.

Several flaws concerning the consistency were found, both in interaction and presentation. Arbitrary use of the keys mars the interaction while presentation suffers from a lack in standards of the style in which to present graphics and text.

Participants displayed poor learning of similar actions in different tasks. Often a course of action was forgotten when it was next needed. This is probably mainly a result of the complex and illogical steps in a sequence of actions making it hard to remember.

6.7 Appendix

A. Briefing

Hej!
Välkommen och tack för att du ställer upp som testperson. Vi gör denna undersökning för att bättre förstå hur vår klients applikationer fungerar samt för att undersöka bästa sätt att testa mobila applikationer i framtiden. För att bättre förstå ditt beteende under testet ber vi dig fylla i följande frågor. Med hjälp av din bakgrund hoppas vi kunna tolka testet av produkten på ett bättre sätt.

Tack på förhand

---

1. Ange kön:

Man ☐
Kvinna ☐

2. Ange ålder:

<19	20-24	25-29	30-39	>40

3. Ange utbildning: _____

_____

4. Ange yrke: _____

_____

5. Markera på skalan hur bra du anser dig vara på att:
(5 står för mycket bra och 1 står för inte så bra.)

	1	2	3	4	5
hantera mobiltelefoner:	☐	☐	☐	☐	☐
hantera datorer:	☐	☐	☐	☐	☐

6. Vilken beskrivning passar bäst in på din inställning till teknik?
(Markera med ett x för det alternativ som passar dig bäst)

* Jag vill alltid ha det senaste inom teknik och försöker få tag på produkten *innan* den är ute på marknaden. ☐
* Jag vill alltid ha det senaste inom teknik och köper produkten *när* den har ☐

kommit ut på marknaden. ☐

* Jag gillar ny teknik och köper produkten så fort jag vet att den fungerar. ☐

* Jag är inte så intresserad av ny teknik och väntar gärna länge innan jag köper en teknisk produkt. ☐

* Jag är inte alls intresserad av ny teknik och köper sällan nya tekniska produkter. ☐

7. Hur ofta gör du följande med din mobiltelefon?

	aldrig	1-3 ggr/ månad	1-3ggr/ vecka	1-3ggr/ dag	Mer än 3ggr /dag
Ringer	☐	☐	☐	☐	☐
Fotograferar	☐	☐	☐	☐	☐
Skickar SMS	☐	☐	☐	☐	☐
Skickar MMS	☐	☐	☐	☐	☐
Lyssnar på musik	☐	☐	☐	☐	☐
Ser på film	☐	☐	☐	☐	☐
Tar tiden, använder alarm	☐	☐	☐	☐	☐
Använder kalendern	☐	☐	☐	☐	☐
Ändrar inställningar	☐	☐	☐	☐	☐
Spelar spel	☐	☐	☐	☐	☐
Lägger in telefonnummer	☐	☐	☐	☐	☐
Använder miniräknaren	☐	☐	☐	☐	☐
Söker information på WAP/WWW	☐	☐	☐	☐	☐

8. Om du skulle köpa en ny telefon, vilka funktioner skulle du då prioritera mest? (5 står för hög prioritet och 1 står för låg prioritet)

	1	2	3	4	5
Kamera	☐	☐	☐	☐	☐
SMS tjänst	☐	☐	☐	☐	☐
MMS tjänst	☐	☐	☐	☐	☐
Musik spelare	☐	☐	☐	☐	☐
Visa film	☐	☐	☐	☐	☐
Spel	☐	☐	☐	☐	☐
WAP/Browser	☐	☐	☐	☐	☐

***9.* Terminologi.** Följande är exempel på termer som förekommer i telefonen. Vänligen förklara, så utförligt du kan, innebörden av dem. (Om det är ett ord du inte känner igen så får du gärna gissa vad du tror att det betyder. Fortsätt sedan med nästa uppgift.)

Homepage:

MMS:

Browser:

Content Manager:

Push message:

Skin:

Select:

BookMark

Draft:

SCENARIO

- för utvärdering av mobiltelefonapplikationer

Läs igenom följande noggrant!

Scenariebeskrivning
Detta är en beskrivning med uppgifter vi vill att du skall lösa. Föreställ dig de situationer som beskrivs och utför därefter de efterfrågade uppgifterna. Du har ingen annan att fråga utan får försöka lösa uppgifterna efter bästa förmåga. När du känner dig färdig med ett scenario, återgå till meny sidan på telefonen. Vi ser gärna att du under testet tillämpar ”tänka högt” metoden, dvs. säger högt vad du tänker.

- Föreställ dig
- Lös uppgifter
- Gå tillbaka när du är färdig
- Tala högt

OBS!
När du löser uppgifterna, ha då i åtanke att du endast ska använda de tre funktionerna i den övre menyraden.

Figur. Den röda rutan visar telefonens övre menyrad.

Scenario:

1. Du har införskaffat en ny mobiltelefon. Det finns lite bilder och filmer i den enligt försäljaren. Titta på en film som finns i katalogen [*svg*].

2. Du är lite nyfiken på webbläsaren som finns i telefonen. Öppna webbläsaren och titta på startsidan.

3. För att locka dig till att köpa din telefon, visade försäljaren dig något du kunde utföra i din telefon. Försäljaren visade dig en webbsida, i din telefon, där du kunde titta på tv tablåer för någon tv kanal. Nu har du glömt bor adressen, men vill gärna titta på sidan igen. Hitta den!

4. Nu har du precis kommit till din arbetsplats, ett kontor i ett flervåningshus. Det är ingen annan där. Plötsligt vibrerar det i din ficka. Du har fått ett meddelande, läs det och följ instruktionerna.

Hurra!
☺ Nu har du nästan gjort en tredjedel av uppgifterna.☺

5. Av erfarenhet vet du att en massa sparade meddelanden i inkorgen skapar oreda och det är aldrig bra. Skapa en mapp i inkorgen där du kan spara alla meddelanden du får från Bente.

6. Ditt fasta nätverk till din PC är tillfälligt ur funktion. Du har planer på att arrangera en picknick för dina vänner. Kontrollera förutsättningarna genom att ta reda på hur vädret blir i morgon. *(Tips: http://se.weather.yahoo.com)*

7. För att slippa ladda sidan varje gång du vill visa någon veckas väder kan du spara sidan och den och visa den ”offline”, dvs. utan att behöva koppla upp dig mot Internet. Spara sidan.

8. Eftersom du är en så ordningsam person och inte tycker om att låta människor vänta på dig utan att veta om du kommer när ni bokat ett möte så skickar du alltid ett MMS där du ursäktar förseningen men att du är på väg. Detta sker ganska ofta och för att slippa skriva ett nytt MMS varje gång så är det bra att ha en mall med ett färdigt MMS redo att skickas. Lyckligtvis finns det en sådan mall (Blir sen...) men för att göra den lite mer personlig så vill du lägga till en rolig bild i mallen. Välj att lägga till en bild från katalogen [*JPEG*].

9. På sidan **www.efd.lth.se/~c01oc/Goobi** finns det en bild du skulle vilja ha i din telefon. Skicka ett MMS till den adressen som finns på sidan till webbmastern och meddela att du kopierar den bilden från sidan.
   (Tips: tecknet tilde (~) skriver du genom att trycka på knapparna <Alt Gr> och <~> samtidigt, tryck därefter mellanslag så visas tecknet.)

10. Kaj tyckte att födelsedags-MMS:et du gjorde till Lisa var fint, men vill gärna att du lägger till låten ”addams” i MMS:et. (*Tips: addams.mid*)

11. När du tittar på vissa webbsidor tar det så lång tid att ladda dem. Det går fortare om man inte behövde ladda alla bilder på sidorna, det är ju ändå texten som är intressant. Gör en inställning som hindrar bilder från att visas automatiskt.

12. I katalogen [*Drafts*] finns det en massa meddelanden som du börjat skriva på men inte avslutat och istället sparat till senare tillfälle. Nu är det dags att rensa bland dem. Ta bort alla utom ett meddelade, som du nog kommer att skriva klart så småningom, i [*Drafts*].

13. Bilden som fanns på **www.efd.lth.se/~c01oc/Goobi** skulle passa perfekt i ett meddelande du tänkt skriva till en kompis. Ladda ner bilden och spara den på lämplig plats i din telefon.

14. Du har fått ett MMS med en ny bild av din kompis Simon. Sedan tidigare har du en bild på honom i katalogen [*work*]. Spara bilden du har i katalogen [work] och bilden du fick i MMS:et i ny katalog med namnet [*Simon*].

