Presentation of criteria in the choice of user testing method with the focus on mobile applications in the field of healthcare

Mostafa Danaei
Yared Habte Yohanes

Institutionen för informationsteknologi
Department of Information Technology
Abstract

**Human computer-interaction - Presentation of criteria in the choice of user testing method with the focus on mobile applications in the field of healthcare**

*Mostafa Danaei, Yared Habte Yohanes*

**Purpose** – The purpose of this thesis was to investigate which criteria that is of importance when choosing a user testing method in the field of healthcare. Furthermore, the authors wanted to examine which user testing methods that are appropriate for mobile applications in the field of healthcare.

**Method** – The methods used in this thesis were literature studies and semi-structured interviews. The qualitative research was conducted through interviewing seven researcher and professionals within the field of human computer-interaction that have been involved with healthcare. Literature studies consisted of literature, scientific articles, and websites within the topic.

**Findings** – The findings consist of appropriate criteria that aids in choosing user testing method in healthcare. Furthermore, suggestions are presented for which user testing methods that could be used in the context of mobile application in healthcare.

**Implications** – The study results showed that there are gaps in knowledge on what aspects are important when choosing a user testing method in healthcare. Furthermore, there is lack of research about which user testing method that is applicable for user testing in healthcare. This thesis was aimed towards researchers and professionals that work in the field of human computer-interaction. Moreover, the thesis is also targeted at healthcare organizations, and companies who develop technology for healthcare.

**Limitations** – The thesis is limited by time-constraints. The initial plan was to investigate which criteria that are of importance, and then to choose a user testing method to test a new healthcare mobile application. To validate the results of the thesis, the authors wanted to conduct a user testing session with users to see if it works in practice. Another limitation is that there was only one interview conducted with a full-time UX-designer. The plan was to get insights for both participants’ groups: researchers and professionals in HCI.

**Keywords** – Criteria, healthcare, user testing, mobile-applications, methods, human, physical limitations, cognitive limitations, elderly, human computer-interaction.
1. Introduction ............................................................................................................................................ 1
   1.1 Background ....................................................................................................................................... 2
   1.2 Problem definition ............................................................................................................................ 3
   1.3 Scope and definitions ........................................................................................................................ 3
   1.4 Purpose and research questions ........................................................................................................ 3
   1.5 Disposition ........................................................................................................................................ 4
2 Theoretical framework ............................................................................................................................ 5
   2.1 Usability in general ........................................................................................................................... 5
      2.1.1 User testing in general ............................................................................................................... 6
      2.1.2 User testing in mobile devices ................................................................................................... 6
   2.2 User testing methodologies ............................................................................................................... 8
      2.2.1 Heuristic evaluation ................................................................................................................... 8
      2.2.2 Focus groups .............................................................................................................................. 8
      2.2.3 Cognitive walkthrough .............................................................................................................. 8
      2.2.4 Think aloud ................................................................................................................................ 9
      2.2.5 Participatory Design .................................................................................................................. 9
      2.2.6 Individual interviews ...............................................................................................................10
   2.3 Physical limitations among elderly people ..................................................................................... 11
   2.4 Users of new healthcare technology ............................................................................................... 13
   2.5 General criteria for user testing methodologies .............................................................................. 14
      2.5.1 Framework for choosing usability evaluation method based on effort and cost ...................... 15
      2.5.2 Criteria for choosing user testing methods .............................................................................. 16
      2.5.3 User testing method comparison ............................................................................................. 19
3 Method and implementation .................................................................................................................. 20
   3.1 Choice of method ............................................................................................................................ 20
   3.2 Research approach .......................................................................................................................... 21
   3.3 Literature review ............................................................................................................................. 21
   3.4 Data collection ................................................................................................................................ 22
      3.4.1 Qualitative interview ................................................................................................................ 22
   3.5 Data analysis ................................................................................................................................... 23
      3.5.1 Interview analysis .................................................................................................................... 24
   3.6 Reliability and credibility ............................................................................................................... 25
4 Findings ................................................................................................................................................. 26
   4.1 Participant information ................................................................................................................... 26
   4.2 Themes of all seven interviews ....................................................................................................... 27
   4.3 Empirical data interpretation ........................................................................................................... 30
5 Analysis ................................................................................................................................................. 34
   5.1 Research question 1: What are appropriate criteria when choosing a user testing method, in the field of healthcare? 34
5.2 Research question 2: Which user testing method is applicable in the context of mobile application in healthcare?

6 Discussion and conclusions

6.1 Results

6.1.1 Ranking of criteria

6.2 Implications

6.3 Limitations

6.4 Conclusions and recommendations

6.5 Future work

7 References
1. Introduction

Users’ approval and management of interactive healthcare applications have often been hindered by their poor design, because of the difficulty and the complexity of healthcare systems. Interactive computer systems designed without considering healthcare professional information processing, are more than likely to disappoint their users. Poorly designed systems can even lead to catastrophe if vital information is not presented in an effective manner (Jaspers, 2009). Han et. al. (2005) and their study on child mortality at a hospital showed that when the hospital shifted to a computerized physician order entry (CPOE) system, child mortality increased compared to the period preceding the shift. Johnson et al., (2005) argues that health care applications must be attentively built, and it is important to consider the different backgrounds and tasks of health care staff. Johnson et al., (2005) also mention that it is not enough to meet the standards and models defined for the professions, but health care applications must also be intuitive and easy to use. The importance of mobile and web applications are increasing for healthcare, for example to support social interactions (Arslan, 2016). According to Arslan (2016) more and more people are using smartphones, to access various applications and services such as managing health data or generating collective actions for a healthier lifestyle. Therefore mobile applications developers need to consider several criteria when developing healthcare applications.

When developing mobile applications, it is vital to take into consideration the target group and their special needs. This is especially true when considering patients within healthcare. It is more likely for patients with a healthcare contact to have both cognitive and physical limitations (Boulos et al., 2011). Boulos et al. (2011) describes two different terms: early adopters and later adopters. Patients have different attitudes to technology, early adopters are people who are used to new technology and have no problems learning about new technology and trends. Later adopters of new technologies may not use them in the same way as early adopters for a number of different reasons; one of the reasons is active resistance, meaning the group of users who actively refuse to adopt to new technology (Satchell & Dourish, 2009).

There are currently a lot of mobile applications available, and the most successful ones are specified towards younger and healthier populations (Boulos et al., 2011). The acceptance of mobile applications by elderly people, and people with physical or cognitive limitations will come with time. Boulos et al. (2011) mention that with good usability, more awareness of the capabilities of the smartphone, and with more devices tailored for the specific need of a user group, the use of mobile healthcare applications will increase by people with physical and cognitive limitations.

User testing of software applications is becoming more popular today. However, there is a variety of challenges due to unique features of mobile devices. Limited bandwidth, environmental factors and unreliability of wireless networks all affect user testing of a mobile application in the field (Zhang and Adipat, 2005). The traditional guidelines and methods used in user testing may be outdated or not applicable to a mobile user interface. Therefore, there is a need for development of convenient methodologies that can be used to evaluate the usability of mobile applications within the field of healthcare.
1.1 Background

In software-engineering it is essential to achieve certain levels of product and process quality. Usability and user experience (UX) are two ways of measuring the quality of a product, especially in interactive products. In order to achieve good usability, developers need to let users do things in a satisfactory, efficient, and effective way (Winter, 2013).

According to experts at the Institute of Medicine (IOM), it is estimated that approximately 98,000 people died from medical errors that appeared in hospitals, using data collected in 1984 (Kohn et al., 2000). Automated information systems are efficient in reducing many types of adverse medical errors (Brennan et al., 1991). Although technology plays an increasing role in modern healthcare, it will not be successful without good usability. Liljegren (2006) mentioned five important criteria which needs to be fulfilled before purchasing medical technology: thoroughness, validity, reliability, cost effectiveness and clarity. The primary focus should be on “difficult to make errors”. The IOM also reported about the use of medical technology used by non-health professionals; they argue the following: “As more care shifts to ambulatory and home settings, the use of medical technology by non-health professionals can be expected to take on increasing importance” (Kohn et al., 2000).

According to Baig et al., (2014) mobile healthcare applications have changed the technology to be applied in a divergent manner. This technology reduces mobile applications that depend on platforms specifically looking at: patient monitoring, early diagnosis, detection, and other aspects (Baig et al., 2014). The biggest challenge for mobile healthcare applications are security and privacy of personal data. Baig et al., (2014) argues that once these issues are resolved, the technology of mobile based healthcare applications will crop up as reliable and trustworthy.

A study by Jaspers (2009) examined three user testing methods and compared them. The methods chosen for Jaspers’ study were heuristic evaluation, the cognitive walkthrough, and the think aloud method. The results show that none of the three chosen methods were uniquely more effective than the others. Instead Jaspers (2009) recommended that a combination of different usability analysis techniques should be used together in order to get the most powerful tool rather than any method used separately. Jaspers (2009) also mentions that in practice the choice of a user testing method will always depend on the following factors: system design stage, the availability and specific skills of usability testers, availability of end users, and finally financial constraints. Furthermore, Jaspers (2009) mentions that there is a need recognized from human computer-interaction researchers for a complete and integrated model specified for HCI design of healthcare technologies. Research resources need to focus on multifaceted, and integrated approaches of usability evaluation of healthcare technologies, in order to expand development for designing high quality interactive healthcare technologies (Jaspers, 2009).

Sainfort et al., (2002) indicate the importance of wireless, handheld and mobile technologies, and that they will be a vital part of healthcare’s information technologies. As mentioned in the introduction section there are several problems with mobile technologies. Some of the problems were listed previously in the introduction section, additionally another aspect to consider is the specific context e.g “mobility” when doing user testing of mobile applications. This aspect can be hard to test, however some researchers have had their users walking on treadmill to simulate walking conditions (Sainfort et al., 2002).

Schusteritsch et al., (2007) mention the difficulty for researchers in practice to observe usability studies conducted on mobile devices. Traditionally usability labs were developed to study PC-based interfaces. The available equipment are often too out-of-date to be able to study mobile devices with small screens.
and forms. Schusteritsch et al., (2007) recommends that a more specialized setup should be used, for studies on mobile devices to enable a successful observation. There has also been a debate on whether mobile applications should be user tested in a traditional lab environment or in the field (Sun and May, 2013). Sun and May (2013) investigated this debate in their study, and came to the conclusion that the number of usability problems identified were similar both in lab and field setting when testing a mobile application.

