

Contextual Usability

Rigour meets relevance when usability goes mobile

Tomas Lindroth, Stefan Nilsson

{tomas, stefan}@laboratorium.htu.se

Laboratorium for Interaction Technology
University of Trollhättan/Uddevalla
Östergatan 18
451 26 Uddevalla
Phone: +46 522 65 60 00
Fax: +46 522 65 60 99

Keywords: Testing issues:fb0602, testing:fb06, user oriented design:fa08,hci:hc0101

Abstract

When we do traditional usability tests on applications using stationary computers the context is controlled and not especially relevant. The computers in the labs are more or less in the same context as when they are used in offices and homes. But for mobile devices, testing might make the result irrelevant since it fails to take the context of its use into consideration. The purpose of this study is to evaluate the usability testing methods and theories from a mobile perspective. This is to find out if and where the conventional usability methods fail and what they fail to detect when applied to mobile devices. How can the usability methods of today be extended to facilitate the testing of mobile devices in its right context? This paper is based on our previous studies and gives a brief overview of our field of work so far. It is written in two parts so to speak where the first part ends at chapter: Further research and that is also where the second part begins. Further research is a rather long chapter with our plans for future studies, it begins with further studies based on the findings from the article and ends with a scenario for our next study.

1 Introduction

The mobile device is seen as remote control for business and pleasure where you can buy, sell, control and supervise any gadget or situation. Without designers with the proper knowledge about HCI (Human Computer Interaction), Information- and User centered design there is high risk of usability flaws sneaking in to the design. With mobile solutions based on devices such as PDAs and cell phones the design of the gadgets and their interfaces are crucial factors for success. If designing for the web is hard with different browsers, screen sizes etc, try designing an interface on a screen with the size of half your credit card that might be used on the run in a dark alley with the rain pouring down. It is a possible scenario, mobile really means mobile, and it really means anywhere, on the bus, at the beach or in a storm. Testing of a new website is a must with different browsers, connections and users. But testing in front of a computer in a controlled environment is one thing, testing for mobility another. Usability testing in a laboratory with controlled situations and tasks works for applications used in stationary solutions. In the lab there is possibilities for video recordings with sound, screen captures, observers and controlled tasks. As expressed by Johnson (1998), this works fine with solutions where the context and environment is of second interest. Now, think of usability testing of a mobile solution where context is a factor.

“It is cold and snowing and you do not know from where your bus leaves in 5 minutes. You pick up your WAP phone to check: The mobile user run to catch her bus, after her run three researchers with cameras and microphones.”

1.1 Purpose

We will in this article explore current usability methods that are used today to test usability aspects of stationary computers. We will also conduct interviews with experts in the fields of mobility. This is done to explore the strengths and weaknesses of the usability methods used today and to aid us in the proposed design of a new way of testing the usability of mobile gadgets in the future.

2 Method

2.1 Scientific approach

We decided to use several different methods to be able to find answers on our question at issue and to be able to achieve our purpose with this article. In our case we used a multi dimensional method, Method Triangulation (Repstad, 1988), which is a kind of hybrid model. We have chosen to combine qualitative interviews, methodology studies and case studies.

2.2 Course of action

The methods used in this study are mainly of qualitative character. We have done interviews, case studies and “on spot” observations. Along with this we have done literature studies concerning the usability field. Books, scientific articles and websites have been our main information sources.

On basis of drawn conclusions and from the tests, we formulated our questions for the interviews that we were going to do. We decided to send out the questions by email. All participating persons were so geographically diversified so we did not have the possibility to meet them in person, mainly because of lack of money. Beside this, some of them were abroad during the time when we were ready to conduct such interviews. The questions were few and quite simple to answer in a few lines, therefore we decided not to spend time, money and effort on telephone interviews.

2.2.1 Method Triangulation

Triangulation is the use of different research methods or sources of data to examine the same problem. If the same conclusions can be reached using different methods or sources then no peculiarity of method or source has produced the conclusions and one's confidence in their validity increases. (Lwin, 1997) There are also other strengths of using method triangulation. These are for example generalizability and method independence. (Sawyer, 2000)

There are different types of triangulation. According to Lwin (1997) these are “Data triangulation”, “Investigator triangulation” and “Method triangulation”. Data triangulation refers to the collection of different data on the same phenomena, for example involving several participants and looking at different phases of fieldwork. Investigator triangulation is data collection that involves more than one researcher, and method triangulation means collecting data using different methods, methods that entail different threats to validity. (Lwin, 1997)

The three methods we used for the triangulation in this article were literature study, expert interviews and empirical tests of usability methods.

2.2.3 Methodology studies

It is of great importance that you study all available methods, when you as a researcher are entering a new scientific area and are trying to develop new methods or evolve old ones. We did a thorough methodology study and came up with twenty-two different methods that were applicable in our case. We analyzed them and tried to sort out methods that would give us a good picture of what the different types of methods did/did not measure in different contexts.

2.2.4 Formal interviews

The formal interviews that we have performed have been structured and sent out by e-mail. The problem with interviews like these is that different people can interpret the material in different ways. (Galtung, 1967) We think that this is not the case in our study since they have been sent to people that are experts in the area of mobility and usability. The questions have been of such character that they have only given their own personal thoughts/opinions about the questions at issue.