15. För att bättre återge vad katalogen innehåller så är det en bra idé att byta namnet på den till [*BilderpåSimon*]

C. Debriefing

PRODUKTEN

1. Beskriv hur du uppfattade produkten du nyss testat!

FRITT:

Navigeringen:

Menyerna:

Ikonerna:

Funktionsnamn:

Antal funktioner:

Feedback:

Färgsättning:

Hur produkten var att använda:

2. Vilka förväntningar hade du på produkten?

Förväntningar:

3. Fungerade produkten som du förväntade dig?

4. Hur ofta brukar du använda mobiltelefon, PDA, dator (ex timmar per vecka;

Mobiltelefon:

PDA:

Dator:

5. **Hur uppfattar du denna produkt jämfört med liknande produkter? Vad jämför du med?**

Produkt:

Uppfattning:

6. **Visste du inför varje scenario hur du skulle lösa uppgiften? (Om nej, hur bar du dig åt för att lösa den?**

7. **Hände det något oväntat under testets gång? (Förväntningar på sådant du trodde skulle inträffa, men aldrig gjorde det. Saker som inträffade fast du inte förväntade dig att de skulle inträffa)**

8. **Liknar ikonerna på simulatorn, ikonerna i den telefon du använder idag? Hur tror du ikonernas utseende påverkar ditt sätt att navigera?**

TESTET

9. **Hur upplevde du testet? (Tips om förändringar)**

10. **Hur upplevde du ”tänka högt” metoden?**

11. Vad anser du om hjälpen från testledaren? (tillräcklig, onödig, överflödig...)

12. Hur går du till väga om du möter problem när du använder din telefon? (Frågar någon ”expert”, provar dig fram, läser manualen, ger upp)

13. Vad använder du för mobiltelefon idag?_____

14. Vad är avgörande för ditt val av telefon? (märke, design, funktioner/applikationer)

D. Checklists

Checklists preparations

Test room

- ❒ Check software*
- ❒ Check equipment*
- ❒ Light on
- ❒ Rearrange furniture (look at picture [PICTURE])
- ❒ Tape a black sheet over the screen
- ❒ Camera arrangement
- ❒ Microphone
- ❒ Remote control
- ❒ Screen

Observer room

- ❒ Check equipment*
- ❒ Lights of

Checklist equipment

Test room

- ❒ PC
- ❒ Software
- ❒ Cameras
- ❒ Microphone
- ❒ Remote control
- ❒ Chair
- ❒ Desk
- ❒ Screen wall
- ❒ Paper / pencil
- ❒ Briefing

Observer room

- ❒ Video recording equipment
- ❒ Sound system
- ❒ Tape
- ❒ Timer
- ❒ Debriefing
- ❒ Logging form
- ❒ Pencil

Checklist software

Unified messaging

- ❑ Inbox contains 1 read message containing a picture
- ❑ Draft folder contains at least 8 messages
- ❑ Sent folder is empty
- ❑ Templates folder contains 1 template called “Blir sen..”

Browser

- ❑ The address wap.tv4.se/tabla.aspx is the only address in history
- ❑ The bookmarks folder is empty
- ❑ The enter address list contains only
    - www.efd.lth.se/~c01oc/Goobi
    - se.weather.yahoo.com
- ❑ Reset settings to show images when browsing

Content manager

- ❑ Reset content in folders by copying the usr_data to session files

Online content

- ❑ Check www.efd.lth.se/~c01oc/Goobi so no network failures has occurred
- ❑ Check se.weather.yahoo.com so no network failures has occurred
- ❑ Check wap.tv4.se/tabla.aspx so no network failures has occurred

*= separate checklist below

E. Data Collections sheets

Spreadsheet for time of completion (TOC) recording

Test person:

Task	Estimated TOC (min)	TOC (min)
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
SUM	0	

Spreadsheet for during-test observations

Test person: _____

TIMESTAMP	COMMENT

7. Intermediate report

for Master thesis

at the Department of Design Sciences, LTH

21/03/2005

Intermediate report
Evaluation of Goobi application suite

-version XXXX

Authors:
Oscar Cosmo
Catia Hansen

7.1 Introduction

This document is an intermediate report concerning obvious usability flaws in the XXXX applications suite and initial analysis of the test results from usability testing of this suite. The analysis is based on experiences during the test sessions and careful studies of seven of the recorded sessions; it is not in any way complete. However, the analysis is in broad outline and conclusions comprised are well-founded in material studied so far.

Problems found are sorted into which application they appear in. Some problems appear in several applications. The summary gives an overview of the suite and contains more examples and some solution ideas.

It is expected that the reader of this document possesses knowledge about the specific applications and it is recommended that reading be supported by visualisation in the Goobi Simulator.

7.2 Terminology

Explanation of special abbreviations used in the document.

- LSK = left soft key
- RSK = left soft key
- CSK = centre soft key
- m = the m key
- c = the c key

In the following text, the soft key options name is written directly after the key. For example if the left soft key has the function ok it is written LSK “Ok”.

7.3 Browsing

Test shows that exit from the browser is done by using the RSK ”Back” until the RSK option “Exit” appears. This term, “Exit”, is not used when exiting from any other application, like the Unified Messaging where back serves the same purpose. Also, the “Exit” has to be confirmed with the LSK “Ok” or the RSK “More” with the options “Cancel” and “Exit”. If the RSK “More” and then
“Exit” is selected nothing happens. It would be better with just a choice between “Ok” and “Cancel”.

When using the RSK “Back” often error messages or notifications appeared. These notifications contained a very long URL address and then a question to reload page, or object, or not to reload. There are problems with these

kinds of messages. They are too long to allow an overview of the full message, besides, most of the information is not relevant or even understandable for the common user. The length also stated a problem when there are several messages piling up on top of each other; the user clicks LSK "Ok" but gets no feedback that the message is removed because the message beneath it appears and it looks almost the same.

Another problem regarding these notifications is the soft key options. The LSK has the option "Ok", which is fine, and the RSK has the option "More", which opens a menu consisting of the options "Stop" and "Cancel". Sometimes though, at random moments, the "More" option contained the options "Exit" and "Cancel" and sometimes the LSK and the RSK both contained the "More" option, however with the options "Select" and "Select" on the LSK. The usage of more is not clear since its menu contents vary. The participants thought that more was a menu option for the various alternatives to be found behind "m" menu. Once the "m" menu key was found the alternative "More" was rarely used.

In browser history, there are too many steps, or perhaps the steps' signification is not clear, until the desired page is loaded. It is not evident that a submenu, where a specific page on the domain can be selected, appears when an address is selected.

One of the scenarios tested the browser's storage function where webpages can be stored for viewing offline. Several participants had trouble understanding what this name meant and did not interpret it as the function they were searching for, even though it was. However the main problem with this function is the feedback. Some users tried to save the page directly, not having loaded it in the browser. This resulted in an error message saying "Could not save page offline". There is no explanation to why it could not be saved or how to do it, which would be very helpful. Some participants, however, managed to save the desired page but pointed out that if there would have been a full list of already saved pages, it would have been impossible to tell whether the page actually was saved or not. There is no notification of a successful save.

The same problem with lack of feedback is present in the browser's bookmark function as well; there is nothing that confirms that the bookmark has been added. This may be because the actual bookmarks are not visible in the first bookmark page, only the options "Add bookmark" and "Manage bookmarks". To se if a bookmark exists one of these options has to be selected. In the first bookmark page there should be actual bookmarks and bookmark folders.

Browser functionality contains means to interact with objects on a website, not only links and forms but also pictures. This function is called “Object mode”. Very few of the test participants realised and what this option was when seen in the menu. However, after testing everything else to be able to save a picture from a website, this option seemed like a last resort. Obviously the name is a problem, it is to technical, but in addition to this there are other problems; When browsing a webpage, the LSK “Select” is used to interact with the page, but when in object mode the “m” menu is used. Although the LSK “Select” option is visible it is not in use. The “m” menu shows all actions applicable to the object but some of the actions seem wrong, like “Add to contact” when a picture is selected. It is also very annoying not to be able to exit the browser without first exiting the object mode.

Messages can be sent to a mail address on a website. When the mail address is selected a menu appears with two choices, “Add to contact” or “Send as”. To send a message to a mail address the “Send as” menu is chosen, and another menu appears, now whit only one choice, “MMS”. These two menus could be combined into one and one step to create a MMS is eliminated.

7.4 Unified Messaging

In the first screen, some folders (Inbox, Sent, and Template) are shown along with some functions (Sort, Delete all, Create new message). The order in which these are presented is not clear, the “Create new message” function is separated from the other three functions, why this easily is overlooked and thought of as a heading instead of a function. Furthermore it is not clear to see that these are functions at all since the folders have folder icons and the functions only have some small dot.

After a message has been opened, exiting from it (pressing RSK “Back”) requires an unnecessary step where details about the message is presented.

There is a consistency error in the “Create folder” menu. First and foremost the input field does not look like any other input field. The whole screen turns into a text edit field, which does not restrict the length of the folder name. Secondly the edited text is confirmed by the LSK where the term “Accept” is used. The term “Accept” is not commonly used throughout the applications, “OK” and “Select” are primarily used.

When editing a message (MMS) there are two instances of “Back”, one on the RSK and one in the “m” menu (or LSK “Select” since these seems to be the same). These two does not have the same functionality:

- if the “m” menu “Back” is used, exit from the message is done.
- if the RSK ”Back” is used the “m” menu is closed (if open, otherwise exit from message).

In addition, when back is pressed and the function is to exit from the message, a dialog is presented asking whether or not to save to drafts. There is no error prevention here. If “Back” is pressed by mistake there is no way to cancel the action and return to message editor. The only way to continue editing the message is to save it to drafts and open it again from there.

When editing a message, LSK “Select” is used to enter the menus of each content type (Add picture, Add text). However the function of “Select” when the text field is selected differs from when one of the other fields (picture, sound, video) is selected. Instead of showing a menu, the text field is directly entered and set in editing mode. There is a problem of consistency here.

The procedure to send a MMS is very complex. Very few of the test participants managed to actually send the message. This is probably a result of many factors:

- **Layout of the screen where recipients are added**
  This factor of course depends on the skin used. However, remarkably few of the participants could identify where to enter a phone number or, in one of the test scenarios, did notice that a recipient already was added. The layout of the screen seemed intimidating and confusing; many participants simply gave up when this screen appeared. Little effort was made to understand the contents of it.