1.2 Problem definition

Research have shown that there is a need for user testing on interactive healthcare technologies because of the complexity and poor design of these systems. Moreover in the field of healthcare there can be users with cognitive and physical limitations which indicates the importance of good usability in interactive healthcare systems.

The previous research on which user testing method that is suitable for mobile applications in healthcare is limited and need further investigation. Furthermore mobile application testing has other aspects such as context and mobility that are challenging to test, when performing user testing on mobile applications. Finally another challenge is to find essential criteria when choosing a user testing method for a usability evaluation of a mobile-based healthcare application.

1.3 Scope and definitions

In order to narrow down the thesis, the focus will be on elderly people that may have cognitive or physical limitations. The purpose of the technology discussed in this thesis is to assist healthcare professionals or private end users. Healthcare is defined as an organization that provides care to individuals and could be applied to different domains (E.g hospital, elderly care, or home care). Mobile applications are defined as an application that is used in a mobile way and can be used professionally by trained staff or for private use.

1.4 Purpose and research questions

The purpose of this thesis is to investigate which criteria that is of importance when choosing a user testing method in the field of healthcare. Furthermore, the work aims to examine which user testing methods that are appropriate for mobile applications in the field of healthcare.

To narrow down the scope of the thesis, two research questions have been formulated:

- What are appropriate criteria when choosing a user testing method, in the field of healthcare?
- Which user testing method is applicable in the context of mobile applications in healthcare?
1.5 Disposition

Chapter 1 presents the introduction to the subject, background, problem definition and research questions.

Chapter 2 contains the theoretical framework of different usability methods, usability attributes and physical limitations for elderly people. This chapter also contains comparisons of usability evaluation methodologies and a usability evaluation framework.

Chapter 3 covers choice of method, and research approach to show how the thesis was conducted in order to answer the research questions and purpose.

Chapter 4 presents the participants, research findings and empirical data interpretation.

Chapter 5 analyzes the empirical data of the thesis.

Chapter 6 contains the results, conclusion, discussion as well as implications and future work.
2 Theoretical framework

The following chapter includes user testing methods, physical and cognitive limitations, and mobile devices aspects such as: testing, guidelines and usability. The chapter ends with general criteria for usability evaluation methodologies.

2.1 Usability in general

ISO 9241-11 Guidance on Usability 1998 in section 8.1 defines usability as: “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (Bevan, 2001).

Based on the standard ISO 9241, human computer handbooks and existing usability studies on mobile applications, Zhang & Adipat (2005) argues that there are nine generic usability attributes.

Table. 1 Usability attributes (Zhang & Adipat, 2005)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learnability</td>
<td>Is focused on how easy a user can finish a task the first time using the application and how fast a user can improve performance levels.</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Is defined as how fast users can accomplish a task while using an application, the difference between learnability and efficiency is that before it is possible to measure efficiency the user should already have experience with the application.</td>
</tr>
<tr>
<td>Memorability</td>
<td>Make reference to the level of ease with which users can recall how to use the application after some time of non use.</td>
</tr>
<tr>
<td>Errors</td>
<td>Can be measured by counting number of mistakes a user make while using the application, the level of severity for mistakes and how readily the user can correct the mistakes.</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Deals with the user attitudes towards an application.</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>Is defined as completeness and accuracy with which users achieve defined goals.</td>
</tr>
<tr>
<td>Simplicity</td>
<td>Is the degree of comfort with which users can find ways to accomplish tasks, the attribute is often used to determine the quality of menu structures as well as navigation design.</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>Measures how easily users can understand content presented on the device.</td>
</tr>
<tr>
<td>Learning performance</td>
<td>Deals with measuring the learning effectiveness of users in education when using applications to facilitate learning.</td>
</tr>
</tbody>
</table>
2.1.1 User testing in general

User testing generally has the following aspects according to Barnum, (2002):

- The primary goal is to improve the usability of a product. For each test, there must be specific goals and concerns that you articulate when planning the test.
- The participants represent real users.
- The participants do real tasks.
- The team observes and records what participants do and say.
- The team analyzes the data, diagnoses the problems, and recommends changes to fix these problems. (Barnum, 2002, p.9)

User testing involves users that are given specific tasks in a test environment. There are several methods available which will be mentioned chapter 2.2.

The urgency of the usability problems identified by doing a user test is very important (Kaikkonen et al., 2005). Dumas and Redish (1993) use a scale from 1-4 when ranking the problems. The first level is the most urgent problems, and the fourth is the least severe. Furthermore, Kallio et al. (2004) split the urgency of problems into different categories; high (very important, task can’t be executed), medium (not so urgent, task can be executed) and low (minor problems that can be fixed later).

2.1.2 User testing in mobile devices

There are two dominant methodologies that are practiced for user testing of mobile applications, laboratory experiments (will be described below) and field studies that will be described later in this chapter (Zhang & Adipat, 2005).

Laboratory testing environment

Traditionally usability tests are conducted in a user testing room. This room usually consists of an office-like area connected to a monitoring area with a one-way mirror. The user testing room is often a quiet peaceful area, where a user can concentrate on the given tasks (Kaikkonen et al., 2005). Lab testing is conducted in a controlled environment and the users are given tasks to be carried out within specified timeframes (Barnum, 2002).

There are some concerns by the usability experts and researchers regarding that laboratory evaluations do not take into account the context where mobile phones are used. There are several factors which could influence a user's performance, such as: interruptions, movements, noise and multitasking. These factors are not present in laboratory tests (Kaikkonen et al., 2005).

The lab shown in the figure 1 below illustrates the typical laboratory environment. There are usually two rooms, that are divided by a one-way mirror between them. One room is for the test participant and an evaluator could be present in the room, to perform the specified tasks. The role of the evaluator can be to read instructions or take notes. In the other room, there is a team observing or a test administrator. Their roles are to observe, record, and sometimes code while recording (Barnum, 2002).
Laboratory studies

There are advantages performing user testing of a mobile application in a controlled laboratory setting. The tester has full control over the environment and the tester can define precise tasks which match the goal of the usability study. The tester can ensure for participants to follow the experimental instructions given (Zhang & Adipat, 2005). It is easy for testers in a laboratory environment to measure usability attributes and interpret the results by controlling for other irrelevant variables which is another advantage (Zhang & Adipat, 2005).

According to Zhang & Adipat, (2005) the laboratory experiment approach is very valuable in usability studies when focusing on comparing multiple interface designs or data input mechanisms, further a laboratory setting enables the tester to use video recording of the screen and subjects’ reaction. Limitations for the laboratory experiment include the mobile context, where factors such as unreliable internet connection and environmental factors are ignored (Zhang & Adipat, 2005).

Field studies

Advantages to conducting usability studies in the field includes the possibility to take the context of mobile use into account. The context may include slow wireless connection as well as environmental factors. Furthermore field studies provide a reliable and realistic experience, compared to laboratory experiments, where you have a staged controlled environment (Zhang & Adipat, 2005).

The disadvantages of performing usability field studies includes the lack of control over the participants. Furthermore, there are three major difficulties found in the context of mobile field studies. Firstly, it can be hard to form realistic scenarios which capture all aspects of the mobile context. Secondly, it is not simple to apply well known and established evaluation techniques such as think aloud and observation when a test is conducted in the field. Thirdly, because the users move around dynamically in the field it is challenging for both data collection and condition control (Zhang & Adipat, 2005).
2.2 User testing methodologies

2.2.1 Heuristic evaluation
According to Barnum (2002) heuristic evaluation is one of two “discount” usability methods. The other one is user testing with a few users. It is called “discount” because it is cheap and effective. Barnum (2002) explained how one evaluator covers 35% of usability problems, therefore a team of three to five evaluators that works together covers most usability problems. Each evaluator work alone to investigate a product against a set of rules or principles. Also, each evaluator goes through the product at least twice, the first time to get acquainted with the product, the second time to inspect the different elements against a set of principles (Barnum, 2002). Since the evaluators do not need to perform real tasks, heuristic evaluation can be used for prototypes that are on paper. Therefore, the heuristic evaluation method could be applicable early in the product development.

All evaluators’ results are collected and gathered in a report. This report produces a list of usability problems, and also explanation of how they violate the principles. Furthermore, some reports have a list of recommendations, although it is not required.

2.2.2 Focus groups
The similarity between regular interviews and focus groups is very high according to Cairns and Cox (2008). A focus group should consist of six to eight participants, and should be comprised of at least three people. If you have a too large group, people are likely to break off to talk in sub-groups and some people are left out from the discussion. And if the group size is too small, it can be hard to maintain a conversation. The decision to conduct focus groups is often down to logistic reasons. It can be hard to get seven engineers from different companies to attend one meeting in the same day.

An advantage of focus groups is that it allows for easier reflection on collaborative experiences (Cairns and Cox, 2008).

The role of the usability investigator is to moderate the discussion with a number of topics and to make sure to record everything. It also includes dividing the speech to all members of the group, in order to let everyone share their views, and not let a single person dominate the meeting.

2.2.3 Cognitive walkthrough
The method cognitive walkthrough was designed in order to evaluate “ease of learning, particularly by exploration” (Wharton et al. 1994). The task is to evaluate the interface in the context of tasks the users should perform (Barnum, 2002). The evaluators must first agree on the user profile and tasks to be analyzed, in order to begin the evaluation. Additionally, the evaluator must receive information about the order of steps that the user must take to complete the tasks (Barnum, 2002). This process is called “crafting the credible story”.

The last step is to do the walkthrough itself by considering which actions that a user would take following a way to solve a problem. This process often uses a question-and-answer approach such as:

- Will users know which key to select here?
- Will users understand what the system tells them at this point?
- Will users be able to determine which pulldown menu contains the action they are looking for? (Barnum, 2002)

This will simulate how the users are expected to complete tasks. The usability team uses this information to tell how the intended user is most likely to solve “problems” of completing tasks (Barnum, 2002). The primary goal of cognitive walkthrough is to evaluate the ease of learning, however other usability issues such as ease of use and functionality of the application gets uncovered with this method. The best advantage of this method is the possibility to identify problems with design early in the development, before testing the actual product on real users (Barnum, 2002). According to Barnum (2002) a disadvantage of this method is the evaluators' need for proper training in cognitive psychology or the process of cognitive walkthrough.