Quotes from the interviews have, if needed, been transcribed and/or translated into English. In this process we tried to stay as close as possible to the original meaning of the statement.

2.3.2 Discussion around the chosen method

It is difficult to adopt and understand a brand new scientific field. None of the authors had ever worked with usability testing when we decided to do this study. To be able to achieve enough knowledge about this field in such short time we realized that we had to use several different methods. The formal interviews gave us a very good ground to start out from and we saved a lot of time since the interviewed persons guided us in the right direction from the beginning. By doing a methodology study on available usability methods we learned how to conduct test, what traps you could fall into, what you can/can not measure in different situations and so on. The case study gave us on hand experience about how to conduct test, how to use a usability laboratory, what problems that can occur, how the lack of context affect the test situation and so on. We think that this was the best way to conduct a study like this.

A legitimate question at this time is if we could have done this study in a different way. The answer is off course, yes, but to what cost regarding time, money and knowledge? In the beginning of this study we thought of some different ways of how to conduct our case studies. We talked about the possibilities to do usability test with reporters and journalists but we realized quite soon that we did not have the right equipment for a study like that. We also talked about the possibilities to spy on people when they are using their mobile devices (Weilenmann, Larsson, 2000). Most methods that we could think of were not possible to perform, mostly because of the fact that we could not watch over the users in a satisfactory way.

3 Available methods (current usability methods)

In this chapter we present all of the applicable methods that we found. They constitute the ground for our methodology study when it came to deciding what sort of usability tests we were going to do.

They belong to three different areas:

- Inspection and evaluation
- Testing
- Inquiry

3.1 Inspection and Evaluation

Method	Purpose
Heuristic Evaluation (Nielsen & Mack, 1994)	Identify usability problems early in the design phase. Guidelines vs. design. You can provide the experts with paper mockups, or even just design specifications, and still get a good amount of usability problems discovered before actual work begins.

Cognitive Walkthrough (Rowley & Rhoades, 1992), (Spencer, 2000), (Wharton et. al., 1994)	Motivating how or why a person would react in a certain situation. Based on assumption about the users background, knowledge and goal. Great for early stages of development because they can be performed using just a system specification as a basis.
Formal Usability Inspection (Kahn & Prail, 1994) (Freedman & Weinberg, 1990), (Gilb et. Al., 1993), (Wheeler, 1996)	A way to detect errors in the code that the design relies on and documentation defects. The inspector performs tasks and reports any found errors and the lines of code causing the problem. The technique is design to reduce the time required to discover defects in a tight product cycle. Great for early stages since the inspector can work with merely a specification or paper mockups.
Pluralistic Walkthrough (Bias, 1991)	Looks into how users react in different situations. Includes user's, developer and usability experts. Best used in the early stages of development, as the feedback garnered from pluralistic walkthrough sessions is often in the form of user preferences and opinions.
Feature Inspection (Nielsen & Mack, 1994)	Find out if the feature of a product meets the users need and demanding. Best used in the middle stages of development. At this point, the functions of the product and the features that the users will use to produce their desired output are known.
Consistency Inspections (Wixon, et. al., 1994), (Nielsen, 1995)	Looks for consistency across multiple products from the same development effort. Best used in the early stages of development, when the initial development work has not progressed to the point where products that require extensive changes to ensure consistency will not require total overhauls.
Standards Inspection (Wixon, et. al., 1994), (Nielsen, 1995)	Standards Inspection ensures compliance with industry standards. Best used in the middle stages of development, as the actual design is being developed with the given standard in mind.
Guideline Checklist (Wixon, et. al., 1994), (Nielsen, 1995)	Guidelines and checklists help ensure that usability will be considered in a design. Usually, checklists are used in conjunction with a usability inspection method. The checklist gives the inspectors a basis by which to compare the product.

3.2 Testing

Method	Purpose
Thinking Aloud (Lewis, 1982), (Dumas & Redish, 1993), (Lindgaard, 1994), (Nielsen, 1994), (Rubin, 1994)	Lets the evaluator understand how the user views the system. The method can be used in any stage of development. Gives a lot of qualitative feedback during testing.
Co-Discovery Method (Dumas & Redish, 1993), (Lindgaard, 1994, (Rubin, 1994)	Idealistic for evaluating groupware programs, CSCW products and other products designed to be used by workers in team environments. Can be used during any phase of development.
Performance Measurement (Nielsen, 1993), (Dumas & Redish, 1993), (Lindgaard, 1994, (Rubin, 1994)	Measures whether a usability goal is reached or not e.g. a kind of bench marketing. Should be used in initial stages of design to provide benchmarks for the design process. It is also used during the design cycle to measure the work done thus far against those benchmarks.