- **Sequence**
  To send a message, first of all the “Send” option in the message editor has to be found. The test shows that this is not easy. Some of the participants were just lucky to find either the “m”-menu “Send” or the LSK “Select”-menu “Send”, which appears only when the selected element is any other than the text field. Once this has been located, the recipient editor screen appears and here the same procedure repeats itself. The “m” menu send has to be found.

- **Consistency**
  As several participants pointed out, a send menu choice in the message editor is expected. This should be consistent with the functions in the “Inbox” that are explicitly stated (Sort, Delete all, Multiple operations). Because of this the user does not expect to find such a basic function as “Send” in some menu. There is no way of telling where the different menu options will appear. Also, when the “Send” option has been found trough the LSK “Select” menu at one time the user expects it to appear whenever the LSK “Select” is used, in the text field is selected as well, which results in repeated entries into the text field.

- **Terminology**
  When choosing to send a message and the “Add recipient” screen appears, the LSK has the option “Ok”. Most participants construed this as an ok-to-send option and automatically pressed the LSK only to get a new menu containing only the one option “Add from contacts”. In the skin used during testing it was not evident that this was a menu post why many users failed to notice it and clicked LSK “Select” with a run-time error as a result. Clearly the “Ok” option is a bad choice of LSK option. Not only is it inconsistent with the commonly used “Select” option, it also invites to usage since send already has been chosen and “Ok” seems to be an confirmation of this.

The participants that did realise that LSK “Ok” here serves the same purpose as the LSK “Select” usually does, instead had a hard time figuring out how to add a desired recipient. The logical thing to do is to select the “Add recipient” option and then add a recipient by entering a number or choosing from a contact list. Instead there is, in addition to the “Add recipient” option (which only has one menu option), an “Enter number” option.

7.5 Content Manager

The content manager works well as a file manager and the test shows that most participants can relate to the conceptual model of a computer’s file system. However, it suffers the same problem of inconsistency in the menus and terminology as the other applications since there are no guiding norms to conform to.

An example of inconsistency is in the media player where the LSK has the function “Options” and the RSK has the function “Back”. Nowhere else the term “Option” connected to the LSK appears, at least not during the test.

In content manager, as in unified messaging, the possibility of selecting several elements exists. The difference between the functions is that in content manager it is called “Enter multiple selection mode” and in unified messaging it is simply called “Multiple operations”. Many of the participants had trouble understanding what multiple operations meant, why the former is the preferable choice.

7.6 Summary

The most obvious problem in this application suite is that there is very poor consistency in and between the applications. The main consistency flaw is in how the different menus differ, or rather not differ, in functionality. There are several ways to carry out the same action. If each of the menus had a distinct and understandable connection with what elements it affect it would be easier for the user to guess where to find certain functions. To give an example of this, look at the menus in the content manager; when a picture is selected the “m” menu contains options that affect the picture, like move and delete file. It also contains options that effects the folders, like create folder and switch view. The same thing appears when editing a MMS, the LSK “Select” menu contains not only actions on the selected elements but also actions that affect the whole message. When the user finds this menu it is not likely that the “m” menu is sought after or used, hence some functionality is unexploited.

An ideal solution to this problem is to reduce the choices in each menu to only contain actions of a certain domain and not include some of the options in both menus and others in just one of the menus. This solution applied to the MMS editor would cause the LSK “Select” in MMS editor to act as a menu for the selected object and the “m” menu to consist of actions concerning the whole MMS, like sending the MMS and setting the page timing. Of course this is only an example, perhaps there are other solutions but the main point is to have a convention that works for all menus so that the user recognises and can guess the menu contents.

In the unified messaging application yet another manner, apart from the menus, of carrying out actions can be found. This manner is to choose from explicit options found among the objects in a folder, like “Sort” and “Delete all” options. This style of putting the menu directly in the visible interface creates problems. First of all, the user expects to find these menu options in every folder even outside the unified messaging application. This constitutes a problem of consistency since they do not exist anywhere else. Secondly, when the folder contains many elements, messages, the options are located at the bottom and consequently out of the visible area. The user has to know that the options exist. Thirdly, the style makes it harder to distinguish the actions from menus since they both look the same, for example in the messenger menu the action “Create new message” looks the same as the menus “Area info” and “Settings”.

There exist many examples of poor consistency concerning the function of the soft keys and the menus. One example is in the multiple operations menu in the unified messaging application and the settings in the browser

application. In multiple operations the CSK and the LSK are both used to check/uncheck the messages and the “m” menu is used to apply the previously chosen action to the checked messages. In the browser settings menu (we tested the “Show images” setting) CSK is used to activate radio button and the LSK is used to confirm the selection. The best alternative is the second where the CSK is used as a select button and the LSK is a confirm or a menu button.

Worth mentioning is also the reactions to the mapping of the back function to the RSK “Back”. In western way of thinking, to go back corresponds to go to the left. This is a product of the way the languages are constructed and thus how reading is done, from left to right. This is obvious when comparing the browsing of web pages to reading the pages of a book. Hence, the back function may be better mapped to the LSK.

8. Test report 2

for Master thesis
at the Department of Design Sciences, LTH
06/04/2005

Evaluation of Test Setups for Goobi project suite

-version XXXX

Authors:
Oscar Cosmo
Catia Hansen

8.1 Background

In this evaluation the client is a leading company in the field of developing and selling information-related applications for mobile phones (e.g.: browsers, messaging clients, file/content management, digital rights management, etc). Traditionally the client has been focusing mostly on the protocol and standards aspects of such applications. Since operators and the market require much more focus on the user interface and usability aspects now and in the future the client requests an evaluation of test methods for an existing application suit and for coming releases.

Previously carried out evaluation of test method resulted in a large number of revealed usability flaws while variations to the method were limited and most purely experimental.

Several questions about the method arose that had not been considered for the first test session; is the simulator the best way to test, are there any benefits to having a test assistant in the room during the test, how do we improve the questionnaires? Ideas for improvement to the test method along with access to new hardware called for a second test using the same software suite but a new test approach.

8.2 Test method

Mainly there exist four types of usability tests; assessment-, comparison-exploratory- and validation test. None of the methods focus on how to achieve a satisfying test method, they all concern testing the product in different parts of the development life cycle [2]. Therefore the aim of the test is rather to see which mixture of factors (e g. environment, device, and support) that will constitute the best test setup for usability test of mobile applications suits.

The goal for this test method is to identify which of the test setups is recommendable to guarantee major usability flaws to be revealed and a certain level of usability to be incorporated in each new release. It is desirable that the designed procedure is re-usable in upcoming projects. Although the current status of the tested system version is *released*, the product test may in future tests be labelled as an assessment test since the objectives shall be to examine and evaluate the product in realistic tasks and identify specific usability deficiencies parallel with development of the product.

An assessment test is usually done during development and is characterised by the task based testing where the test person has little or no inter-

action with the test monitor. The tasks have to be carefully developed to provide just enough information to the test person. If too much information is incorporated in the task description it may affect the test person's manner of carrying out the tasks. If too little information is provided the test person may experience unnecessary difficulties understanding and solving the tasks at hand.

To find the best setup for evaluating mobile application suits, at least two dimensions are varied, order of test device and interaction with test monitor. First dimension, order of test device (straight or reversed), will be strongly connected to the technique of an informal comparison test which compares different designs and interaction styles. The scenarios will be carried out on following devices in noted order;

Straight Order: remote control, PC, and mobile telephone.
Random Order: mobile telephone, remote control, and PC.

Second dimension concerns varying the possibility to interact with the test monitor (ITM) which in some way can be associated with exploratory- contra assessment tests. Where exploratory test refer to a technique requiring extensive interaction between the participant and the test monitor, an assessment test insist on less or no interaction at all (No ITM) [2]. The concept of extensive test person/test monitor interaction is used to explore preliminary "concepts" and allows closer study of the test person's behaviour.

The test monitor will be present in the test room and interact with the test persons in the ITM group. The interaction will consist of feedback and encouragement to think aloud and discus ways of action. In opposite to ITM the members of No ITM will receive little help and with the test monitor not present in the test room, communication is only possible through an intercom. To ensure reliable results and to provide big enough groups for each variation to the test, twelve test participants are used.

	Straight Order	Reverse Order
ITM	*3*	*3*
No ITM	*3*	*3*

Table 12. Number of participants in each group. N=12.

Groups are numbered from left to right. Group #1 does the test in straight order with ITM, group #2 in reverse order with ITM while group #3 does

the test in straight order with No ITM and group #4 in reverse order with No ITM.

In this test the users' are not grouped by their background or expected level of knowledge. Answers from the briefing questionnaire will later show whether their background affected the test results.

A pilot test is carried out to identify any problems with the test plan itself. If there is an obvious problem with it, it will be revealed and the test plan can be revised. The pilot test follows the test plan but observers and test monitor focus on how the test itself works and how tasks and questionnaires work instead of focusing on usability aspects of the product tested. These results are *not* included in the evaluation.

8.2.1 User profile

A typical user of this system is difficult to identify since it has a very large target group. The tested version of the software is official, which allows the test to be performed on persons whom have no connection to the developing company, without the need for a disclosure agreement to be signed. Since the test environment is located within a technical university, it is easy to engage students in the testing procedure.

8.2.2 Test environment and equipment

Testing is carried out in the usability test laboratory at Ingvar Kamprad Design Centrum (IKDC) and facilitates a classic testing laboratory setup. Separated rooms for testing and monitoring characterize a classic testing laboratory setup. The separated rooms are divided by a one-way mirror, which allows the test person to be observed without feeling the direct presence of observers during the test.