2.2.4 Think aloud

The think aloud method requires users to talk aloud while solving a problem or performing a task (Jaspers et al., 2004). The method is a direct way to gain insight in how humans solve problems. The method should be used to get to know about cognitive processes and how to build systems on the basis of these insights (Jaspers et al., 2004). The think aloud method consist of two steps:

- Collect think aloud protocols (what the participant thinks) in a systematic way.
- Analyze the protocols in order to gain a model of cognitive processes that take place in tackling a problem (Jaspers et al., 2004)

The protocols are gathered by instructing users to solve a problem while “thinking aloud”, the participants tell the researcher what they are thinking while performing tasks. These protocols are later used as raw data, and require further analysis and interpretation in order to gain deeper insight in the way users perform tasks.

The advantages with using the think aloud method is that it is used early in the development process which gives the usability team user task behavior analysis before any prototype is built. This may lead to less system design iterations, and result in more efficient and sufficient computer systems (Jaspers et al., 2004).

2.2.5 Participatory Design

According to Luck (2003) the definition of Participatory Design is “Design processes that involve user participation concern issues of representation in the early stages of design, when users’ needs and expectations are being expressed.”

Participatory design is research, however it has been seen as a design approach involving users (Spinuzzi, 2005). The implementation of this method vary in attention to rigor and validity. But the method follows a set of investigative principles. This method involves both designing and doing research. Just like ethnographic research, participatory design makes sure that users’ interpretations are an important factor in the research (Spinuzzi, 2005).

Participatory design is a method, with a wide range of practical techniques to enable users in the design phase (Crabtree, 1998). A few techniques will be mentioned below:

**Future workshops:** A session with both the users and designers in order to identify problems of work and alternatives from a user perspective.
**Studies of work:** Most likely pre-studies of the workplace, that needs further investigation, in order to find important aspects that requires support.

**Mock-ups:** Cardboard designs with the purpose of envisioning the future work, and enabling users to experience and adjust design solutions.

**Prototyping:** The development of the future, through preliminary and iterative design of systems enabling experience and alteration by prospective users (Crabtree, 1998).

Major disadvantages with the participatory design method is that it is time-consuming, takes a lot of resources, and requires institutional commitment to pull off (Spinuzzi, 2005).

According to Participateindesign (n.d) a participatory design is important because it gives insight in users’ needs and interests. Participatory design also show researchers the reality rather than making assumptions. Furthermore, it gives a community a chance to form a solution, rather than just giving people a solution, they are involved in the solution (Participateindesign, n.d).

**2.2.6 Individual interviews**

Individual interviews are used in order to understand users’ attitudes, beliefs, desires and experience better. Interviews can take place in person, via telephone, video conference or with instant messaging systems. In an interview the interviewer talks with participants individually for 30 to 60 minutes (HHS, n.d). Interviews ideally should be conducted in the start of the development cycle while reviewing the goals of the software.

Interviews can be used in order to supplement online surveys. Performing interviews before a questionnaire is distributed, aids in refining the questions for the survey. Conducting interviews after a survey, helps the researcher probe for reasons and details behind the answers. (HHS, n.d)
2.3 Physical limitations among elderly people

Table 2 below summarizes some of the physical limitations for elderly people. The different categories of physical limitations are discussed in more detail in the subsections below.

<table>
<thead>
<tr>
<th>Physical limitations for elderly</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision and aging</td>
<td>The individual's ability to see small details and their contrast sensitivity is declining after the age of 40 years old.</td>
</tr>
<tr>
<td>Speech and hearing</td>
<td>The ability to detect high pitch sounds is diminishing, especially in attention getting sounds with peaks over 2500 Hz.</td>
</tr>
<tr>
<td>Psychomotor abilities and older people</td>
<td>With age comes lengthening of response time on complex motor tasks.</td>
</tr>
<tr>
<td>Attention automated response and aging</td>
<td>Older adults have difficulties in maintaining attentions over a long period of time.</td>
</tr>
<tr>
<td>Memory and learning</td>
<td>Memory retrieval times increase with age.</td>
</tr>
</tbody>
</table>

Vision and aging

Problems with vision tend to appear in the early forties. At that age people start noticing difficulties adjusting focus for near vision. The individuals’ ability to see small details and also their contrast sensitivity is declining after the age of 40 years old (Fozard, 1990). Around the age of 60, individuals may encounter a reduction in the width of their visual field meaning that peripheral stimuli must be stronger and/or closer to the center of the visual field to be detected (Cerella, 1985). The ability to recognize figures that are embedded within other figures is reduced and also there is a decline in the ability to recognize fragmented objects (Frazier & Hoyer, 1992).

According to Hawthorn (2000), there is need for design that helps older users find items, and keep their attentions focused on the item. Problems with partial and embedded figure recognition may disadvantage older individuals, when looking at multiple overlapping windows on a desktop or laptop screen. Hawthorn (2002) mentions design recommendations that support visual search. Important aspects such as layout simplicity, clarity, consistency and bolder search cues should be considered.
Speech and hearing in older people

Hearing declines with age, and about 20% of people between the age of 45 and 54 have some kind of hearing impairment. This rises to 75% for people between the age of 75 to 79 (Fozard, 1990). The ability to detect high pitch sounds is diminishing especially in attention getting sounds with peaks over 2500 Hz (Hawthorn, 2000).

Applications that use sound as a means to capture the users’ attention will need to use lower frequency sounds for older individuals. The same is true for recorded voices where one should make use of speakers with low pitched voices (Hawthorn, 2000). Brief computer generated spoken messages can be useful as an option, to give hints and explanations of buttons. Furthermore, sound can be used instead of the traditional fly-over hints used in many applications (Hawthorn, 2000).

Psychomotor abilities and older people

With age comes lengthening of response time on complex motor tasks. This can be improved with training in skills such as finger tapping speed. Moreover, older adults show poorer performance when asked to track a target. A possible reason for this can be the reduced ability to hinder interference from neural noise (Hawthorn, 2000).

Furthermore, older adults are reported to having less ability to control and inflect the forces they apply. Dixon et al. (1993) found that handwriting quality gets worse with age (Hawthorn, 2000).

Older people can be expected to be slower in the use of a mouse and require larger targets to be comfortable (Hawthorn, 2000). Text input should allow for older users to shift to larger fonts, this could accommodate difficulties in point and click manipulation of small texts (Hawthorn, 2000).

Attention automated response and aging

Attention is defined as: The ability to focus on the items attention needed to perform a task (Hawthorn, 2000). Older adults have problems maintaining attention over long periods of time. Tasks that require rapid or continuous scanning are fatiguing for older adults (Vercruyssen, 1997).

There is a need for designers to use graphics carefully, and in a way that is relevant for the task rather than decoration for elderly users. Multi-media heavy web pages may disadvantage elderly users (Hawthorn, 2000).

Memory and learning for older people

Short term memory is used to hold the event in the immediate past. Information stored in short term memory gets replaced with new information when it is not transferred to long term memory (Hawthorn, 2000).

In a study conducted by Anders et al, (1972) they found age-related differences between three different age groups: young, middle aged and old individuals. They found that retrieval time increased with age. This was a result of increased time to search through the contents of short term memory and to initiate the search or to generate a response. Working memory include the act of simultaneously holding and using short term information and is central to effective and conscious action (Hawthorn, 2000).

Long term memory deals with storing information for longer than 60 seconds and is divided into several components: Episodic memory, procedural memory and semantic memory (Hawthorn, 2000).
Episodic memory is used for a specific event, procedural memory for holding information of how a task is carried out and semantic memory which holds information on the meaning of the world and the general knowledge an individual has accumulated over time (Hawthorn, 2000). Findings suggest that it is common for elderly people to have age-related deficits in episodic and procedural memory; semantic memory could also be affected at extreme age (Hawthorn, 2000).

Where it is possible, designers of an interface should provide opportunities for users to off-load memory requirements on to the system. Since time delays in the flow of the task leads to short term memory losses it is important to avoid delays and emphasize simplicity in the interface (Hawthorn, 2000).

2.4 Users of new healthcare technology

Users are using new technology differently; there are two kinds of users: *early adopters and late adopters*. Table 3 below will further explain the term late adopters which can be connected to users in the field of healthcare.

**Early adopters**

Early adopters are described as individuals who are risk-tolerant, well-educated, socially connected, science-minded and generally affluent. In the context of healthcare technologies, early adopters are those people who are more likely to engage with innovative self-tracking or medical devices (Cheung et al. 2016).

**Late adopters**

Late adoption is defined as a resistance to technology, late adoption of technology could depend on factors such as lack of user resources described as disenfranchisement or by active choice by the user. There are several categories of late adopters, some of them are described below in table 3.

*Table 3. Types of late adopters (Satchell & Dourish, 2009).*

<table>
<thead>
<tr>
<th>Lagging adoption</th>
<th>Defined with respect to some expected patterns of technology adoption and diffusion. The view of non-use does not describe the people who does not use technology, but rather people who does not use the technology yet.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active resistance</td>
<td>Describes the group of users who is actively refusing to adopt a technology. These users are later forced to use the technology anyway. Concerns over privacy, control over personal information, a political stance concerning corporate or state responsibilities, educational, environmental and health considerations are motivations for actively resisting technologies.</td>
</tr>
<tr>
<td>Disenchantment</td>
<td>Refers to an active refusal and is associated with nostalgic wistfulness for a world passing out of existence.</td>
</tr>
<tr>
<td>Disenfranchisement</td>
<td>Illustrates the way particular social groups are incapable of using particular kinds of technology. Socioeconomic disenfranchisement often arises especially as a consequence of the technology being developed.</td>
</tr>
<tr>
<td>Displacement</td>
<td>A type of indirect use of technology. The user is not directly using the technology but rather relies on people with access to the technology to forward for example messages from the telephone to the user.</td>
</tr>
<tr>
<td>Disinterest</td>
<td>Could be explained as people that are not interested in the technology. This occurs when the topics that are investigated are not of significant relevance to the broader population.</td>
</tr>
</tbody>
</table>

2.5 General criteria for user testing methodologies

Since there is a gap in the research of criteria for user testing methodologies in healthcare, this chapter presents general criteria for user testing methodologies, which can be applied to all fields. As previously mentioned, Jaspers (2009) discovered several important factors when choosing a user testing method. The system design stage, the availability and specific skills of usability experts, the availability of end users and lastly, the time and financial restraints.
2.5.1 Framework for choosing usability evaluation method based on effort and cost

Lind (n.d) has produced a framework for choosing usability evaluation method based on effort and cost. According to Lind (n.d) the framework is about how well the methods fit the formulated questions and how much human effort that is required to perform the test. The scale is from 1 to 5, where 5 is the best. The human effort is ranked from a to e, where are ‘a’ means least human effort, and ‘e’ means most human effort.