3.3 Inquiry

Method	Purpose
Contextual Inquiry (Holtzblatt & Beyer, 1993), (Holtzblatt & Jones, 1993), (Beyer & Holtzblatt, 1995), (Beyer & Holtzblatt, 1997),	Contextual inquiry is used to get a broad knowledge about the environment that you are producing the program or device for. This technique is best used in the early stages of development, since a lot of the information you will get is subjective--how people feel about their jobs, how work or information flows through the organization, etc.
Ethnographic Study/Field Observation (Hammersley & Atkinson, 1995), (Wixon & Ramey, 1996)	Ethnographic Study is used to get a broad knowledge about the environment that you are studying. This technique is best used when you are studying complex situations where ordinary methods would miss to detect important details, for example "unspoken acting" i.e. tacit knowledge. A lot of the information you will get is subjective--how people feel about their jobs, how work or information flows through the organization, etc.
Interviews and Focus groups (Greenbaum, 1997), (Nielsen, 1997), (Templeton, 1994)	This technique can be used at any stage of development, depending on the questions that are asked. Interviews and focus groups are often held at very early stages of development thou, when the product requirements are still not firm. Focus groups are then held to extract user requirements prior to initial design.
Customer Research Groups (Lynch & Palmiter, 2000)	Customer Research Groups is an effective alternative to focus groups with the same purpose. (Se above)
Journalled Session (Nielsen, 1993)	Journalled sessions bridges usability inquiry, where you ask people about their experiences with a product, and usability testing, where you observe people experiencing the product's user interface. This technique is best used in the early stages of development, probably even pre-development, where the information you are attempting to gather is more preferential than empirical.
Incident Diaries or Self-Reporting Logs (Nielsen, 1993)	Finds out what kind of problems a user has had during a period of time or what they have used the system/device for.
The Valuation Method	Finds out how important a feature is to a user.
Logging use (Nielsen, 1993)	Gathers information about use and problems without the user knowing about it.

4 Theories

4.1 Rigour vs. Relevance

According to Mason (1988) there exist two primary attributes of knowledge producing activities in controlled experiments. He identifies them as: tightness of control and richness of reality. These attributes are taken generally to be in opposition to one another at the same level of knowledge, called the iso-epistemic curve. Hence, researchers must ultimately make a trade-off between them.

The larger the number of factors that is under control in an experiment, the more scientific rigour is emphasized. The more natural like the experimental setting is, the more relevant and applicable the results will be. (Järvinen, 1999)

4.2 User Centred Design

To make usable products, tools and applications there are several methods and theories that help the designer to reach her goal. They all focus on the user, her needs and requirements. User Centered Design (UCD) is a process that puts the user and her tasks in focus from the very beginning. The alternative to UCD has been a problem based approach where the user has to fit the solution, UCD demands deep understanding of the users needs and goals (Shneiderman, 1998). A big part of UCD is the iterative design cycle where a solution is designed, tested and modified repeatedly like a spiral (Rubins, 1994). The focus in this article is not UCD in particular but a part of it is the testing and evaluation of the software, device or other product that measures the usability of the same.

UCD means many things and goes by different names, but they are all names of the same concept, design that focus on the user (Rubins, 1994)(Nielsen, 1993). According to Nielsen usability is part of UCD and stands for the evaluation, change and improvement of a system, product or gadget Usability are not UCD, but one of the techniques to secure a user centered design.

4.3 Mobility

Defining what mobility is can be a difficult task (Kristoffersen & Ljungberg, 1999). In one sense, we are always mobile, we move around all the time. In another sense, while working in front of a computer, we are bound by our computer to sit in one place. Dahlbom & Ljungberg (1998) introduces three modalities of mobility – wandering, visiting and travelling.

They argue that while sitting in front of a computer on a desk, we are indeed mobile, we are wanderers. They say that while this setting can be seen as static, we still move. For example we walk in a corridor to get coffee.

The second modality of mobility is visiting. They give an example of a consultant that works at a client, using a computer there.

The third modality is people traveling between sites, for example by bus or car. They face an unpredictable context, usage of a mobile device for a traveler can be anywhere in the world.

When we talk about mobility we mean the usage of mobile devices anywhere and everywhere. A truly mobile device should support all three of these modalities (wandering, visiting and travelling). (Ljungberg & Kristoffersen)

4.4 Usability

Usability is the process of testing with a handful of techniques to gain learnability, efficiency, memorability, less errors and satisfaction (Nielsen 1993). These five attributes are the basics of usability engineering according to Nielsen (1993). There are others with their own definition of attributes like Rubins (1994) for instance. He outlines four similar attributes, usefulness, effectiveness, learnability and attitude (Booth, 1989 in Rubins, 1994). These are similar to Nielsens but with a slightly different definition. Without further discussion we choose Nielsens definition because it is the most widely known of these two (Olsson, 2000).

- Learnability

It should be easy to learn a new system so the user can start working quickly.

- Efficiency

A system should be efficient to use so the user achieves high productivity.

- Memorability

A casual user should not need to re-learn between times, the system needs to be logical.

- Errors

The system should stop the user from doing errors and if the user makes errors she should easily be able to recover.

- Satisfaction

Using the system should be pleasant. The user should want to return and like to use the system.

Here we use these five attributes as our definition of usability engineering. Any method or theory that supports and enhance one of these attributes would fit into the description of Usability Engineering. These attributes and theories are meant to support rigour.

In the method part above a vast amount of different methods was lined up with a short description of how they work. These methods are merely tools to measure the five attributes above. The product of the different tests is for some methods lists of errors made and for other methods it is videotapes from where you can collect user statements and interesting observations.