Further measures to disguise the test performed are by creating a school setting instead of the sterile testing environment. This type of setting is typically static and this helps to ensure that the different tests will not be influenced by environmental disturbances. The test person is supposed to recognise the surroundings and feel comfortable with the situation. This aims to prevent the test person from acting like a test is performed and trying to fulfil expectations connected to the test, instead of using the device as intended. Since the user profile is indefinite and mainly students will be used to conduct the test a school environment is probably familiar to the entire group of test people.

The tested products are used in various situations but since the test is divided into three parts, using a PC, a PC with remote control, and a mobile phone, a school environment will be easy to recreate using existing furniture. School milieus vary between schools but the traditionally holds a whiteboard (or blackboard) and a couple of desks and chairs.

Test on PC (Setup A):
The test will be performed on a PC using a mouse. For entering text and numbers the user is allowed to use the keyboard. On the screen a simulation of a mobile phone, running the test software, is presented.

Test on PC with remote control (Setup B):
Evaluating mobile applications is a bit different from evaluating PC applications. To avoid operations normally connected to the conceptual use of a PC, for example using the right mouse button which is not supported on a mobile phone, the mouse is removed and replaced with a remote control. The remote control is slightly larger than a regular mobile phone handset and consists mainly of a touch sensitive LCD screen where the key set of a mobile phone is simulated. The screen contains, apart from the key set, an area with buttons which is not used in the test; therefore it should be covered during the test to prevent usage. Entering text is done using the keyboard.

On the face of the remote are some physical buttons not to be used in the interaction; these buttons will not be covered. The user will be instructed not to use them. Since the remote communicates with the computer through an infrared receiver it has to be fixed in a position that allows direct line of sight between the remote and the receiver and this position should also simulate the position relative to the screen of a regular handset; the remote is placed as close beneath the mobile phone screen as possible.

Test on mobile phone (Setup C):
For this test a Nokia 6630 with the test software suite installed will be used. A manoeuvrable surveillance camera is attached to the ceiling above the test person's position. To avoid unnecessary camera manoeuvring from the control room, the test person will be instructed to keep the mobile phone inside an area marked on the table.

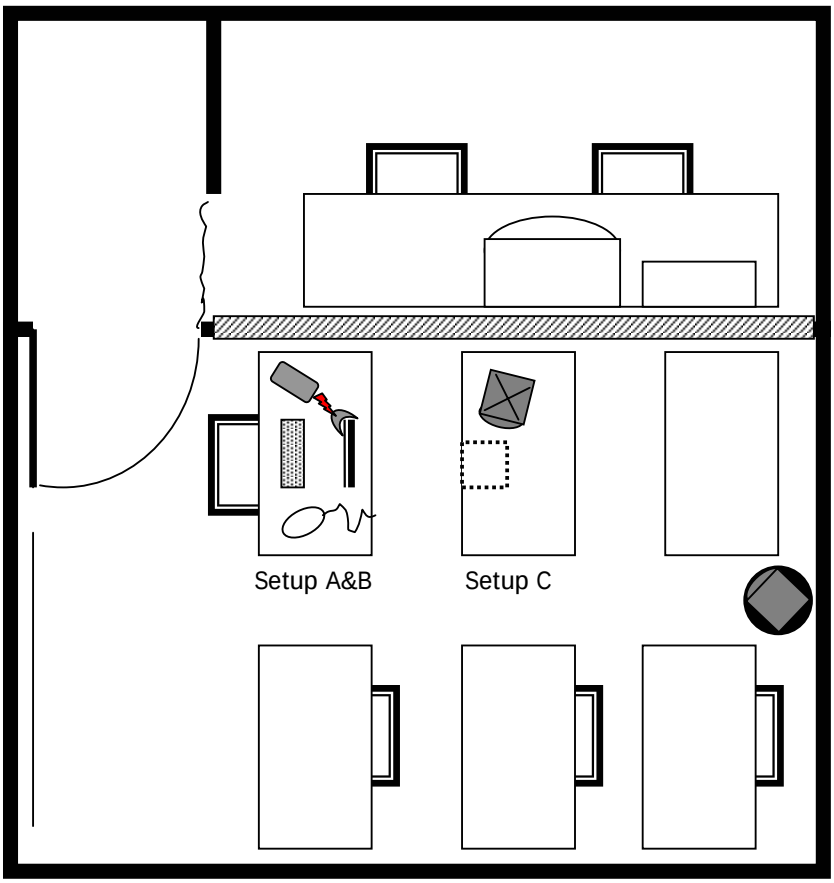

Table 13. Classic Testing Laboratory Setup with equipment for setups A, B and C.

8.2.3 Roles

During the test there will be different roles;

- **Test monitor:** The test monitor has the ultimate responsibility for the test in conduct.
- **Timer:** Incorporated in the test monitor role. Times actions specified in the test objectives.
- **Test assistant:** Interacts with test person. Introduces the test person to the test, objectives and goals. Aim to make the test person feel relaxed. Responses to help calls
- **Data logger/Recordings operator:** Responsible for collecting data and operate data collection instruments; video- sound recorders.
- **Observer:** Invited guests with special interest in the testing procedure. Observes the testing without interacting.
- **Product expert:** Responsible for technical aspects of the product being tested.

There will be two personnel sharing these roles, observers not included. The roles are divided so that interference between them is minimized; for example can the responsibilities of the recordings operator role interfere with the responsibilities of the test monitor role.

8.2.4 Performing the test

The test is divided into three parts, before, during, and after the test. Below each part is described by specific activities.

Before: Once the test person arrives the test assistant greets personally and by giving a calm impression the assistant should make the test person to feel comfortable and relaxed. The test person is asked to fill out a simple questionnaire gathering information about which mobile phone functionality the test person regularly uses, how frequent, and in what situations. At this time the test person is informed of the video/sound recording, observation wall, confidentiality of participation (test person will have to sign a paper as approval) and the golden rule for every test; it is the product being tested, not the test person.

When the questionnaire is filled out the assistant escorts the participant to the test room, explains the background and objective of the test according to a prewritten script. Also the test person is informed about some of the equipment used and the restrictions, such as not to use the buttons on the remote and to not move the mobile telephone out of the restricted area. The last thing the test assistant does before the test begins is to hand the

test person the task list, explaining that the tasks are to be solved without or without assistance depending on which group the test person is a member of, and that the test person is welcome to explain or comment every step taken to solve a task according to the “think aloud” principle.

During: Depending on which group the test person is a member of; different approaches will be taking during the test.

> **Variant I:** Once the test monitor has left the test room and observe from behind the mirrored wall and the video/sound recording has commenced, the test assistant will be instructed through the intercom to start the test. Then the test assistant tells the test person to start. Presence of the test assistant in the test room allows more interaction with test person. The test monitor is allowed to inform the test person when maximum time of completion is reached and instruct the test person to continue with the next task.
>
> During the test it is important to encourage the interaction and therefore the test assistant responsibility is to frequently as questions like; “what are you thinking right now?”, “what did you expect?” and “how did you solve that task?”. The test assistant also inform about the m-button on each device.
>
> **Variant II:** Once the test monitor and test assistant have left the test room and observe from behind the mirrored wall and the video/sound recording has commenced, the test person will be instructed through the intercom to start the test. In this variant the test person will receive no help. However the test monitor is allowed to inform the test person when maximum time of completion is reached and instruct the test person to continue with the next task.

Both variants will be divided in to two separate subdivisions; straight order or reverse order.

Straight Order:

1. Using PC steering with a remote control.
2. Short debriefing, to catch main thoughts/opinions about the test.
3. PC steering with a mouse.
4. Short debriefing
5. Mobile telephone,

Reverse Order:

1. Mobile telephone
2. Short debriefing, to catch main thoughts/opinions about the test.
3. Using PC steering with a remote control.
4. Short debriefing
5. PC steering with a mouse.

After: When the actual test has ended, because of maximum time of test is reached or all tasks fulfilled, the test person will together with the test assistant and test monitor discuss the test, aiming to capture the attitude towards the test and the applications. As a token of appreciation the test person is handed a small gift.

8.2.5 Data collection

During testing various data will be collected. Collection will be done automatically by video/sound recording and manually through questionnaires and logging during the test. Logging is supported by a spreadsheet where timestamps of interesting events is recorded. Data is either qualitative or quantitative and the way it is collected depends on what type it corresponds to and present phase of test.

Information about the test person's background is collected in the initial briefing questionnaire. During the test both types of data is gathered. Quantitative data by recording time of completion for specific tasks, time to recover from errors, time to realize and recognize an error, number and percentage of tasks completed correctly with and without assistance, number of errors, and counting incorrect selections. Qualitative data is gathered through study of the video recordings, observing the test person's behaviour and expressions; like hesitations, irritated behaviour, and quotes. Qualitative data is also gathered before the test person change device (for example from mobile telephone to PC).

8.3 Goal of usability testing

Request from the client is specified in a confidential document. A primary interest is a general evaluation of usability test methods for mobile application suits that can be applied to coming releases. During the evaluation it is desirable to obtain results about how well IKDC's usability lab fits evaluation of this kind of products, mobile phone applications.

Furthermore the client desires a method to efficiently test mobile applications and guarantee a certain level of usability to be incorporated in each new release. Procedures used during this test should be reusable in the testing of future development projects.

Specifically, different test devices and setups are evaluated and compared. Testing focuses on the participants' opinion of the different devices and performance of the individual groups. Also quality and relevance of collected data is evaluated.

8.3.1 Problem statements

Application related statements

- Does any result not already found in the application test appear during any of the variants tested?