Table 4. Framework for UEMs based on effort and cost

<table>
<thead>
<tr>
<th>Stake</th>
<th>Method</th>
<th>Goal</th>
<th>Users</th>
<th>Context</th>
<th>Efficiency</th>
<th>Effectiveness</th>
<th>Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>What should the system help with, and why?</td>
<td></td>
<td>What are the prerequisites for every defined user-group?</td>
<td>What are the usability characteristics in every case?</td>
<td>Can every user-group satisfy their use-goal with the help of the system?</td>
<td>Why can not the user group reach the goals?</td>
</tr>
<tr>
<td>a</td>
<td>Heuristics evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>Participatory heuristic evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>Formative evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>Summative evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>Specified context of use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>Contextual inquiry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>Vision seminar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>System usability survey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.5.2 Criteria for choosing user testing methods
NNgroup (2014) reports that the field of user-experience has a lot of research methods available ranging from methods such as lab-based usability studies to unmoderated online UX assessments. The key question is what to do? and when to do it? NNgroup (2014) proposes a 3D framework with the following axes, see figure 2 below:

- Attitudinal vs Behavioral
- Qualitative vs Quantitative
- Context of use

![3D Framework over User Research Methods](image)

Figure 2. Picture of the 3D framework over user research methods (NNgroup, 2014).

The Attitudinal vs. Behavioral Dimension

The difference between the two dimensions could be summarized as “what people say” versus “what people do” (NNgroup, 2014). The purpose of attitudinal research is to get information about people's stated opinion, which is why this dimension is very popular in marketing departments (NNgroup, 2014). On the opposite side, is the behavioral dimension that focuses on what people do. For example, eye-tracking can be used to see how users visually interact with the system by measuring eye movements. Furthermore, there are two different methods that utilize both dimensions (E.g usability studies and field studies). They use a mixture of attitudinal and behavioral data (NNgroup, 2014). Field and usability studies give the researcher a chance to observe how people use or not use technology. As a researcher you can ask questions based on behavior. You can even adjust the study
protocol if it meets the objectives better. The analysis of field and usability studies is not usually mathematical (NNgroup, 2014).

**Qualitative vs. Quantitative**

According to NNgroup (2014) qualitative studies generate data about behavior or attitudes based on observation directly. Quantitative studies gather data about behavior or attitudes indirectly with the help of a measurement, or an instrument such as survey or other analytics tools (NNgroup, 2014). The analysis of quantitative studies are similar to mathematical analytics, since the data collection gathers a large amount of data.

Due to the differences between these studies, qualitative studies are more applicable when answering questions about why or how to fix a problem. It is suitable to use quantitative studies in order to answer questions about how many, how often etc. (NNgroup, 2014). Figure 3 below gives an illustration over how the behavioral and attitudinal dimensions affect the type of questions that can be asked.

![Figure 3](image-url)

**Figure 3, A chart over behavioral and attitudinal dimensions, and how they affect the study questions. (NNgroup, 2014).**

**The context of use**

The last aspect is about how and whether participants in the study are using a product/service in question, it can be described as following:

- Natural or near-natural use of the product
- Scripted use of the product
- Not using the product during the study

- A hybrid of the above

If the participants in the study have natural or near-natural use of the product, the objective is to decrease interference from the study in order to get an understanding about behavior or attitudes. This gives greater validity, but less control over the new topics you learn about (NNgroup, 2014).

A scripted use of the product is done in order to get insights on a specific use aspect, such as a newly redesigned flow. Studies when a product is not used in testing, explore issues that are broader than usage and usability. It can be a study about a brand or larger cultural behavior. With regards to a hybrid method, the goal is to use a creative form of product usage to meet the objectives of the study (NNgroup, 2014).

The time dimension

Other important aspects to consider when choosing a user testing method are the following: phase of product development and the objective of the test (NNgroup, 2014).

There are three different phases in the product development stage that are described below:

**Strategize:** The start phase of the product development. The user testing methods in this phase can vary a lot.

**Execute:** This is a period where you iterate on the current design direction that you have chosen. The user testing methods in this phase are formative.

**Assess:** The product or service is available for users in this phase, and this phase can measure how good the product is. The user testing methods are summative (NNgroup, 2014).

Figure 4 below shows the goal, approach and typical method for each phase.

<table>
<thead>
<tr>
<th>Product Development Phase</th>
<th>Strategize</th>
<th>Execute</th>
<th>Assess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal: Inspire, explore and choose new directions and opportunities</td>
<td>Inform and optimize designs in order to reduce risk and improve usability</td>
<td>Measure product performance against itself or its competition</td>
<td></td>
</tr>
<tr>
<td>Approach: Qualitative and Quantitative</td>
<td>Mainly Qualitative (formative)</td>
<td>Mainly Quantitative (summative)</td>
<td></td>
</tr>
<tr>
<td>Typical methods: Field studies, diary studies, surveys, data mining, or analytics</td>
<td>Card sorting, field studies, participatory design, paper prototype, and usability studies, desirability studies, customer emails</td>
<td>Usability benchmarking, online assessments, surveys, A/B testing</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: A table over the three phases of product development (NNgroup, 2014)
2.5.3 User testing method comparison

Koutsabasis et al. (2007) mentions several essential criteria when comparing UEMs (Usability Evaluation Method). They have identified five criteria that needs to be considered when choosing an usability evaluation method: Realness (or relevance), Validity (or accuracy), Thoroughness (or completeness), Effectiveness, and Consistency.

When defining the criterion realness Koutsabasis et al., (2007) refers to whether an usability finding is a real problem or not. Realness is also used to see the severity of the problem. The problems detected often get rated, from the most urgent problems to less important usability issues that could be fixed in the future.

Validity could be described as the number of real usability problems compared to the total number of findings in order to identify real and ‘false alarms’ for each application of UEM (Koutsabasis et al., 2007).

Koutsabasis et al. (2007) explains the criterion thoroughness as the ratio of the number of (real) usability problems found, compared to the total number of usability problems that exist in the targeted system. To reach good validity the total number of real problems that has been found, has to be cross-examined with all UEMs.

Effectiveness is explained as the accuracy and completeness with which users achieve specified goals (Koutsabasis et al., 2007).

The criterion consistency can be defined as the term repeatability, which means that multiple applications of different usability methods produce similar results (Koutsabasis et al., 2007).
3 Method and implementation

The objective was to find appropriate scientific methods, that would help to fulfill the purpose and to answer the research questions with a scientific support. In this chapter, the authors describe how the research questions were answered from a scientific point of view. Furthermore, which methods that have been used, and why they were chosen. Figure 5 below shows a processing map over the structure of the thesis.

3.1 Choice of method

A literature review was conducted in order to understand and examine what has been studied previously. The literature review established a lack of criteria, when choosing user testing method in mobile applications in the field of healthcare. To answer the first research question: What are appropriate criteria when choosing a user testing method, in the field of healthcare? seven semi-structured interviews with researchers and professionals in the field of HCI were conducted. The participants have previous experience in working within the field of healthcare. The empirical data collected were then compared to the literature review to verify if they matched or not. This resulted in an opinion regarding which criteria that are important, in order to choose which user testing method that is suitable for mobile applications in healthcare.

To be able to answer the second research question: Which user testing method is applicable in the context of mobile applications in healthcare? a comprehensive investigation of common user testing methods in the field of HCI was conducted. A list of user testing methods with advantages and disadvantages was presented (see chapter 2.2). Furthermore, data from the same seven semi-structured interviews aided in answering the research question.
According to Eliasson (2013) qualitative research is suitable for a deep understanding of the context and subjects studied. Direct feedback from interested individuals is central to human computer-interaction (Lazar et al., 2010). Surveys are limited in comparison to interviews. The respondents only answers questions that are asked, open ended questions are likely to go largely unanswered (Lazar et al., 2010). This was the reasoning behind the choice of semi-structured interviews.

3.2 Research approach

According to Patel and Davidsson (2011) the most common concepts that helps the researcher relate theory with empirical data are inductive reasoning, deductive reasoning, and abductive reasoning. To summarize these concepts, you could say that deductive reasoning follows the path of proof. A researcher that uses inductive reasoning follows the path of discovery. A third way of reasoning is abductive reasoning, which is a combination of inductive and deductive reasoning. Patel and Davidsson (2011) mentions that an advantage with abductive reasoning is that the researcher will not be as limited to either path of reasoning, which can be the case when working either inductive or deductive.

However, there is a risk with abductive reasoning. The researcher could choose a research topic from previous experience, additionally writing a hypothesis that excludes other interpretations. The chosen research approach for this thesis was abductive reasoning, because the authors wanted to compare existing theory with empirical data, and to come to a conclusion. Another reason was that the authors did not have previous experience with the research topic, hence the choice of the research approach. Lastly, the choice of abductive reasoning would contribute with new knowledge in choosing a user testing method, in the field of healthcare. As mentioned previously, the literature review revealed that there is a lack of research in the chosen topic.

3.3 Literature review

After searching in several databases on what aspects that need to be considered before choosing a user testing method in the field healthcare, the authors found that there is a gap in the research. For this reason, the focus shifted to find general criteria, when choosing a user testing method in the field of healthcare instead.

According to Bryman (2012) existing literature represents an important component in all research. When a topic is of interest, the researcher must read further to determine the following things: What is already known about the topic? What concepts and theories have been applied to the topic? What research methods have been used? Are there any controversies and key contributor to the research in the topic? A literature review gives the researcher an overview of what has been done, and if there is any gap that needs further investigation. It is an important and appropriate way of demonstrating credibility and contribution of the research (Bryman, 2012).

Patel & Davidsson (2011) describe that the usual sources when seeking information are books, articles published in scientific journals and reports. These sources are often printed, but it is often possible to find a copy that could be read in digital form. When conducting a literature review the researchers often come across sources that are either in printed form or digital documents. In books the reader is more likely to find a summary about the different problem areas. To find the most recent literature within the subject of interest it is recommended to search for articles, reports and conference articles, since publishing a book often requires more time. Literature is most easily found via university
libraries as well as through other online sources and journals. It is time consuming to conduct a literature review as searches could generate a large number of hits. This means that the researchers need to delimit and sift through the literature in order to find the proper literature (Patel and Davidsson, 2011).