5 Empirical Study

In total we had about twenty different methods to use and from these we tested three methods that ranged from laboratory environment to group discussions.

We produced three different tasks to evaluate the usability methods selected. They were designed to be carried out on a PalmV. Each test was quite simple and we estimated that the whole test would be carried out in less than thirty minutes. The tests were not supposed to be used as a test of the PalmV, but rather a tool for us to explore the limits of a static laboratory when it comes to mobile devices. It was also a tool for us to see what information we missed when the mobile device was used in a natural environment. A researcher with usability experience approved the tests that were to be carried out.

The first task was to add a person to the address book. The second task was to schedule two different lessons that were occurring every other week repeatedly for a period of twenty weeks. The last task was to create a business card. The user supplied their own personal information and transmitted their business card over to another PalmV.

5.1 Performance Measurement

The first method that we evaluated was Performance Measurement (Nielsen, 1993). We engaged five users to participate in our usability tests in the usability laboratory in Aalborg. They ranged from beginners to experienced user and they had very different backgrounds, from a Spanish music composer to an English architect. There were four men and one woman.

The users participated on voluntary basis and they were told that they could interrupt the test at any time if they felt uncomfortable. Before the test took place we introduced them to the laboratory and showed them how the equipment would be used. They were allowed to "play" around with it, all this to make them comfortable. We also explained that they were not the subjects of the test, rather we were testing the method.

The laboratory consists of three rooms. One control room where all the technical personal is sitting and controlling the cameras and other effects like background noise and so on. One more control room where the test leader is sitting and doing the recording. The test leader is in control of the test situation and helps the user if some problems occur. The control rooms are placed on each side of a test room. They are separated by windows and were sound isolated.

When the user said that they were ready we lead them into the test room. Inside there, we told them what they were allowed to do and not. In our case they had to sit in a special angle to the table and they were not allowed to move the PalmV outside specified marks on the table. The three tasks that they were going to do were presented on a laptop in front of them. All usability tests were conducted in one day and recorded on digital video (DV). After the test we asked the each user if we could keep the business card that they had transmitted to us in the last task. We also asked them if we could contact them by mail if we needed to ask the questions that we did not think of during the time we worked with them. All of them were, fortunately, happy to participate.

5.2 Co-Discovery Method

The second method to evaluate was Co-Discovery Method (Dumas, Redish, 1993, Rubin, 1994, Lindgaard, 1994). We gathered four new participants. We used the three tasks once again as a tool for evaluating the method. The users sat down at two tables and formed two groups. Each group were given the tasks and told to perform them in pairs on one PalmV. They were told to speak out loud during the test. The tests were recorded on DV (Digital Video).

5.3 Pluralistic Walkthrough

The third and last method that we evaluated was Pluralistic Walkthrough (Bias, 1991). We gathered a new group of PalmV users; in total there were three participants. They ranged from intermediate to advanced users. Once again we used the three tasks as a tool for evaluating the method. We, the authors, acted as moderators and usability experts. Our role was to look at the users while they were performing the tasks and to ask them questions about what they were doing. The users were told to talk out loud and keep up a discussion about what they did and why. After each task we asked them if there was anything to remark upon and if they thought that the task would be able to perform on the run. We also asked them if they would have done it another way if they were on the move. The whole test and discussion was recorded on MD (Mini Disc).

5.4 Expert Interviews

The expert interviews were all conducted through an e-mail based question form. The questions were more of in the character of "thoughts", and we asked the selected persons to comment on these thoughts. This was done to better explain to them the theme of the study. Since all were professionals working in the field of mobility and usability, they all had a deep insight into the field of this article. The "thoughts" we presented to them can be seen as to have influenced them in their answers, therefore compromising the validity and reliability of the interview. Our view is that since this is a group of people with long experience in the theme, they all already have clear view of their field and does not get influenced by our thoughts on the subject.

In September 2000 we sent out e-mails to five researchers within the field of mobility/usability and asked them if they wanted to participate in an interview about Mobile Usability. In early November the four questions were sent out and we asked them to answer before Christmas Eve. We received answers from all the recipients with thoughts and reflections.

The answers was mainly what we had expected and was very much in line with our own thoughts and presumptions, but with some more depth and experience. One of the purposes of asking researcher already in the field was to balance our own lack of experience of fieldwork.

6 Discussion

6.1 Method Evaluation

It became clear to us rather soon that a lab like that was not designed to test mobile gadgets. We had numerous technical problems related to the small size of the gadgets. The cameras used in the laboratory were unable to get a good focus of the gadget. And when we had managed to get an acceptable view of the gadget, we could not move it since it then had been moved out of scope for the camera. We also had problems with the lighting in the laboratory. It constantly gave us reflections in the mobile gadget's display, and thus we could not see what the user was doing with it. This forced us to place the gadget and the person using it in an unnatural way that was nothing like the way they normally would use it.

Another problem not directly related to the technology used was that the test subjects had to read the instructions of what to do in the task. This clearly differs from real world use of a mobile device. You do not always get information that is going to be put into the mobile device in written form.

The time to perform a task varied greatly amongst the users. Also, the subjects learned from each other while performing the tasks. This test was performed indoors in a controlled office environment. The authors often asked the subjects if they would perform the tasks in another way if they had been outdoors, or if they were doing other things at the same time. The answer wearied from task to task, but many times the subjects answered that they would do the task completely different on the run.