Test related statements

- What is the best way to test mobile application?
- How well does each test method work for testing mobile applications?
- Which is the best equipment to perform mobile application tests on? (mobile phone, PC, remote control)
- How many test persons are needed to ensure reliable results?
- How many individual test needs to be done to capture the most important flaws?
- How does the environment affect the test? (Camera quality, camera position, disturbance from environment, test monitors location)
- Is the conceptual understanding affected by the device being used?
- Is the understanding for navigation affected by the equipment being used
- Which test persons are the representative end users? And how can it be verified?

- How does help from test assistant influence the test results?
- Does performed task order affect test results?
- How does think aloud affect the test?
- Should the test allow cooperation between test person - test person or test person – test monitor? Which is the best number of participants to test at each test?

8.3.2 Goals attained

To keep the size of the test, cost- and time wise, at a realistic level some of the test related statements could not be addressed. Conceivably the combinations of test variants to address every aspect would be far too large. Focus is to evaluate the possibilities of the specific test environment and the applicability of validated theories to it.

No results related to the application related statement were found.

8.4 Summary of test results

Results from this test focuses on which test setup that gave the best response from the test participants and how different test setup parameters affect the result, not to find usability flaws in the tested applications. Data concerning the participants' response to the setups has been collected by the briefing and debriefing questionnaires and discussion.

The participants are divided into four groups as explained in the test method chapter. Following is the participants sorted by group affiliation:

Group 1 (straight order, ITM):	Participants 1, 8 and 10
Group 2 (reverse order, ITM):	Participants 2, 3 and 12
Group 3 (straight order, No ITM):	Participants 4, 6 and 7
Group 4 (reverse order, No ITM):	Participants 5, 9 and 11

8.4.1 Briefing results

The briefing questionnaire is evaluated through a simple point scale applied to each question concerning previous experience with related technology. Hence questions 1 – 4, 6 and 11 are excluded. To rate the participants according to technical knowledge the scale is applied so that the highest point is given for the alternative that conceivably has the most positive impact on technical knowledge. An example; in question eight the participant is asked to estimate personal usage of a number of specific mobile phone functionalities along a scale from *never* to *every day*. Since the answer *never* has no positive impact on technical knowledge it is rewarded with zero points while the answer *every day* accumulates four points. The accumulated score represents an estimate on each participant's relative technical knowledge or experience.

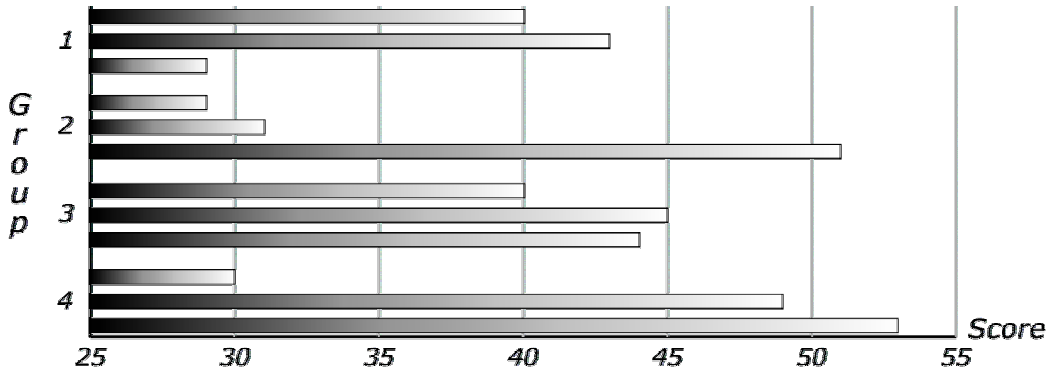

Diagram 17. Technical knowledge score extracted from briefing questionnaire sorted by group.

Presented below is a summary of the briefing questionnaires.

Question	Summary
1	Seven male and five female participants.
2	Ten of the participants filled out the 20-24 alternative, one 25–29 and one 30-40.
3	All of the participants were currently studying at university level.
4	Same as above.
5	Most of the participants (six) claimed to be very interested in new technology while five answered to be moderately interested. Only one marked the alternative for interest in future technology.
6	Alternative #3 (I buy a new product I want when I can afford it) is dominant with ten marks. Probably closely related to question 4. The two remaining marks are divided between alternative #2 and #4.
7	Only one of the female admits to have used Internet from a mobile device and one male claim not to have tried it.
8	Eleven participants use the telephone every day for ordinary voice calls and almost as many uses SMS just as often. Most never uses MMS or camera or seldom do so. High end functionality such as music- or movie player is never used, perhaps because most affordable mobile phones do not provide this.
9	Four participants claim using the Internet from an ordinary computer more then eight hours a week; none uses it less than 2 hours.
10	Participants using the Internet more than eight hours also use chat services or instant messaging more than eight hours a week. Two participants never use these services.
11	Windows is the most commonly used operating system among the participants followed by Unix and Linux. Only five of the participants ranked Mac OS and all ranked it as least used.

12	Seven of the participants could be considered to know the terms or at least to know enough to be able to guess accurately. All could explain what a MMS is accurately enough and most knew what a browser is.

Table 14. Summary of each question ordered by straight order questionnaire.

8.4.2 Debriefing results

Debriefing was carried out before switching test station after each section of scenarios was finished. Questions and discussion aimed to capture the participants' immediate reactions to the test station. Most of the questions in the questionnaire are of multiple choice type and only a few are discussion topics why these are left out of the statistic presentation of the results and instead a summary of the most common comments to each question is presented. Questionnaires for both straight and reverse order are found in Appendix C, beginning with the straight order questionnaire.

8.4.2.1 Straight / reverse order comparison

To reveal any differences between participants who carried out the test with the different set ups, answers to the questionnaire are compared between those who carried out he test in straight and reverse order and also between with our without the presence of a test assistant.

Below is a diagram of the debriefing result, comparing straight and reverse order. The left hand column of each question represents the answers from participants who did the test in straight order and the right hand column is the corresponding answer from participants who carried out the test in reverse order.

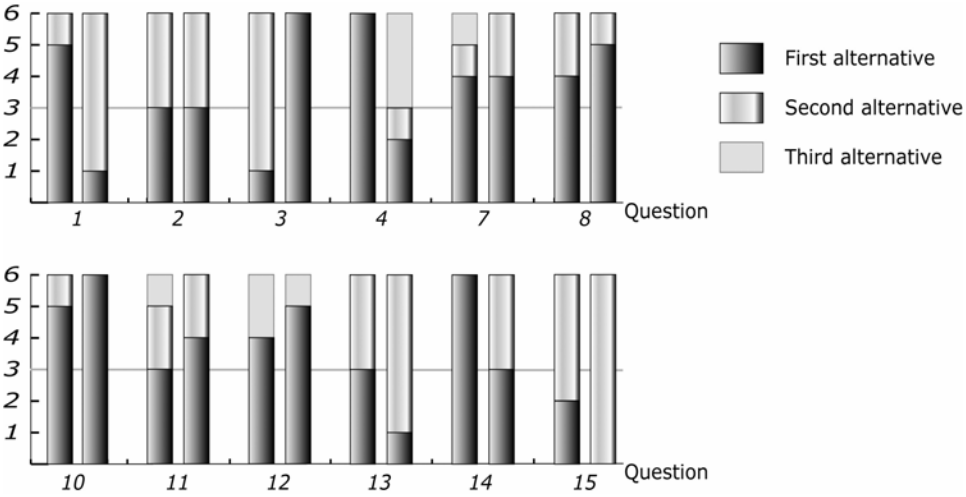

Diagram 18. Diagrams showing the debriefing result for straight and reverse test order. Question numbers as well as alternatives are taken from the straight order debriefing questionnaire; Appendix C. Alternatives correspond to the numbering of the answer alternatives in the answer alternatives in the questionnaire. N=12.

Noticeably questions 1, 3 and 4 display a big difference between straight and reverse order. Question 1 concerns the apprehension of the simulator as a phone or computer. Participants who did the test in straight order, simulator before mobile, answered that they thought of the simulator as a phone and not as a computer while reverse order participants thought just the opposite.

The result for question 3 indicates better understanding of the buttons' application among the reverse order participants.

Three of the reverse order participants could not answer or did not understand question 4, why the *don't know* alternative is overrepresented.

Another notable result is that when the participants are asked to rank the test devices, question 14, all of the straight order participants prefer the mobile phone while the reverse order participants are divided equally between computer and mobile phone. None of the participants preferred the remote control even though this alternative has some shares in the comparison between mouse and remote control, question 7.

8.4.2.2 ITM/No ITM comparison

Results presented in straight/reverse order comparison are affected by how much interaction with the test assistant the participants received. To weigh

these against each other, questionnaire results are compared between participants in the No ITM group, left hand columns, and participants in the ITM group, right hand columns.