Being able to link your own research questions, findings, and discussion to the existing literature is an important and useful way of demonstrating the credibility and contribution of your research. The written literature review is expected to be critical. This does not necessarily mean that you are expected to be highly critical of the authors you read, but it does mean that you are supposed to assess the significance of their work and how each item fits into the narrative about the literature that you construct when writing a literature review.

3.4 Data collection

Primary data is information that is collected by the authors, and secondary data is information that is already collected by others. Eriksson and Widersheim (1991) write that the researcher should consider expense, quality, and accessibility. It is hard to find a method which meet these three criteria, hence the authors have to do a trade-off. The choice depends on the situation, and has to be adjusted after it (Eriksson and Widersheim, 1991).

The authors have collected both primary and secondary data. Primary data is collected via semi-structured interviews with researchers and professionals within the field of HCl. Secondary data were collected through an extensive literature review, where sources of information were compared against each other in order to determine consensus.

The authors have mainly used literature from the library of Uppsala University. Digital articles have been collected from the search engine provided by the library of Uppsala University, ACM Digital Library, Google Scholar, and the Google search engine. These search engines have been used to get a scientific insight, and to filter out irrelevant data, in order to get reliable quality in this thesis.

Several keywords were used in order to get reliable hits. Keywords such as “user testing methods”, “usability evaluation method”, “usability evaluation in healthcare”, “criteria usability evaluation” to name a few. These keywords have then been matched with other words to get a fewer more precise hits, and also to filter out irrelevant data.

3.4.1 Qualitative interview

Patel & Davidsson (2011) highlight the problem with summarizing the meaning of a qualitative interview. The area is complex and can be done in multiple ways. Qualitative interviews often have a low degree of structure, meaning that the questions asked by the researcher gives the interviewee the possibility to answer in his or her own words. The interview could also be conducted with a high or low grade of standardization. With a high grade of standardization, the researchers will ask the questions in a set order. Low standardization means that the researcher asks questions in the order that seems fit according to the situation. The purpose of using qualitative interviews is to discover and identify properties of a specific problem or subject, for example the thought and perceptions of a phenomena.

According Moll (2013), performing semi-structured interviews aids in the interpretation of observational data since they give insight in participants’ own views and opinions of what went on during the study. This data is valuable, because it is not possible to be gathered through observations or by analyzing quantitative data. Therefore, they add a new magnitude to the results (Moll, 2013).
As mentioned previously primary data were collected through seven semi-structured interviews with researchers and professionals in the field of human computer-interaction.

The authors asked each interviewee for the permission to audio record each session. The reason was to not miss out on important information, and to be able to cite the participants accordingly. Additionally, the authors took notes while conducting the interviews. The authors used low standardization when asking questions, in order to let the participant, speak freely based on their experience and opinions. Furthermore, to have the possibility to ask another question if the interviewee is not able to answer a specific question. Each session took about 30-40 minutes, and the location was in the participants’ office in a quiet environment. However one session was conducted through Skype.

**Interview participant overview**

The requirements for the interview participants were the following:

- Minimum five years of experience in the field of Human computer interaction (HCI) or similar.
- Experience in healthcare domain
- Researcher or working professional

The reasoning behind this was that the authors wanted a “developer” perspective that have insights in both healthcare and user testing knowledge. Both HCI researchers and professionals fulfilled these requirements. The recruitment of participants was firsts done via email, a number of participants recommended other people as potential participants. The recommendations helped the authors in the recruitment process of participants for this thesis. Each participant was given the thesis topic, no further preparations before each interview was done.

**Interview questions**

After doing the literature review, the authors had enough knowledge about which areas that needed more information. The literature review revealed that there is a gap in criteria for choosing user testing method in the domain of healthcare. This resulted in asking professionals and researchers in the field of human computer interaction about their experience of user testing in healthcare. The literature also revealed that there are difficulties when performing user testing. This is especially true in the domain of healthcare, hence one of the interview questions were about challenges present when doing user testing in healthcare. Furthermore since the literature review showed that there is no single user testing method capable of testing all aspects, the authors asked interview participants about what they thought was important in the choice of usability evaluation method in the field of healthcare. All interview questions can be found in appendices (see Interview Questions).

**3.5 Data analysis**

Solvang and Holme (1997) explain that you can separate text analysis in two different forms: complete analysis and partial analysis. A complete analysis, examines all of the collected data. Interviews get a purpose when put in a context. In a partial analysis sentences with a certain claim related to an appearance that is in the focus in the research. The analysis is then categorized or is described in tables that lead to an interpretation of the topic that is in focus.
A complete analysis was performed after the interviews. The authors’ examined their own notes and also the recorded material. According to Ryan and Bernard (2003) there are several tasks to do when analyzing text. They mention discovering themes and subthemes, deciding which themes that are important, building hierarchies of themes, and linking themes into theoretical models. The authors have chosen to focus on the first task: discover themes and subthemes, since the task was to identify the important data and to sort out irrelevant data. Ryan and Bernard (2003) argues that it is important to discover themes in qualitative data because without thematic categories, researchers would have problems describing, comparing and explaining. According to Ryan and Bernard (2003) a theme is an abstract construct that link expressions found in text, images, sounds and objects. To be able to find a theme, you can ask the question: What is this interpretation an example of? (Ryan and Bernard, 2003). Themes can be found both from collected data, and from the researcher’s theoretical understanding of a phenomenon (Ryan and Bernard, 2003).

After analyzing the data, there was an interpretation session to specify the results of the study. The results were then applied to the thesis purpose and research questions to see if they have been fulfilled. Quotes have been used to provide a deeper understanding of the problem area and to strengthen the findings.

When the literature review was finished, the authors started the analysis of that data, to filter out which data that is needed to answer the research questions. Moreover, to see which data is unnecessary and irrelevant. The next step in the analysis was to compare the general criteria from the literature review with the collected empirical data. The authors also did a comparison of the user testing methods. A list of pros and cons was conducted to find a suitable method that is applicable in the context of mobile application in healthcare.

### 3.5.1 Interview analysis

The analysis of the empirical data was done in three steps. The first step was to identify themes that were important in all of the collected data. This was done by listening through the recorded material, and reading the authors’ own notes from the interviews. The authors first listened through each interview separately to identify which themes that were discussed in all interviews. To be able to identify the most important themes, the authors had an interpretation session after each interview. In the session the authors discussed which themes that were considered important.

The next step in the analysis was to collect what have been said. In this step the authors listened to each interview together to be able to interpret what data that was essential. Each participant had different views and explained various phenomena differently, thus the authors had to interpret what they meant. It resulted in sentences and concepts that will be described below in 4.3.

The third step was to code the collected data. To be able to do this, the authors listened through each interview twice. This was done to secure that no data would be left out. The coding was done by counting each time a sentence, or concept were mentioned in each interview. For example, the concept “context” were mentioned twenty times, this was coded as x 20 in the analysis. To add a further dimension to the analysis the authors also counted in how many interviews a sentence or concept were mentioned. For example, the sentence or concept of “Too little user testing in healthcare” was mentioned in four out of seven interviews. This was coded as (4/7). Additionally, each sentence or concept is marked with different colors. The colors that were used in the coding of the data were: red, yellow and green. Red means that a sentence or concept was mentioned once or twice out of seven interviews. Yellow means that a sentence or concept was mentioned at least three to five times out of all seven interviews. Finally, green means that a sentence or concept were mentioned in six out of seven interviews.
interviews or in all seven interviews. If a sentence or concept were mentioned in all seven interviews, the authors coded it as (*7/7*). The tables below give a thorough overview of how the coding was done. The first column includes each theme with sentences and concepts that is connected to that theme. The second column is how many times a sentence or concept was mentioned. The third column summarizes how many times a sentence or concept was mentioned, out of the total number of interviews.

3.6 Reliability and credibility

Reliability is concerned with the question of whether the results of a research are repeatable (Bryman, 2012). In qualitative research, reliability can be divided into two aspects: external reliability and internal reliability. External reliability, meaning the way a study can be replicated, is a difficult criterion to meet in qualitative research since it is impossible to freeze a social setting (Bryman, 2012). The term internal reliability means whether there is more than one observer, and if the members of the research team agree on what they see and hear (Bryman, 2012). In order to achieve good internal reliability, the authors have written down answers obtained from interviews individually, the notes have then been compared, to minimize the risk of misunderstanding. Interview sessions have also been audio recorded.

To increase the reliability of the thesis, the authors have read and discussed the notes from the interviews. Furthermore, the recorded interviews have been inspected several times, in order to not miss out on any important information. The authors have also reviewed each other’s writing in this thesis, to increase the reliability. Moreover, the authors have investigated the sources from websites, to confirm the credibility of the organizations behind the websites. Furthermore, the participants that were interviewed have considerable experience, which indicated that the collected empirical data is credible.
4 Findings

The authors of this thesis conducted seven semi-structured interviews with researchers and professionals in the field of Human-Computer Interaction. In order to reach a high level of confidentiality, the authors have chosen to anonymize all participants. Each participant will be given a participant number, and is going to be referred by that number in this report from now on (See section 4.1). To be able to answer both research questions: “What are appropriate criteria when choosing a user testing method, in the field of healthcare?” and “Which user testing method is applicable in the context of mobile applications in healthcare?” the authors chose to interview both researchers and professionals working in the field of Human-Computer Interaction. The reasoning behind this was to get different perspectives from both of the participant groups. The findings will be presented in three parts: a brief description of the participants, followed by an interview analysis and finally an interpretation of the findings.

4.1 Participant information

Table 5 presents the participants interviewed in our study. The table contains information on the participants’ age, gender, number of years in the field, speciality within the field, working experience within the field of human computer interaction (not in a research capability) the date when the interview was conducted and if the interview was audio recorded.

Table 5. List of participants that participated in the study.