This shows that the users use the gadget in different ways depending on the situation. The mobile gadget might work fine in the office environment without stress or other contextual challenging factors, but this does not say much about how it might work in different situations on the run.

6.2 Thoughts of findings

The goal of traditional usability to increase learnability, efficiency, memorability, less errors and satisfaction would still be the same, but needs to be applied to new or modified methods in a mobile situation. Many of the methods mentioned above would be difficult if not to say impossible to use in an open environment but how can we modify these methods to work in a mobile scenario to create the possibility to gather the data we need?

Using a method like Pluralistic Walkthrough where you ask the users to solve a couple of tests, encouraged to talk to each other and solve the test while the researchers asks questions, revealed in our case, a lot of bad design within the software of the product. These discussions are easy to record but the problem with mobility is that it is mobile! Mobile users make it hard to record and store conversations, to do that you need wireless microphones that might feel uncomfortable for the user to wear. You also need video to record how the user handles the device physically and that is not an easy task if you, at the same time, want to capture what happens on the screen. You also do not want to interfere with the user in any way. In doing so you would undoubtedly alter the way the user reacts in an given situation.

In this case it is not the methods that needs to be modified but rather our data collection tools that needs to be reconsidered.

When we apply usability methods we try to measure how usable an object is in a given situation. We harvest the data that the method is digging out from the situation and try analyse it for proper understanding and how to make a more usable product. But is the method bringing the right data to the surface or are we missing something out?

As Fagrell (2000) express it:

I believe that it is more important to establish techniques to capture and evaluate IT use concepts. This is in contrast to the typical CHI community usability study that quantitatively compares the speed of use between two systems. The types of usability study (in a wider sense) that I like is validation in practice.

[Fagrell, Henrik 2000]

So what is it that we miss out in a mobile situation? With the Palm V that we made our tests with it was obvious that the time it took to do a certain task was not paid enough attention. In a real situation when you are writing down a person's address in the Palm while he stands in front of you, seconds feels like minutes.

Also we had trouble with how we would let the users read the task list. The user's concentration was totally focusing on the Palm and on the paper with the tasks during the test. In that situation the task-paper becomes a major actant that do not exist in the real world. In a mobile situation there would be an even greater problem if the user would hold the paper in his hand!

There is of course workarounds to these problems and maybe you only need to be a little creative to solve them. Our suggestion, that we have not tested, is some type of role-play where the user is told to walk down the street and interact with the people contacting him. The people confronting him on his way are of course part of the test. They take on different roles such as an old classmate that the user has not met for a while and the classmate (actor) gives the user his address for him to put into the palm.

We also see a need for methods inspired of ethnographical methods where we observe the user and the use of a mobile device in a real world situation. This could be done in several ways. One of the most common would be to let the user observe her self and write it down at a daily basis in a diary. This is one of the methods used in Nielsen and Ramsay's evaluation of WAP in September 2000 (Ramsay, 2000). Taking it a bit further, the next thing to do would be Weilenmann's method of listening to and watching the user when using the mobile device without their knowledge (Weilenmann, Larsson, 2000).

We believe that it is in these types of situations where the device is used in the right context, on the run, while interacting with others and while being carried that you find another set of problems. It also depends on the purpose of use and if the situation for example is under pressure or not.

"Give the palm to, for example, a nurse or doctor at a hospital who were forced to use it as an journal or something, and you will find other faults. If i were to use it right now I do it in one way, pick it up in half an hour and continue. But if the patient could die, it would have another consequence and you would find other types of faults in the gadget."

[Skov, Mikael, 2000]

With these solutions for testing in context there is a loss of what we here address as rigour. We loose control over the given situation where the actual test is taking place. The number of factors that possibly affects the test increases and might affect the result in unpredictable way. Though we do not see this as a major drawback. We see control and rigour as a very important factor but not at the price you have to pay when you loose relevance.

Most of what is mentioned above might sound obvious for the experienced usability professional. With only a little creativity you would probably think of alternative methods when doing usability tests on mobile devices. But if you study existing literature you will find very little of this creativity in the usability books. There are examples in articles, like Weilenmann, Larsson (2000) and Ramsay (2000) but nothing gathered in book form that we could find.

7 Further Research

This study makes a very good ground for further research within the usability field. Mobile devices will be even more common in the near future and we see a great need for a different design. We will, in the next step of our journey, evaluate our methods of practice mentioned above and compare the result with traditional methods. When that stage, the second, is finished there should be enough empiric knowledge to start creating a framework for design of mobile devices. In future work we will also concentrate of the combined techniques of a pda and a mobile cell phone. The framework we will try to develop is targeted towards this hybrid of a communication device and a digital filofax.

With methods such as technomethodology, developed especially for the design of artifacts and generalization of human behavior, we will try to define the framework for the design of this hybrid personal mobile device.

In the near future we plan to re-do above mentioned tests in a bigger scale, with at least 20 participants. This time we will be focusing both on the method and the result of the test. This is done to get additional data about the methods and also to harvest data about the actual use of the artifact. This time also mobile phones will be evaluated with the traditional methods.