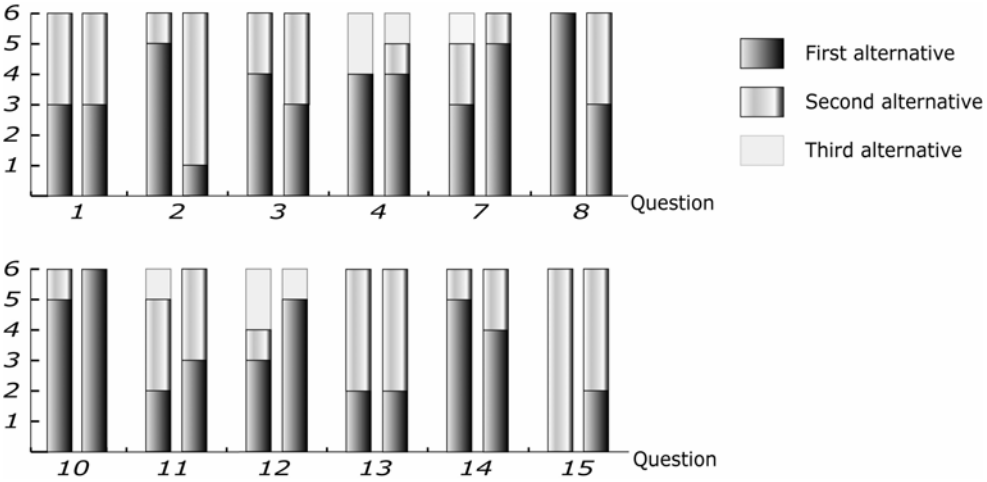

Diagram 19. Diagram showing the debriefing result for ITM and No ITM. Question numbers as well as alternatives are taken from the straight order debriefing questionnaire; Appendix C. Alternatives correspond to the numbering of the answer alternatives in the questionnaire. N=12.

In the above diagram the participants are evenly distributed in question 1 and 3. This implies that ITM or No ITM has little influence to the results of these questions previously found in straight or reverse order comparison (*Diagram 18*).

Conversely to the question 2 results found in comparison between straight and reverse order (*Diagram 18*), where participants are distributed equally across the groups, comparison between ITM and No ITM shows that ITM participants claims to understand the icons to a greater extent than the No ITM participants.

Unlike the result of question 14 in straight/reverse order diagram (*Diagram 18*), there is little difference between the groups in the ITM/No ITM comparison; the preferred device is the mobile phone in both groups.

8.4.2.3 Summary of debriefing discussion questions

A summary of the most common comments to the debriefing discussion questions, Appendix C, divided by test order. Quotes have been translated from original language.

Straight order

Question	Comments
5	"Not good". "Complicated". Inconvenient to move focus between screen and remote control. Some, though, claimed it felt more natural than a mouse, mouse phone-like.
6	Poor feedback. Hard to know what button is pressed. Very sluggish.
9	Smoother navigation. Better overview and nicer graphics. Looks more finished. Too many buttons!
16	Most claimed that they had no problem with their setup. If alone, that's good and if with test assistant, that's good. Participants in the ITM groups felt more secure, "if you mess up I know help is right beside me", while participants in the No ITM groups thought it would have been stressful with the presence of test assistant.

Table 15. Summary of the most common comments from straight order participants. N=6.

Reverse order

Question	Comments
1	"Easy to test on". Hard to navigate. Some of the participants said they were too used to their own mobile phone and it was hard to switch to this.
11	Hard to get an overview. Feels awkward, not natural. Very sluggish.
12	"Poor feedback". Hard to hit the right button.
17	Comments were much like in the ones in question 16 in the straight order setup.

Table 16. Summary of the most common comments from reverse order participants. N=6.

8.5 Source of error

To get reliable results the test setting and procedure are required to be exactly the same in all related test sessions as dictated in the test plan. Sometimes circumstances force a deviation from the test plan and these circumstances has to be taken into account during data analysis.

A deviation from the test plan was invoked when, during two tests, web access from the simulation phone was lost. Since two of the scenarios are depending on web access these scenarios had to be improvised. However the main focus of the test is not to reveal usability flaws but to test different test setups. Hence the scenarios could be changed and still give the participant enough experience to make a comparison of navigation between the different devices.

Software and hardware behaviour outside the scope of the test can influence the test participant's conceptual understanding of the test device and environment. Consequently this may alter the participant's behaviour and navigation.

When scenarios were carried out on the mobile phone some participants pressed the red button which causes the software to return to Nokia's start menu. Return to the main menu is the intended action but not the expected menu appears why conception of the test software's start menu as a main menu is lost. This caused the participants to avoid using this button and the natural way to solve a task and use the system is effected.

Some actions using the simulated mobile cause program failure, the simulator crashes. When this occurs, the participant's reliance in the simulator is damaged and otherwise naturally exploratory behaviour is restrained by fear of causing another system failure.

The test results present the remote control as the least preferred device among the three. One reason for this may be that it is not fully functional. Two of the buttons did not work on the remote and had to be pressed using the mouse and simulated key set on screen.

8.6 Analysis and conclusions

Analysis of the test results given the source of error. The analysis is based on the personal observation and apprehension of the test personnel and the statistical data presented in the *summary of test result* chapter. Personal observation and apprehension is very important since not all information about the results is displayed by the statistic results. Observation of the participants' reactions and discussions outside the scope of the briefing and debriefing helps to create a sense of how the test could be improved and what parts that did not work as intended.

8.6.1 ITM or No ITM

ITM or No ITM should be decided depending on what kind of data to be extracted from the test. The two tends to present different kinds of data, qualitative data presented by ITM and more quantitative data presented by the No ITM. Of course this is not black or white; No ITM presents some quantitative data as well but perhaps not as clear and to find.

Having a test assistant or monitor present in the test room during sessions allows closer observation of the participants' actions. However, for the test assistant to notice and put down every observation and reflection a very fast handwriting is required. Hence it is not recommended that the test assistant has a very observational role but simply functions as help and support to the participant. The test assistant may in that way be very helpful in stimulating aloud thinking and encouraging the participant to explain certain interesting actions and thoughts. This provides a great deal of qualitative data, attitude data, while the quantitative data may suffer from the fact that the interaction with the test assistant consumes time, may provide better conceptual understanding and above all; it is very hard to control the amount of help and the type of help each participant receive.

Some of the participants commented that it felt safe to have the test assistant in the room to help out if something went wrong and to correct mistakes. If this were completely true a more exploratory approach would have been more widely used among the ITM participants. Instead the opposite was observed. ITM participants seemed more eager to do the right thing and avoid getting stuck than the No ITM participants. Despite this ITM participants claimed not to feel observed or watched by the test assistant while the No ITM guessed they would have found it stressful with an observer in the room.

As seen in debriefing question two in the comparison between ITM and No ITM, most ITM participants claim to understand the icons. This seems to be

a result of interaction with the test assistant. Participants may have asked indirectly about the icons, been encouraged to reflect over the icons, or unintentionally guided by the test assistant to the meaning of each icon while the No ITM participants did not reflect over the icons or simply had trouble understanding the meaning of them.

Sessions carried out without the test assistant present in the test room proved to be very fruitful. Participants who were properly introduced to the environment and the think aloud principle provided lots of results, mainly due to adoption of the exploratory approach sought after in the ITM participants' behaviour. It is also easier to control the amount of help and the type of help provided from the test crew. A deliberation among the test personnel behind the one-way mirror is possible throughout the test.

8.6.2 The Device

Choosing which device to test on is not an easy thing to do. The question to ask is; how important is the recorded data? Since there is a big difference in how much of the user's action can be observed using a camera and monitor this question may very well decide which device to utilise. Findings in this test suggest that the simulation on a computer and navigation with a mouse presents the most lucid view while it is virtually impossible to perceive what the participant is doing when the remote control is used for navigation. When testing on the mobile phone the view of the participants' actions is to some degree limited by the placement of the camera relatively to the phone's display; however an adequate view was obtained during this test. It was easy enough to see what was happening on the phone display and which button the participant pressed even though some participants had to be reminded not to block the view by sitting leaned forward over the phone.

Most preferred device among the test participants in total was the mobile phone. Mainly due to its natural feel, this is how the software is supposed to be presented and used. When presented on a mobile phone display the system looks and feels more completed and invites to an exploratory approach when solving the tasks. When looking more closely at which group preferred the mobile phone the straight order group is unanimous while the reverse order group is divided. A reason for this may be that the straight order group had to learn how to navigate in the system using the simulation and therefore found this setup to be harder to use than the following. Conversely the reverse order group had to learn how to navigate on the mobile phone but still fifty percent preferred this device.

Also the reverse order group display a better understanding for the buttons than the straight order group, at least fewer of the straight order participants claimed to understand the buttons' functionality. This could be a result of the participants' mostly high accustomedness to mobile phones in general. Placement and look of the buttons provide clues to what their function is, even though when using this system the buttons were not properly mapped. On the simulated mobile phone the buttons has no individual feel and look except for the label under each button. Mapping can be done more accurately but the natural *in-thumbs-reach* feel (given the thumb is used to navigate on a mobile phone key set) is lost when a mouse is used. Furthermore the simulation is a non-existing phone which does not support the test person by previous experience with the key set or look. This has its benefits when testing software that is intended for multiple hardware configurations though, focus is on the software. However nothing in the test results show that the Nokia used should have influenced the participants' behaviour significantly. A reason for this might be that since the phone used is a recently released model and quite expensive none of the participants had any previous experience with it. A more common phone may have caused confusion among the participants since a concept of how to use it already exists.

Participants in the straight order group answered that the simulator is conceived as using a mobile phone while reverse order participants on the contrary answered that it is conceived as using a computer. Obviously the reverse order participants are influenced by using the real mobile phone why the same system on a computer feels more like using a computer. Even if the simulator was perceived as a mobile phone some actions more affiliated with the use of a computer occurred, such as double clicks. A problem with feedback can also be observed; sometimes participants clicked several times on the same button even though nothing happened the first time, sometimes in frustration and sometimes because experience tells them that sometimes one click is not enough. This seldom occurred when using the mobile phone.

8.6.3 Summary

Using the briefing questionnaire and adherent scoring system proved to be an efficient way of deciding how representative for the end user and suitable to the test each participant was.