<table>
<thead>
<tr>
<th>Participant No.</th>
<th>Age</th>
<th>Gender</th>
<th>Number of years in the field</th>
<th>Speciality within the field</th>
<th>Working experience in years</th>
<th>Interview date</th>
<th>Audio Recorded</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54</td>
<td>M</td>
<td>32</td>
<td>Disability research and task analysis and UI design</td>
<td>3</td>
<td>17-04-26</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>M</td>
<td>9</td>
<td>Interaction design and design theory</td>
<td>1</td>
<td>17-04-26</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>M</td>
<td>37</td>
<td>Information visualisation, user testing, user-centered design</td>
<td>40</td>
<td>17-04-26</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>63</td>
<td>F</td>
<td>30</td>
<td>Human-computer Interaction (Visual Design)</td>
<td>30</td>
<td>17-05-04</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>F</td>
<td>20</td>
<td>Human-computer Interaction</td>
<td>25</td>
<td>17-05-09</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>F</td>
<td>9</td>
<td>HCI - Usability testing in healthcare</td>
<td>5</td>
<td>17-05-10</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>31</td>
<td>M</td>
<td>6</td>
<td>UX-designer</td>
<td>6</td>
<td>17-05-15</td>
<td>Yes</td>
</tr>
</tbody>
</table>
4.2 Themes of all seven interviews

This chapter presents a summary of the seven interviews, the authors have categorized important concepts into themes. Each theme will be presented with the most common concepts that were mentioned in all seven interviews. The table also shows how many times a certain concept was mentioned, and lastly in how many interviews it was mentioned.

The first column includes each theme with sentences and concepts that is connected to that theme. The second column shows how many times a sentence or concept was mentioned. The third column summarizes how many times a sentence or concept was mentioned, out of the total number of interviews. Additionally the themes are color coded in the following colors: red, yellow, and green. Red indicates that a sentence or concept were mentioned once or twice out of seven interviews. Yellow shows that a sentence or concept were mentioned at least three to five times out of all seven interviews. Finally green indicates that a sentence or concepts were mentioned in six out of seven interviews or in all interviews. If a sentence or concepts were mentioned in all interviews, the authors coded it as (*7/7*).

Table 6. Themes Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Number of mentions</th>
<th>Mentioned in x out of 7 interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase in development</td>
<td>x13</td>
<td>(6/7)</td>
</tr>
<tr>
<td>Contextual usability</td>
<td>x12</td>
<td>(6/7)</td>
</tr>
<tr>
<td>Access to end-users</td>
<td>x18</td>
<td>(4/7)</td>
</tr>
<tr>
<td>What is the user-group?</td>
<td>x9</td>
<td>(4/7)</td>
</tr>
<tr>
<td>Since the context is in healthcare, security is essential</td>
<td>x14</td>
<td>(3/7)</td>
</tr>
<tr>
<td>What do you want to test?</td>
<td>x13</td>
<td>(3/7)</td>
</tr>
<tr>
<td>Depends on how much resources are available</td>
<td>x3</td>
<td>(2/7)</td>
</tr>
<tr>
<td>There is a need for user-involvement</td>
<td>x3</td>
<td>(2/7)</td>
</tr>
<tr>
<td>Is the usability test short-term or long-term?</td>
<td>x3</td>
<td>(2/7)</td>
</tr>
<tr>
<td>The usability method needs to be lightweight</td>
<td>x2</td>
<td>(2/7)</td>
</tr>
<tr>
<td>Actability</td>
<td>x4</td>
<td>(1/7)</td>
</tr>
<tr>
<td>More informants are needed in high security systems</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
</tbody>
</table>
### Table 7. Themes mobile application testing

<table>
<thead>
<tr>
<th>Mobile application testing</th>
<th>Number of mentions</th>
<th>Mentioned in x out of 7 interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>x22</td>
<td>(<em>7/7</em>)</td>
</tr>
<tr>
<td>Test different kinds of devices</td>
<td>x4</td>
<td>(3/7)</td>
</tr>
<tr>
<td>Mobility</td>
<td>x2</td>
<td>(2/7)</td>
</tr>
<tr>
<td>It is very important to write good scenarios</td>
<td>x3</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Mobile application testing needs a lot of manual work</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
</tbody>
</table>

### Table 8. Themes Challenges

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Number of mentions</th>
<th>Mentioned in x out of 7 interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>x22</td>
<td>(<em>7/7</em>)</td>
</tr>
<tr>
<td>Hard to test, fragment of a whole system.</td>
<td>x11</td>
<td>(4/7)</td>
</tr>
<tr>
<td>End-users technical skills</td>
<td>x10</td>
<td>(4/7)</td>
</tr>
<tr>
<td>User testing comes late into development phase, hard to change core-functionality.</td>
<td>x6</td>
<td>(4/7)</td>
</tr>
<tr>
<td>Usability expert’s competence</td>
<td>x5</td>
<td>(4/7)</td>
</tr>
<tr>
<td>High risk environment</td>
<td>x2</td>
<td>(1/7)</td>
</tr>
<tr>
<td>No flexible architecture</td>
<td>x2</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Healthcare professionals are often busy, hard to recruit them for testing.</td>
<td>x2</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Physical limitations/impairments in health care leads to hard to recruit end-users.</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Be clear that it is the application that is going to be tested, not the end-user.</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
</tbody>
</table>
### Table 9. Themes Improvements

<table>
<thead>
<tr>
<th>Improvements</th>
<th>Number of mentions</th>
<th>Mentioned in x out of 7 interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>A way to test the whole system</td>
<td>x8</td>
<td>(4/7)</td>
</tr>
<tr>
<td>Early in the development phase</td>
<td>x6</td>
<td>(4/7)</td>
</tr>
<tr>
<td>More user testing in healthcare</td>
<td>x6</td>
<td>(4/7)</td>
</tr>
<tr>
<td>Do a list of heuristics specified for healthcare, and educate healthcare staff about usability.</td>
<td>x2</td>
<td>(2/7)</td>
</tr>
<tr>
<td>More resources to usability evaluation in the field of healthcare</td>
<td>x2</td>
<td>(2/7)</td>
</tr>
<tr>
<td>Application needs to be user tested and suited for elderly/impaired people.</td>
<td>x2</td>
<td>(2/7)</td>
</tr>
<tr>
<td>There is a need for visible results.</td>
<td>x2</td>
<td>(1/7)</td>
</tr>
<tr>
<td>The application should be sustainable.</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>User tested certification</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Separate database and user-interface, competition will give better usability.</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>How to engage users in improving usability.</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Look at the pre-conditions, before the user testing.</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
</tbody>
</table>

### Table 10. Themes User Evaluation methods

<table>
<thead>
<tr>
<th>User evaluation methods in healthcare</th>
<th>Number of mentions</th>
<th>Mentioned in x out of 7 interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation study</td>
<td>x8</td>
<td>(4/7)</td>
</tr>
<tr>
<td>Heuristics</td>
<td>x5</td>
<td>(4/7)</td>
</tr>
<tr>
<td>Think aloud</td>
<td>x5</td>
<td>(4/7)</td>
</tr>
<tr>
<td>Interview</td>
<td>x5</td>
<td>(4/7)</td>
</tr>
<tr>
<td>Informal User testing</td>
<td>x2</td>
<td>(2/7)</td>
</tr>
<tr>
<td>Focus group</td>
<td>x3</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Cognitive walkthrough</td>
<td>x2</td>
<td>(1/7)</td>
</tr>
<tr>
<td>User centered design</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Survey</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Field studies</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Participatory design</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>PACT framework</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Theory of use</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Action research</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Diary studies</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Lo-Fi prototyping</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
</tbody>
</table>
Table 11. Themes physical limitations

<table>
<thead>
<tr>
<th>Physical limitations</th>
<th>Number of mentions</th>
<th>Mentioned in x out of 7 interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size/shapes/attributes of buttons and text for elderly</td>
<td>x4</td>
<td>(4/7)</td>
</tr>
<tr>
<td>Motor controls</td>
<td>x6</td>
<td>(3/7)</td>
</tr>
<tr>
<td>Not many features, keep it simple for elderly.</td>
<td>x2</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Physical capability</td>
<td>x2</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Dexterity</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Vision</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Hearing</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
<tr>
<td>Big icons</td>
<td>x1</td>
<td>(1/7)</td>
</tr>
</tbody>
</table>

Table 12. Themes Contextual usability

<table>
<thead>
<tr>
<th>Contextual usability</th>
<th>Number of mentions</th>
<th>Mentioned in x out of 7 interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context awareness</td>
<td>x22</td>
<td>(<em>7/7</em>)</td>
</tr>
<tr>
<td>Healthcare staff have behaviors and habits without thinking about it.</td>
<td>x4</td>
<td>(2/7)</td>
</tr>
<tr>
<td>Hard to test the system because healthcare staff have different professions and end goals.</td>
<td>x2</td>
<td>(2/7)</td>
</tr>
<tr>
<td>Difference in technical skills between developers and end-users gives different subjective opinion about the system.</td>
<td>x2</td>
<td>(2/7)</td>
</tr>
</tbody>
</table>

4.3 Empirical data interpretation

The findings will be presented by going through each theme, with a summary of the important findings. The authors will discuss the most common and mentioned concepts in each theme.

Criteria (see Table 6)

Every participant agreed that the concept of “context” is crucial to consider when choosing a user testing method in the field of healthcare. Context was identified in several themes such as: criteria, mobile application testing, challenges and contextual usability. The concept of context was mentioned by all participants in different themes and situations. The importance of context is high since it was mentioned twenty-two times in all seven interviews. The following quote from participant 1 highlights the importance of context: “There is a lot of aspects that would not be caught unless you actually test the context. I would say contextual usability or something like that. And we need that especially much in those kind of environments that the healthcare system works in.”

Another aspect that was considered important according to six out of seven participants is “phase in development”. This can be explained as what time in the development that a user testing session is planned. It can be early in the development, or it can be after core-features are implemented, or it can be
to assess a complete product after the development. This aspect decides what user testing methods that are available and suitable. This criterion can be explained by participant 6 with the following quote: “When do you have access to them [participants] is it in the development phase when they can influence stuff? Or is it later on? ”

Two other critical phenomena that were discussed by four out of seven participants were: What is the user-group? and Access to the user-group? The first criterion can be explained by the following quote from participant 2 “The first thing I would try to find out more about who are the intended users. That is the most important thing. Also I would try to understand what kind of motivations do they have for using this one. “

The second criterion “Access to the user-group” clarifies if the testers have access to the end-users that are going to use the system or if they have to consider other user-groups for the test. An example of this can be user testing of a journal system in healthcare. The users’ using the system are nurses, but the only users that are available for user testing are doctors. This can be the case sometimes, hence the importance of “access to the user-group”. This criterion was considered essential, especially in the field of healthcare because of the complexity of many different systems working simultaneously.