In parallel with the new tests of pdas and mobile phones with traditional methods there will be additional tests with above proposed methods such as as role-play, diaries and direct observation.

Role-play is a method sometimes used when designing new artifacts were the test subjects do not have a mental model of such a “non existing” device. The devices we plan to evaluate are existing and we do not use this method because of a weak mental model but rather because of the traditional methods lack of context awareness. A role-play could look like this:

“We are standing in front of the local shopping mall. The test subject is told that she will walk through the mall and interact with the persons that confront her. As she walk through the crowed equipped with a Palm 5 a person approaches and says: - Hello, is that really you??? Linda??? Oh, I haven't seen you since 5th grade, but I have to catch a bus, beam me your address and give you a call... Here she hopefully picks up the Palm and beam the address over”

During this conversation someone is recording the interaction on video for later analyses. From this we expect to gain knowledge of how persons handles the palm under stress and in a quite real situation where we still have the possibility to record the event. We are still in the development of this test and it might be re-designed at a later state. Does it work? The who lives will find out!

Diaries will be used because wants the user to reflect over their use of the device and compare this to how they actually use it in role-plays and in direct observation. The user will write in this diary for two weeks where we also will provide a cell phone or a PDA. If the user is not used to handling such a device we will give a short introduction of critical functions. This because we do not want them to stop using the device because of poor usability. In this case we are not primarily

interested in how to make the actual device a more usable product but rather how to make such device truly mobile. To direct the users comments in the direction of mobility we will provide some short questions to consider when writing.

The direct-observation method is quite simple in theory, but intrusive and the ethical aspect can be discussed. When we say direct observation we mean observing the user without the users knowledge at a café, on the bus or at a shopping mal. Then we record this with either video or just simple notes. From this we hope to gain real use that we can compare with the data from the other methods.

Problems we will encounter is in many ways related to selection of everything from mobile devices to users. The devices we choose to use will have a great affect on the users actual use! For example, a Palm 5 affords a different use than an iPAQ and the use of a Motorola cell phone will differ from the use of an Ericsson.

When doing traditional tests we have the possibility to choose our respondents. This means that we can have a target group of , let say, technique savvy persons between the ages 15-30. In direct observation it is much harder to have this sort of selection because we do not know whom the user is.

To be able to refine the methods we propose above we need to observe both real settings, like the real observation, and in settings we control. Below is a is a proposal for future studies.

7.1 Application of methods: Mobile e-learning, anytime, anywhere

7.1.1 Thoughts, challenges and motivations

To be able to test above mentioned methods and thoughts in a good way we will design a system where we have the possibilities to observe mobile users and their behaviors. A possible scenario is outlined below. The scenario is basically about using the possibilities in modern mobile devices for situated learning. That is using a PDA as a multimedia device to interact and trigger thoughts with the user.

7.2 Scenario

A possible scenario for todays type of mobile multimedia device with educational capabilities could today involve a PDA like the Compaq iPaq, a wireless LAN at an airport or hotel like Telias Homrun and a university who delivers the course. Or if we just offer it to students, we do not need the wireless LAN, only the cradle to be able to sync. There are many types of courses that could benefit from this type of device, here we will describe on possibility. To understand certain concepts it is better to discuss and experience them than to hear about them in a classroom. For instance, concepts as affordance, conceptual models and conventions are not always easy to explain. But everyday we are exposed to them without noting. At the University of Trollhättan Uddevalla there is a 2 credits course about these topics. The examination of the course involves reading the book *The Design Of Everyday Things* (Norman, 1988) and related

articles. The course will now be enhanced with recordings of seminars that will be delivered through movie clips. Also speeches could be delivered with related PowerPoint-like slides. To be able to get this material at a daily basis there should be synchronization cradles at the entrance of the school where students could download next day's tasks and videos. The students that would benefit the most from this kind of service would be commuters from Gothenburg or other places where you could go by bus or train. To be very concrete the course could be delivered by a combination of media's. Movies could be compressed using Microsoft's Streaming media file format where you can adjust the movie to the limits of the iPaq in a very easy way. Also available is Flash for Pocket PC, a limited version of the very common browser plug-in from Macromedia. With these different types of medias we can deliver sound, video and animations for the student to watch at slow moments. As we said, we believe that the material has to be somewhat amusing to really attract the students. When traveling you have several different choices, you can sleep, read the newspaper or as we suggest educate your self. The material that we offer has to be provoking or very informative to avoid the feeling of boredom, common in classroom settings. Often you have to attend to class and students feel an obligation to do so, we doubt that if the material is boring they will feel the same way about the mobile classroom. As outlined by Cadiz et al (1999) an important aspect of education and learning is interaction with tutor and other students. This means that there need to be very good collaboration possibilities for the students. We see a scenario where the student are online at all times with their PDA and are able to watch a video sequence in synchronization with other students where they watch the video clip at the same time. Now they would have the possibilities to pause and discuss the video via the PDA. This is not possible today but there should be good possibilities to simulate this situation with a wireless LAN.