During the test it worked very well with intermediary debriefings. The debriefings provided useful feedback and a pause for the participant as well as valuable information to the test team. In this case a natural break occurred when switching test device, in a regular test situation there might

not be such a natural reason to interrupt the scenario. An idea may be to divide the scenarios into smaller sections and distribute one section at a time with an interruption for intermediary debriefing.

Intermediary debriefings also offer a chance to stimulate usage of the think aloud technique. Some participants are by nature very quiet and to constantly remind them to think aloud not only cause unnecessary interruptions to the participant's problem solving but could also damage the participant's self-confidence. When the think aloud technique is properly adapted by the participants it may provide comments to actions otherwise hard to interpret and spontaneous comments about the user interface.

When it comes to the environment, most of the participants claimed not to reflect upon it at all. However, many reacted on the interrogation-room style mirrored wall and actually commented on it and many looked for the cameras both during the briefing when informed about it and during actual testing. If the environment affect the participants' behaviour is very hard to tell from this test but seemingly participants were conscious of that they were being observed and despite their claims not to be affected by this a slight tendency to nervous behaviour can be observed in some participants. Disturbances during the test such as the test assistant entering the room, however, seem not to affect the participants materially.

One thing to consider in future test is to prepare test participants by briefing in a different environment than the test room. The sterile environment may make participants uneasy and aware of that they are being observed. Use a more familiar environment like a cafeteria.

8.7 Recommendations

As mentioned above not all of the test related statements were addressed but it is very clear that different test setups achieve different results. The tested setups provide god results when applied on software with obvious flaws but to extract results from software designed with usability as first priority and hence no obvious flaws is a harder task which may call for more elaborate techniques.

Recommendations based on the addressed test related statements in this test are to carry out the tests on the mobile phone with No ITM. This will provide both satisfying quality to the recorded data and more easily conducted tests.

Some improvements to the setup are recommended. The participant's placement in the room relative to the mirrored wall and the camera can be improved to avoid the participants' negative reactions of an interrogation room and to straighten the image of the mobile phone display on the monitor.

8.8 Appendix

A. Briefing

SCENARIO

- för utvärdering av mobiltelefonapplikationer

Läs igenom följande noggrant!

Scenariebeskrivning
Detta är en beskrivning med uppgifter vi vill att du skall lösa. Föreställ dig de situationer som beskrivs och utför därefter de efterfrågade uppgifterna. Du har ingen annan att fråga utan får försöka lösa uppgifterna efter bästa förmåga. När du känner dig färdig med ett scenario, återgå till menysidan på telefonen. Vi ser gärna att du under testet tillämpar ”tänka högt” metoden, dvs. säger högt vad du tänker.

- Föreställ dig
- Lös uppgifter
- Gå tillbaka när du är färdig
- Tänk högt

Läs igenom följande noggrant!

Scenariebeskrivning
Detta är en beskrivning med uppgifter vi vill att du skall lösa. Föreställ dig de situationer som beskrivs och utför därefter de efterfrågade uppgifterna. Testassistenten kommer att finnas vid din sida under hela testet. Diskutera gärna uppgifterna och fråga om det är något du undrar. När du känner dig färdig med ett scenario, återgå till menysidan på telefonen. Vi ser gärna att du under testet tillämpar ”tänka högt” metoden, dvs. säger högt vad du tänker.

- Föreställ dig
- Lös uppgifter
- Gå tillbaka när du är färdig
- Tänk högt

Du har precis lyssnat på en tråkig föreläsning i denna sal. De andra entusiastiska deltagarna har gått ut för att ta en paus och diskutera sina nyförvärvade kunskaper. Eftersom du inte är fullt så intresserad, spenderar du tiden med att utforska innehållet i din nya mobiltelefon.

16. Du är lite nyfiken på webbläsaren som finns i telefonen. När du öppnar webbläsaren och tittar på sidan **www.efd.lth.se/~c01oc/Goobi** hittar du en bild du skulle vilja ha i din telefon. Spara bilden i din telefon.
    (Tips: tecknet tilde (~) skriver du genom att trycka på knapparna <Alt Gr> och <~> samtidigt, tryck därefter mellanslag så visas tecknet.)

17. Det vibrerar i din ficka. Du har fått ett meddelande från Kaj, läs det och följ instruktionerna.

18. Pelle vet inte riktigt vad content managern innebär, han har inte någon sådan i sin telefon. Hur skulle du förklara den för honom?

OBS!
Vänta på ytterligare instruktioner.

19. När du tittar på vissa webbsidor tar det lång tid att ladda dem. Det går fortare om man inte behövde ladda alla bilder på sidorna, det är ju ändå texten som är intressant. Gör en inställning som hindrar bilder från att visas automatiskt.

20. Du har fått ett meddelande från Lisa, läs det och följ instruktionerna.

OBS!
Vänta på ytterligare instruktioner.

21. På sidan **www.efd.lth.se/~c01oc/Goobi** finns adressen till webbmastern, använd den och meddela, med ett MMS, att du tycker sidan är inaktuell.
    (Tips: tecknet tilde (~) skriver du genom att trycka på knapparna <Alt Gr> och <~> samtidigt, tryck därefter mellanslag så visas tecknet.)

22. För att slippa ladda sidan varje gång du vill titta på den, kan du spara sidan och visa den ”offline”, dvs. utan att behöva koppla upp dig mot Internet. Spara sidan ”offline”.

Du har precis lyssnat på en tråkig föreläsning i denna sal. De andra entusiastiska deltagarna har gått ut för att ta en paus och diskutera sina nyförvärvade kunskaper. Eftersom du inte är fullt så intresserad, spenderar du tiden med att utforska innehållet i din nya mobiltelefon.

1. Du är lite nyfiken på webbläsaren som finns i telefonen. När du öppnar webbläsaren och tittar på sidan **www.efd.lth.se/~c01oc/Goobi** hittar du en bild du skulle vilja ha i din telefon. Spara bilden i din telefon.

2. Det vibrerar i din ficka. Du har fått ett meddelande från Kaj (se nedan). Läs meddelandet och följ instruktionerna.
   [Meddelandet]: ”Tjena! Kaj här. Kan du skicka en bild på Simon till mig; Anna vill se hur han ser ut.”

3. Pelle vet inte riktigt vad content managern innebär, han har inte någon sådan i sin telefon. Hur skulle du förklara den för honom?

OBS!
Vänta på ytterligare instruktioner.

4. På sidan **www.efd.lth.se/~c01oc/Goobi** finns adressen till webbmastern, använd den och meddela, med ett MMS, att du tycker sidan är inaktuell.
   (Tips: tecknet tilde (~) skriver du genom att trycka på knapparna <Alt Gr> och <~> samtidigt, tryck därefter mellanslag så visas tecknet.)

5. För att slippa ladda sidan varje gång du vill titta på den, kan du spara sidan och visa den ”offline”, dvs. utan att behöva koppla upp dig mot Internet. Spara sidan ”offline”.

OBS!

Vänta på ytterligare instruktioner.

6. När du tittar på vissa webbsidor tar det lång tid att ladda dem. Det går fortare om man inte behövde ladda alla bilder på sidorna, det är ju ändå texten som är intressant. Gör en inställning som hindrar bilder från att visas automatiskt.

7. Du har fått ett meddelande från Lisa, läs det och följ instruktionerna.

C. Debriefing

(A summary of the debriefing questions can be found in section Debriefing results starting on page 134)

Namn: Datum: Order: Straight

PC:

1. **Hur kändes det? (Som en telefon eller en dator?)**

   1. Telefon **O** 2. Dator **O** 3. Annat **O**

   Kommen-

   tar:_____

   _____

   _____

2. **Visste du vad ikonerna betydde?**
   1. Ja **O** 2. Nej **O** 3. Vet ej **O**

   Kommen-

   tar:_____

   _____

   _ _____

3. **Visste du vad knapparna betydde?**

   1. Ja **O** 2. Nej **O** 3. Vet ej **O**

   Kommen-

   tar:_____

   _____

   _ _____

4. Förväntade du dig att ”den” skulle fungera som en vanlig telefon.

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommentar: _____

_____

_____ _____

ÖVRIGT:

_____

_____

_____

_____

FJÄRR:

5. Hur kändes det?

_____

_____

_____ _____

6. Bortsett från att knapparna inte fungerade riktigt som de ska, hur tyckte du att fjärrkontrollen fungerade?

_____

_____

____ _____

7. Vilket var lättast, styra med mus eller fjärrkontroll?

1. Mus **O** 2. Fjärrkontroll **O**

Kommen-

tar:_____

_____

8. Såg du kopplingen mellan texten och knappen?

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommen-

tar:_____

_____

_ _____

ÖV-

RIGT:_____

_____

_____

_____

_____

MOB:

9. Hur kändes det?

_____

_____

____ _____

10. Tycker du att du hade lärt dig från förra ”stationen” hur du skull navigera?

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommen-

tar:____

____

_ ___________________________________

11. Kände du igen hur du skulle styra ”mobilen” för att du hade testat på den andra stationen precis eller för att man styr den så i din telefon?

1. Lärt av test **O** 2. Från egen telefon **O** 3. Annat **O**

Kommentar:_____

_____

_ _____

12. När var det lättast att styra telefonen/navigera. När du fick hålla telefonen i handen eller när den skulle ligga på bordet?