Furthermore, another aspect to consider when choosing a user testing method in the field of healthcare is: What do you want to test? Three out of seven participants explained that user testing is broad, and you can test different parts of a system or application. For example you can test learnability, effectiveness, satisfaction, or if the system prevent errors. Participant 7 explained the following regarding this principle: “The aspect you want to test is central in choosing an appropriate method, you need to think hard about your choice of method depending on what you want to find out.”

Participant 5 introduced a new term called “Actability”. Participant 5 claimed that it was vital to consider this term before considering user testing method in healthcare. The following quote explains the term further: “I think that the method actability is very very important, that the method gives results that the organization is ready to cope with. That the usability evaluation results in things that is possible to do something about.”

Finally, three out of seven participants argued that since the context is healthcare, security is imperative. Since medical records are sensitive data, aspects such as identification and preventing errors are important. For example user identification and preventing errors like administering an incorrect dosage of medicine. Moreover it is important to do thorough usability tests in the domain of healthcare because system-errors can lead to life-and-death situations. Participant 1 described the following about security: “When you have this kind of system where you have a lot of requirements on the safety and security you have to actually spend more money. It has to cost money to actually do this user testing here. Because an error will make big, if you have an error you feed in 1000mg of something instead of 1mg that is a very big mistake and it could actually cost lives. “

Mobile application testing (see Table 7)

As mentioned in the previous section the concept of “context” was considered vital when doing mobile application testing. All participants agreed on the fact that context needs to be considered before doing user testing on mobile applications. The reason for this was that mobile devices are used differently than normal desktop computers. Participant 5 mentions the difference between traditional and mobile application testing with the following quote: “It is a huge difference between if you take a stationary computer and mobile phone, you need to think about completely different methods.”
Furthermore, it is vital to test different kinds of devices when doing mobile application testing. Participant 6 explains: “I have also thought about the importance of usertesting on different devices, because that was something we noticed now, to test on Android and Apple and different sized phones. Thus testing tablets and phones, they have different sizes. I think it matters for others and for elderly it’s a big deal because, yes… how big the text size is simply speaking. “

Finally, the term “mobility” was mentioned by two out of seven participants. An explanation to the term can be that mobile devices are used in different environments, contexts. Furthermore, mobile devices are often used while doing something else (e.g taking a walk), thus you should test mobile devices in that specific situation and context to have ecological validity.

**Challenges (see Table 8)**

The findings show that there are challenges present when doing user testing in the field of healthcare. Three challenges that were mentioned by at least four out of seven participants were: It’s hard to test healthcare systems because they are often a fragment of a whole system, end-users’ technical skills, and user testing comes in too late into the development phase, it’s hard to change core-functionality.

The first challenge “It’s hard to test healthcare systems, because they are often a fragment of a whole system” can be explained by the following quote from participant 2: “User testing is ideal but I think it’s for systems that are reasonably complex it does not solve the problem and it should rather be done by the companies that are selling the systems“.

The second challenge “end-users’ technical skills” is referring to the different technical competences of all healthcare staff. It can be staff that works extra shifts, nurses, doctors or elderly staff that are less tech-savvy. This results in doing many tests with all users-groups, which can be tough since staff are often occupied with their own work, hence they are unavailable for user testing.

The third challenge concerns the stage of development phase, and specifically that user testing is prioritized too late, thus problems with changing core-functionality. Participant 2 explains the following about this subject: “When you have chosen a system then you can start, when you’re sort of rolling it out into practice then they are doing this pilots like they roll it out with one pilot first then, before they do the whole, but the feedback they get from the pilot cannot affect the core functionality anymore“.

**Improvements (see Table 9)**

In this theme participants gave their opinions about improvements when doing user testing in the field of healthcare. An improvement they suggested were to do more user testing in the field of healthcare. Participant 5 mentioned the following: “We did an interview study with companies in Sweden who are working with agile methods me and [x] we should… we asked how do you evaluate with users, that was the research question. This was a couple of years ago now, and we had a crisis afterwards because the answer was we do not.”

As mentioned in the “Challenges” section healthcare systems are often part of a bigger system, and it is hard to test them all. They advise a way to test the whole system integrated in healthcare to catch all aspects that needs evaluation.

The last improvement is also mentioned in the previous section and concerns the user testing in the development phase. Four out of seven participants agreed on that user testing must come early in the development phase in the field of healthcare.
User evaluation methods in healthcare (see Table 10)

All participants agreed on that choosing a specific method to do user testing is not enough. Other variables such as context, access to users, what needs testing, and resources needs to be considered. However, some user testing methods that were mentioned, and could be applicable in the field of healthcare were: Observational studies, Interviews, Think Aloud, and Heuristic evaluation.

Physical limitations (see Table 11)

Four out of seven participants agreed on that there has to be a consideration of size, buttons, shapes and text that is suitable for elderly people. Participant 4 expressed the following regarding this topic: “Healthcare, then the users are not one, represent not the healthy active part of the population maybe, but they are people who are a little older, are a bit sicker, and in general it's usually harder to get them to participate.”

Furthermore, another thing to look at is “motor controls”. This aspect is vital to consider when doing user testing with elderly people that have limited mobility.

Contextual usability (see Table 12)

As mentioned previously (see Criteria and Mobile application testing) context awareness was also considered essential in this theme. The term is explained earlier in this chapter (see Criteria). Two out of seven participants mentioned that healthcare staff often have behaviors and habits without thinking about it. They further explained that healthcare staff get “knowledge” in the workplace, and after some years they just do things without thinking about them. For this reason, some habits and behavior can be missed out if you ask healthcare staff directly. Participant 7 describes the following: “You can’t just ask the healthcare staff about what they would like to have in a system, they don’t know because they tend to forget about their daily routines”.

Finally, another point to recognize is the fact that “healthcare staff have different professions and goals”. For example, nurses need to access some systems while doctors and reception staff are working with other systems. This adds another dimension and further complicates user testing in the field of healthcare.
5 Analysis

In this chapter the findings of the thesis will be compared against the theoretical framework to see if the data corresponds or differs. The analysis will be presented with a summary of the key findings from chapter 4.3, followed by the theoretical findings.

5.1 Research question 1: What are appropriate criteria when choosing a user testing method, in the field of healthcare?

The empirical data from this study shows that the participants agree on that two important criteria are context and phase in development. Context can be described as in what kind of environment and situation something is being tested. Phase in development can be explained as in what stage of the development user testing is planned.

According to NNgroup (2014) and Barnum (2002) the context of use is essential to consider when choosing a user testing method. Barnum (2002) explains that one gets the actual environment the user works or lives in which provides insight into their workspace, lightning, type of hardware and other aspects of use. Furthermore, you get insight in how a user uses the product in a real-world setting, instead of a lab which is a controlled environment. Moreover, Jaspers (2009) also mentions the system design stage as a factor that plays part when choosing a usability testing method.

With these facts given, the general criteria match the criteria produced in this thesis. It is relevant to state that two important criteria that needs consideration before choosing a user testing method in the field of healthcare are: context and phase in development. It is safe to claim that these criteria are relevant for most fields, however the importance of these criteria may vary.

Other criteria that conform are: What is the user-group? and Access to the user-group? The first question asks about which is the intended user-group and the second question asks if the tests would have access to the end-users who are going to use the system. Jaspers (2009) also mentions this in his study. The conclusion of his study was that no single user testing method has shown any significant results indicating it is singularly effective in all circumstances. Therefore, Jaspers (2009) propose a combination of different usability techniques. However, Jaspers (2009) also mentions that in practice factors such as availability and representativeness of end-users plays a big role in the choice of user testing method. Jaspers (2009) also recognize other factors such as: usability-experts’ skills, time and financial constraints. After comparing criterias found in literature with our collected empirical data, we found that the two previously mentioned criteria, context and phase in development can be worth inspecting before the consideration of a user testing method in the field of healthcare. As mentioned previously these criteria can also be applied to most fields, but these are especially important in the field of healthcare.

Furthermore, another aspect that were commented in the empirical data findings was the term actability. Actability refers to if the chosen method gives results and actions that the organization is ready to cope with. The other aspect that was considered critical was security. This term was explained as: Healthcare systems must be able to prevent fatal errors that could lead to life-and-death situations.

These criteria are not mentioned and recommended in the general criteria by NNgroup (2014) and Jaspers (2009) and Barnum (2002). The term actability was a new term introduced to the authors. The term is considered vital, because even if user testing results in a list of usability problems, the organization has to be ready to fix them. There is no point in identifying problems, without a plan to fix
them. The criterion security is mentioned by three participants in this study. However, it is not mentioned in the general criteria. An explanation for this could be that security is considered essential in the field of healthcare, but the importance of it is limited in the general criteria for usability evaluation methods.

Finally, the last criterion that was agreed on both in the empirical findings and the theoretical framework was: What do you want to test? This criterion was described as: since user testing is broad, you can test different parts of a system. According to the empirical data findings there has to be a definition of what is intended to test, before the choice of a user testing method in the field of healthcare.

Barnum (2002) recommends that a user test should include: specific goals and concerns. Furthermore, Barnum mentions that the test should have real users, real tasks and that usability evaluators should observe and record what users do and say. Since both Barnum (2002) and the empirical data findings correlate, it is relevant to say that the criterion; what is intended to test, is of importance, when choosing a user testing method in the field of healthcare.

5.2 Research question 2: Which user testing method is applicable in the context of mobile application in healthcare?

According to the empirical data findings there was not a single user testing method that could be specifically applied in the context of mobile application in healthcare. However, there were several user testing methods mentioned by several of the participants: Observational studies, interviews, think aloud and heuristic evaluation.

As stated previously the empirical findings show that no single user testing method is best suited for mobile applications in the field of healthcare. According to the participants’, aspects such as: context, access to users, what needs to be tested, and resources available needs to be contemplated before choosing a user testing method.

Jaspers (2009) states that there is no single user testing method that showed any significant results indicating that is it singularly effective in all circumstances. Jaspers (2009) recommends a combination of different user testing methods that complement each other. However, Jaspers (2009) argues that the choice of user testing method is different in practice where it depends on several factors such as: system design stage (phase in development), the availability and skills of usability experts, the availability and representativeness of end-users and finally financial constraints.