7.3 Learning

We learn in every situation either we want it or not, in schools and everywhere else (Säljö, 1999). But in everyday life we do not always have someone to ask (help), and we cannot stop the experiment (as in school) to start over. But if we can combine these aspects of learning and of everyday life would we increase the quality? If we could enter a "learning space" when we are wandering or traveling (Dahlbom, 1998) and still moderators, teachers and discussants would be available, would we use such a service?

The type of mobile learning we want to explore opens up for a new possibility - "contextual learning". Contextual learning would open possibilities to interact with the surroundings in different ways. Take photos of bad usability for your HCI course, as an example. But there is also a different view of this where the teacher can give another type of tasks, questions and exams. There could be tasks where you should deliver an interview within the hour or observe mobile phone behavior during your day with a report at the end with pictures, movies etc. There are too many factors involved to say anything about if people will use the service or not. One part is how we frame the problem that we try to solve.

Also, as history tells us, when new technology is introduced into learning environments there has always been a short burst of enthusiasm over how revolutionizing this new technology will be for education. For example, motion pictures was seen as a revolutionizing educational tool in the mid-20th century, but was soon discarded (Berg, 2000) The same happened for the early uses of computers in education. Much was promised, but it never lived up to their expectations.

Since mobile users often tend to use their mobile gadget as a tool and as entertainer there should be some thoughts about the physical- and the software design of the mobile artefact. The discussion about users and visitors within the field of HCI separates users from visitors depending on the goal with gadget. If the goal with the device is leisure and entertainment the application should accommodate the visitor rather than provide good usability (Hedman, 2000)(Nielsen, 1999). This basically means that usability is still important but there should also be a certain level of fun.

7.4 Challenge

We must not directly transfer the use of computer aided education used with stationary computers directly into the mobile device. We probably need to do very specific changes, due to for example the smaller screen, limited input and context of use (could be anywhere). But not only do we need to take away features when we go from a stationary computer to mobile devices. We also need to study what a mobile device adds from a stationary computer and take those additions into consideration (the most obvious feature is the nature of mobility of the device). We need to have an understanding of the mobile device as a medium. As Berg (2000) says when it comes to computers as educational tools, "...what is unique about the computer medium? What are the specific advantages for education." We haven't yet mastered the stationary computer as a tool for education. Transfer this quote and focus it towards mobile computers and mobile devices, what are the specific advantages of a mobile device for education? Then we have a question that is not easy to answer. We probably need to have several ways of accommodating learning in context, taking into considerations all situations that might occur for a travelling e-learning person.

For the problem of providing the feeling of a classroom, we need some sort of awareness of the other people taking the course. We need both asynchronous and synchronous awareness to both emulate and enhance the environment of a classroom. The synchronous awareness would simulate "seeing" the others and participate in real-time ephemeral discussions on the topics of the course. The asynchronous would on the other hand enhance the classroom with historic and persistent discussions, where the participants can go anytime and discuss and read what others have contributed with. With a multimedia device this could be sound clips and videos illustrating a problem, you can watch the clips live or view them later, something that is not common in traditional education.

8 References

Berg (2000) Gary A. Berg, Human Computer Interaction (HCI) in Educational Environments: Implications of Understanding Computers as Media. *Journal of Educational Multimedia and Hypermedia* (2000) 9(4), 349-370

Beyer, Hugh, and Holtzblatt, Karen, "Apprenticing with the Customer: A Collaborative Approach to Requirements Definition," *Communications of the ACM*, May 1995.

Beyer, Hugh, and Holtzblatt, Karen, *Contextual Design: A Customer-Centered Approach to Systems Designs*, 1997, Morgan Kaufman Publishers, ISBN: 1558604111

Bias, Randolph G., "The Pluralistic Usability Walkthrough: Coordinated Empathies," in Nielsen, Jakob, and Mack, R. eds, Usability Inspection Methods, 1994, John Wiley and Sons, New York, NY. ISBN 0-471-01877-5

Cadiz et al JJ, Bakachandran A, Sanocki E, Gupta A, Grudin J, Jancke G. Distance Learning Through Distributed Collaborative Video Viewing. ACM Library, 1999

Customer Research Groups (2000)

Invented by Gene Lynch and Sue Palmiter of Design Technologies.

<http://stc.org/pics/usability/topics/focusgroups.html>

Dahlbom, B. and F. Ljungberg: Mobile Informatics, Scandinavian Journal of Information Systems, 1998, 10(1&2):227-234

Dumas, JS, and Redish, Janice, A Practical Guide to Usability Testing, 1993, Ablex, Norwood, NJ, ISBN 0-89391-991-8

Freedman, Daniel, and Weinberg, Gerald M, 1990, Handbook of Walkthroughs, Inspections, and Technical Reviews : Evaluating Programs, Projects, and Products, Dorset House, ISBN: 0932633196

Galtung, J. Theory and methods of social research, Universitetsforlaget, Oslo, 1967

Gilb, Tom, Graham, Dorothy, and Finzi, Susannah, Software Inspection, 1993, Addison-Wesley Pub Co, ISBN: 0201631814

Greenbaum, Thomas L., The Handbook for Focus Group Research, 1997, Sage Pubns; ISBN: 0761912533

Hammersley, Martyn & Atkinson, Paul 1995: Ethnography. Principles in practice. London: Routledge.