1. Hand **O** 2. Bord **O** 3. Vet ej **O**

Kommentarer:_____

_____

_ _____

13. Tänkte du medvetet på att du inte skulle skymma telefonen?

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommentar:_____

_____

_ _____

HELA TESTET:

14. Vilken station var lättast att testa på och varför?

1. Mob **O** 2. PC **O** 3. Fjärr **O**

Kommentar:_____

_____

_____ _____

15. Tänkte du på att du blev filmad?

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommen-

tar:_____

_____

_ _____

16. Hur kändes det att det satt någon bredvid dig? / Hur kändes det att sitta här inne själv?

_____

_____

____ _____

Namn: Datum: Order: Reverse

MOB:

1. **Hur kändes det?**

_____

_____

_____

_____

2. **Visste du vad ikonerna betydde?**

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommentarer:_____

_____

_____

3. **Visste du vad knapparna betydde?**

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommentarer:_____

_____

_____

4. **När var det lättast att styra telefonen/navigera. När du fick hålla telefonen i handen eller när den skulle ligga på bordet?**

1. Hand **O** 2. Bord **O** 3. Vet ej **O**

Kommenta-
rer:____
____
____

5. Tänkte du medvetet på att du inte skulle skymma telefonen?

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommentarer:_____
_____
_____

ÖVRIGT:_____
_____
_____
_____
_____

PC:

6. Hur kändes det? (Som en telefon eller en dator?)

1. Telefon **O** 2. Dator **O** 3. Annat **O**

Kommentar:_____
_____
_____

7. Vid vilken station var det lättast att styra ”telefonen”, MOB/PC?

1. Mobil **O** 2. PC **O**

Kommentar:_____

_____

_____

8. Tycker du att du hade lärt dig från förra stationen hur du skull navigera?

1. Ja **O 2.** Nej **O** 3. Vet ej **O**

Kommen-

tar:_____

_____

_ _____

9. Kände du igen hur du skulle styra ”mobilen” för att du hade testat på den andra stationen precis eller för att man styr den så i din telefon?

1. Lärt av test **O** 2. Från egen telefon **O** 3. Annat **O**

Kommen-

tar:_____

_____

_ _____

10. Förväntade du dig att ”den” skulle fungera som en vanlig telefon?

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommen-

tar:_____

_____

_ _____

ÖV-

RIGT:_____
_____
_____
_____
_____ _____

FJÄRR:

11. Hur kändes det?

_____

_____

_____

12. Bortsett från att knapparna inte fungerade riktigt som de ska, hur tyckte du att fjärrkontrollen fungerade?

_____

_____

_____

13. Hur var det lättast att styra PCn, med mus eller fjärrkontroll?

1. Mus **O** 2. Fjärrkontroll **O**

Kommentar:_____

_____

_ _____

14. Såg du kopplingen mellan texten och knappen?

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommentar:_____

_____

_ _____

HELA TESTET:

15. Vilken ”station” var lättast att testa på och varför?
1. Mob **O** 2. PC **O** 3. Fjärr **O**

Kommen-

tar:_____

_____

__ _____

16. Tänkte du på att du blev filmad?

1. Ja **O** 2. Nej **O** 3. Vet ej **O**

Kommen-

tar:_____

_____

_ _____

17. Hur kändes det att det satt någon bredvid dig? /Hur kändes det att sitta här inne själv?

_____

_____
_____

D. Data collection sheets

Task	Est TOC	Start Time HH:MM:SS	End Time HH:MM:SS	TOC MM:SS	PASSED	Termination (OK, ERR, MTC Quit)
1						
2						
3						
4						
5						
6						
7						

	Briefing Start HH: MM:SS	Briefing End HH:MM:SS
B1		
B2		
B3		

Comments:

Time	

E. Checklists

Checklists preparations

Test room

- ❒ Check equipment
- ❒ Check software
- ❒ Light on
- ❒ Rearrange furniture
- ❒ Camera arrangement
- ❒ Microphone
- ❒ Remote control
- ❒ Screen

Observer room

- ❒ Check equipment
- ❒ Lights off

Checklist equipment

Test room

- ❒ PC
- ❒ Nokia phone
- ❒ Cameras
- ❒ Microphones
- ❒ Remote control
- ❒ Chairs
- ❒ Desks
- ❒ Screen walls
- ❒ Briefing
- ❒ Scenario

Observer room

- ❒ Video recording equipment
- ❒ Sound system
- ❒ Tape
- ❒ Timer
- ❒ Debriefing
- ❒ Logging form
- ❒ Pencil

Checklist software

Simulator and mobile phone

- ❒ Storage is empty
- ❒ Reset show images in browser
- ❒ Enter address history contains www.efd.lth.se/~c01oc/Goobi
- ❒ Check www.efd.lth.se/~c01oc/Goobi so no network failures has occurred
- ❒ Remove saved picture in content manager

9. References

[1] Norman, Donald A. 1998. The *Design of Everyday Things*. The MIT Press.

[2] Rubin, Jeffrey. 1994. *Handbook of Usability Testing*. John Wiley & Sons, Inc.

[3] Preece, Jennifer & Rogers, Yvonne & Sharp, Helen. 2002. *Interaction Design – beyond human-computer interaction*. John Wiley & Sons, Inc.

[4] Lauesen, Søren. 2002. *Software Requirements – Styles and Techniques*. Pearson Education Ltd.

[5] Weiss Scott. 2002. *Handheld Usability*. John Wiley & Sons, Inc.

[6] Kiljander, Harri. 2004. *Evolution and Usability of Mobile Phone Interaction Styles*. Helsinki University of Technology.

[7] Scoop Eye-Innovation Channel Vol.02. 2003-08-25. *LCD Driver ICs–Keeping pace with the ever-changing mobile phone*. 2005-05-10. http://www.necel.com/en/channel/se/e_se0825_1.html

[8] Research Methods Knowledge Database. 2005-05-13 http://www.socialresearchmethods.net/kb/datatype.htm

[9] Holler, Robert. *Mobile Application Development: A Natural Fit with Agile Methodologies*. 2005-05-13 ttp://www.versionone.net/pdf/MobileDevelopment.pdf

[10] Norman, Donald A. 2002. *Emotion & Design: Attractive Things Work Better*. 2005-05-12. http://www.jnd.org/dn.mss/emotion_design_at.html

[11] Izdebeski, Jacek & Johnsson, Stefan. Bachelor thesis. 2005. *User-Centered Design i Industrin – en kartläggning inom användbarhet.* LTH

[12] SDSU. Center for Learning, Instruction, & Performance Technologies. 2005-05-10. clipt.sdsu.edu/posit/tx/gloss.htm

[13] ACM Special Interest Group on Computer-Human Interaction (SIGCHI). 1996 (Updated 2004-06-03)*. Curricula for Human-Computer Interaction*. ACM, Inc. 2005-05-10. http://sigchi.org/cdg/cdg2.html#2_1

[14] Sommerville, Ian. 2001. *Software Engineering, 6th Edition*. Pearson Education Ltd.

[15] ISA. 2004. *Twenty Years of Mobile Phones.* 2005-05-15. http://www.obigo.com/PSUser/mediacache/1702/1710/20_years_mobile_nov_2004_low.pdf

[16] *White Paper, The Story of Obigo.* 2005-01-10. http://www.obigo.com/PSUser/mediacache/1702/1710/The_Obigo_Story.pdf

[17] Mobilforum. 2005-05-09. http://www.mobilforum.se/?nav=facts_history

[18] *The Mobile Phone User Guide-History: Radio*. 2005-05-12. http://www.mobileshop.org/history/radio.htm

[19] Smith, Brian*. The Story of Reginald Aubrey Fessenden*. 2005-05-03 http://www181.pair.com/otsw/Fessenden.html

[20] Yahoo India. *SMS History*. 2005-05-03. http://in.mobile.yahoo.com/smshistory.html

[21] Antenn. ELFAs Tidning Vol.4. 2004. 2005-05-14. http://www.elfa.se/antenn/arkiv/4_04.pdf

[22] Radio National. *The Science Show*. 2003-06-07. ABCOnline. 2005-05-12. http://www.abc.net.au/rn/science/ss/stories/s871809.htm

[23] UMTS World. *UMTS / 3G History and Future Milestones.* 2005-05-10. http://www.umtsworld.com/umts/history.htm

[24] Oxford City Council. 2005. *How Does A Mobile Phone Work*. 2005-05-12. http://www.oxford.gov.uk/services/mobile-phones-work.cfm

[25] *Nobel Lectures, Physics 1901-1921*, Elsevier Publishing Company http://nobelprize.org/physics/laureates/1909/marconi-bio.html

[26] Connected Earth. *The First True Mobiles.* 2005-05-12. http://www.connectedearth.co.uk/Galleries/Frombuttonstobytes/Mobilecommunications/Thefirsttruemobiles/index.htm

[27] Nerlich, Kjell. 2001. *Telegrafins Historia.* ESR. http://www.scag.se/prov/cwhist.htm

[28] Löwgren, Jonas. 1993. *Human-Computer Interaction – what every system developer should know.* Lund: Studentlitteratur.

[29] Natt och Dag, Johan – Madsen, Ofelia S. 2005. *An Industrial Case Study of Usability Evaluation*. LTH.

[30] Preece, Jennifer - Keller, Laurie. 1990 *Human-Computer Interaction: selected readings.* Prentice Hall International Ltd.

[31] Course Material for the course *Human-Computer Interaction*. LTH. 2004-12-05. www.eat.lth.se/Kurs/Material/MAM061.

[32] Olsson, Caroline. 2004. Master's Thesis: *Design & Development of Web-based Information Systems for Port Operations.* LTH