Lind (2017) also mentions several questions that needs an answer, to determine which user testing method is appropriate. Lind (2017) proposes a framework with several criteria including: goal, users, context, efficiency, effectiveness, and satisfaction. After comparing the empirical data findings with the theoretical framework, it is relevant to state that there is no single user testing method that is singularly effective in all circumstances. The choice of method that is applicable in mobile applications in the field of healthcare must first consider the previously stated factors, and first answer questions such as the following:
- What is the goal?
- Who are the intended users?
- What is the context?
With that said, if the previous factors are known, there are a few user testing methods that are appropriate to use when doing user testing on mobile application in the field of healthcare: Observational studies, interviews, focus groups, and heuristic evaluation.
6 Discussion and conclusions

After analyzing the empirical data with the theoretical framework, the results will be further discussed in this chapter.

6.1 Results

6.1.1 Ranking of criteria

In this section a ranking of the identified criteria will be presented. The authors had the results of literature review and semi-structured interviews in mind, to be able to rank the criteria. Furthermore this ranking should be seen as subjective based on the findings of the thesis. The table below consists of criteria that are of importance, in the choice of user testing method, in the field of healthcare. The authors have ranked the criteria on the scale 1-3. A rating of 1 means that the criteria is mandatory, and is fundamental in the choice of user testing method in the field of healthcare. A rating of 2 indicates that the criterion is essential, and needs to be considered before choosing a user testing method. Finally, a rating of 3 imply that the criterion is slightly less important, and should be optional. All of the criteria below in Table 13 are of relevance and importance. However, the most important criteria will be discussed in chapter 6.4. The other criteria will not be discussed further than the analysis (see chapter 5.1).

What are appropriate criteria when choosing a user testing method, in the field of healthcare?

Table 13. Criteria when choosing user testing method in the field of healthcare

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Explanation</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>In what kind of environment and situation is something being tested?</td>
<td>1</td>
</tr>
<tr>
<td>What do you want to test?</td>
<td>What is the goal of the test? What part of the system do you want to test?</td>
<td>1</td>
</tr>
<tr>
<td>Security</td>
<td>The system should be able to identify a user and prevent errors.</td>
<td>1</td>
</tr>
<tr>
<td>Phase in development</td>
<td>In what stage of the software development user testing is planned.</td>
<td>2</td>
</tr>
<tr>
<td>Access to the user group</td>
<td>Access to the user group who are intended to use the system?</td>
<td>2</td>
</tr>
<tr>
<td>Actability</td>
<td>Refers to if the chosen method gives results and actions that the organization is ready to cope with.</td>
<td>2</td>
</tr>
<tr>
<td>What is the user group</td>
<td>Who are the intended user group?</td>
<td>3</td>
</tr>
</tbody>
</table>
6.2 Implications

The authors’ findings suggest that there is a need for better understanding of the important aspects, when choosing a user testing method within the field of healthcare. Furthermore, the analysis showed that it is not enough with a single user testing method. There are several factors that need consideration before choosing a user testing method for mobile applications in the field of healthcare.

The thesis is aimed towards professionals and researchers within the field of human computer interaction. It is also addressed to people interested in the topic, for example healthcare organizations, and user testing organizations. The results of this thesis could also be applicable to other domains such as: mobile-application testing in general, mobile application in the fitness industry etc.

Moreover, this thesis could also be used to get knowledge about different user testing methods, user testing in mobile devices, and information about different physical limitations elderly may have, that needs consideration as a developer.

There is also an improvement section in chapter 4.2 see Table 9, there are subjective opinions from the participants of the study, but are relevant for researchers in HCI that focuses on healthcare technologies.

6.3 Limitations

This thesis was divided in two parts: Firstly to find criteria and a suitable method for mobile applications in the field of healthcare. Secondly to test the findings in practice with end-users, in order to verify the results of the thesis. Due to time constraints, the authors did not have the time to perform the tests, hence there is no verification of the results.

Furthermore, the end result of the thesis could have been different if the authors chose another scientific method to examine and analyze the work. For example if the method was survey the end result could have been different with more respondents but with less specific answers. Also the authors would have to do more interpretations of the answers, since there are no follow-up questions included in surveys.

The authors’ objective was to interview both researchers and professionals in the field of HCI. Some of the participants were both researchers and working as a professional. They had insights on both perspectives. However only one interview was held with a full-time UX-designer, and the authors think that this could limit the results. Also the number of interviews was finally set to seven semi-structured interviews. The goal was to hold ten interviews; this is also a limitation.

6.4 Conclusions and recommendations

The purpose of this thesis was to investigate which criteria that are important when choosing a user testing method in the field of healthcare. Furthermore, the authors wanted to examine which user testing methods that are appropriate for mobile applications in the field of healthcare.
What are appropriate criteria when choosing a user testing method, in the field of healthcare?

The general conclusion of this question is that context is in focus, when choosing a user testing method in the field of healthcare. The analysis showed that healthcare is a complicated domain with many perspectives to cover. Healthcare environments are often hectic and busy, and healthcare staff needs interactive healthcare technologies that will aid them in their work. The technology should not hinder their work, but rather be a tool that assists the healthcare staff in their daily work and routines. Therefore, usability evaluators need to see the real environment, how healthcare staff works, what systems they interact with, in order to understand the whole picture.

Furthermore, the analysis showed that context awareness should be focused on by all usability evaluators, when choosing a user testing method. The authors think that the criterion context is applicable in all domains, and that more user testing methods must be developed, especially methods that are centered around context. Furthermore, the authors recommend that healthcare stakeholders should put a lot of resources into user testing early in the development phase.

The field of healthcare is greatly regulated and is in a high risk environment; therefore the criterion security is vital. The analysis showed that systems must be thoroughly tested, in order to guarantee that errors and unauthorized access to the system is prevented. Another aspect that the analysis revealed is the fact that healthcare often has sensitive data about patients. There are ongoing debates about medical data: Who can access that medical data? And which information should be available for patients? The authors’ opinions about this matter are that new methods are needed, in order to be able to discover usability problems that concerns security in integrated systems. Another point that the authors would like to recommend for companies that develop healthcare technologies, is to think about a feature that detects potential errors. For example, if a user inputs a number that is considered too high, a popup could tell the user that the entered number is not in the interval of the recommended dosage.

The last general conclusion for research question 1 concerns what do you want to test? The analysis revealed that this criterion needs to be considered. The definition of the mentioned criteria is that before choosing a user testing method, usability evaluators must define the following: What is the goal of the test? and Which aspect of the system needs testing? The authors agree with the importance of this criteria, moreover this criterion needs to be acknowledged in all domains that requires user testing. Without this criterion, it is not possible to do proper user testing.

Moreover there are a few criteria that were not identified in the empirical data, according to the theoretical framework five important criteria when comparing UEMs are: realness, validity, thoroughness, effectiveness, and consistency. These criteria are more applicable when comparing different usability testing methods with each other. These criteria could be seen as parameters and to be able to give scores to the UEMs. In this thesis the focus was on mobile applications in the field of healthcare, and thus this could be an explanation of why there is no mention of the five previously mentioned criteria.

Which user testing method is applicable in the context of mobile applications in healthcare?

The analysis showed that there is no single user testing method that is capable of finding all usability problems, in the context of mobile applications in healthcare. Furthermore, the analysis showed that there are a few methods that could be used when conducting user testing on mobile application in the domain of healthcare. The analysis revealed that the following user testing methods could work when testing a mobile application in the field of healthcare: Observational studies, interviews, think aloud and heuristic evaluation. The analysis also showed that some of the standard user testing methods was
not mentioned a lot. User testing methods such as: Surveys, focus groups, field studies, and cognitive walkthrough were mentioned briefly. A possible reason for this could be that there is limited user testing in healthcare. Another explanation could be that the healthcare domain is complex, and the usual UEMs are not suitable for user testing in healthcare.

The analysis also presented an alternative approach when choosing a user testing method in mobile applications in the field of healthcare. The analysis further showed that there are several factors that need consideration: Context, what is the goal of the test? and who are the intended users? These factors need consideration in all fields when performing user testing; it is not specific for mobile applications in the field of healthcare. Defining these factors, would simplify the choice of user testing method for mobile applications in the field of healthcare. The analysis also showed that there are too little user testing conducted in the field of healthcare. The authors believe that actability is a big part of a usability evaluation. Even if a test is successful and many problems are identified, the healthcare organizations should be giving enough financial resources into solving the discovered usability problems. Furthermore, the authors’ think that there is a need for more research to test a whole system. Since healthcare often has many systems that are integrated with each other, there is a need to be able to test all systems combined, instead of testing one system separately. Finally, usability testing methods should be integrated with agile software development methods, in order to be able to make changes in early stages in the development process. This will prevent big changes that would affect core functionality, which may be very costly for stakeholders.

6.5 Future work

For future work the authors would perform user testing with real users, using one of the mentioned user testing methods (see 5.2 Analysis). This would verify if that specific method works in practice and thus is appropriate to use in the field of healthcare.

As previously mentioned by Jaspers (2009) there is a need for usability testing methods that are suitable for the field of healthcare. This domain has in the last years developed and have a lot of interactive technologies. These technologies need proper user testing, and the authors urges researchers to include contextual testing, as the analysis showed the importance of that criteria.

The authors also believe that the term “Actability” need to be introduced more into the HCI community. Even if a company or organization have resources for user testing, they have to be ready to give even more resources to deal with the problems.

Finally, the produced criteria give researchers in HCI guidelines of what to think about in the choice of user testing method, but the authors’ feel that there is a need for further research in the area. The domain of healthcare is broad; it is not limited to hospitals. Healthcare technologies can be applied to home environment, as well as fitness applications in mobile devices.
7 References

Websites


Lind, Mats; Professor at Department of Information Technology, Division of Visual Information and Interaction, Uppsala University. n.d.


**Pictures/Figures**

Appendices

Interview Questions

1. Can you tell us about your experience within the field? Can you describe a typical assignment?
2. Are you familiar with usability in the field of healthcare?
3. Do you have experience performing user testing on mobile applications? What methods are applicable in user testing on mobile applications? Are there any differences compared to traditional user testing?
4. What aspects are important before considering an usability evaluation method? (Can you rank them from 1-5, with 1 being not important and 5 is very important?)
5. Which challenges are present when performing user testing? (Are there any specific challenges, in the context of healthcare?)
6. Do you think there is a need for improvements, when performing user testing in the field of healthcare?