Hedman(2000) Anders Hedman, Visitor Oriented Design – Three Studies of Visitor Accommodation and a Call for Action. In proceedings of NordiCHI2000, Stockholm October 23-25 2000

Holtzblatt, K., and Jones, S. "Contextual Inquiry: A Participatory Technique for System Design." in Schuler, D., and Namioka, A. (eds.) Participatory Design: Principles and Practice. Lawrence Earlbaum, Hillsdale, NJ. 1993

Holtzblatt, Karen, and Beyer, Hugh, "Making Customer-Centered Design Work for Teams," Communications of the ACM, October 1993.

Johnson, P. (1998). "Usability and Mobility; Interactions on the move." First Workshop on Human Computer Interaction with Mobile Devices. GIST Technical Report G98-1. 21-23rd May 1998. Department of Computing Science, University of Glasgow, Scotland.

Järvinen, Pertti, (1999), *On research methods*.- Tampere, Finland: University of Tampere.

Kahn, Michael, and Prail, Amanda, "Formal Usability Inspections," in Nielsen, Jakob, and Mack, R. eds, *Usability Inspection Methods*, 1994, John Wiley and Sons, New York, NY. ISBN 0-471-01877-5

Kristoffersen, S. and F. Ljungberg. *Mobile Use of IT*, In the Proceedings of IRIS22, Jyvaskyla, Finland.

Lewis, C. (1982). Using the "thinking aloud" method in cognitive interface design. IBM Research Report RC 9265. IBM Thomas J. Watson Research Center, Yorktown Heights, New York: Author.

Lindgaard, G., *Usability Testing and System Evaluation: A Guide for Designing Useful Computer Systems*, 1994, Chapman and Hall, London, U.K. ISBN 0-412-46100-5

Lwin, CPD (1997) Lwin, Thein *Designing A Research Study* (1997)
<http://www.students.ncl.ac.uk/thein.lwin/edd1.html>

Mason, R. O. (1988) *Experimentation and knowledge – A pragmatic perspective*, *Knowledge: Creation, Diffusion, Utilization* 10, No 1, 3-24

Nielsen, Jakob, *Usability Engineering*, 1993, Academic Press/AP Professional, Cambridge, MA ISBN 0-12-518406-9

Nielsen, Jakob, and Mack, R. eds, *Usability Inspection Methods*, 1994, John Wiley and Sons, New York, NY. ISBN 0-471-01877-5

Nielsen, Jakob, "Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier, 1994" online Web page at
http://www.useit.com/papers/guerrilla_hci.html.

Nielsen, Jakob, *Usability Inspection Tutorial*, 1995, CHI '95 Proceedings

Nielsen, Jakob, "The Use and Misuse of Focus Groups" 1997
<http://www.useit.com/papers/focusgroups.html>

Olsson, C, (2000). "The usability concept re-considered: A need for new ways of measuring real web use". Proceedings of IRIS 23, *Laboratorium for Interaction Technology*

Ramsay, M. and Nielsen, J. (2000). "WAP Usability – Déjà Vu: 1994 All Over Again". Nielsen Norman Group. California, USA

Repstad, P. (1988), *Närhet och distans*, Lund: Studentlitteratur

Robinson, B, Doing Case Studies in Educational Research and Evaluation in 'The Open University (1992). Research and Evaluation'

Rowley, David E., and Rhoades, David G. "The Cognitive Jogthrough: A Fast-Paced User Interface Evaluation Procedure." CHI `92 Proceedings, (May 3-7, 1992):

Rubin, Jeffrey, Handbook of Usability Testing, 1994, John Wiley and Sons, New York, NY ISBN 0-471-59403-2

Shneiderman, B. (1998) Designing the User Interface, third ed. Addison-Wesley Longman, Inc, Massachusetts, USA

Spencer, Rick. " The streamlined cognitive walkthrough method." CHI 2000 Proceedings, (April 1 - 6, 2000):

Säljö. Lärande i praktiken, Studentlitteratur, 1999 Sweden.

Templeton, Jane F., The Focus Group : A Strategic Guide to Organizing, Conducting and Analyzing the Focus Group Interview, 1994, Probus Pub Co; ISBN: 1557385300

Wharton, C., Rieman, J., Lewis, C., and Polson, P., "The Cognitive Walkthrough Method: A Practitioner's Guide." in Nielsen, Jakob, and Mack, R. eds, Usability Inspection Methods, 1994, John Wiley and Sons, New York, NY. ISBN 0-471-01877-5

Wheeler, David A. (Ed.), Software Inspection : An Industry Best Practice, 1996, IEEE Computer Society, ISBN: 0818673400

Weilenmann, Alexandra & Larsson, Catrine (2000) On Doing 'Being Teenager', Proceedings of IRIS 23, Laboratorium for Interaction Technology

Wixon, Dennis, et. al., "Inspections and Design Reviews: Framework, History, and Reflection," in Nielsen, Jakob, and Mack, R. eds, Usability Inspection Methods, 1994, John Wiley and Sons, New York, NY. ISBN 0-471-01877-5

Wixon, D., and Ramey, Judith (Eds.), 1996, Field Methods Casebook for Software Design, John Wiley & Sons, Inc., New York, NY. ISBN: 0-471-14967-5

