Usability Testing of Software Products
- A case study about usability testing and how to improve the practice

Alice Hällkvist

Systems Sciences, bachelor's level
2019

Luleå University of Technology
Department of Computer Science, Electrical and Space Engineering
ABSTRACT
Usability testing is an important task during development of software products such as web and mobile applications. It is a growing and evolving concept but is often left out during development due to time and costs. Developers are also expected to handle the usability testing but are given minimal training to do so.

The aim of this thesis is to examine how Tieto uses and perform usability testing when developing their software products. The study is about identifying any issues in the practice and compare it to presented theory and come up with suggestions of change that Tieto can take into consideration.

The method in this thesis is single case study done with qualitative data collection through observation and semi-structured interviews, to be able to answer the purpose and get a deeper knowledge about what some of the employees at Tieto thinks of usability testing.

By comparing the result of the study with the presented theory the conclusion can be drawn that the practice of usability testing consists of a few flaws but that the main problem is the organizational culture.
SAMMANFATTNING

Användbarhetstester är en viktig del i utveckling av mjukvaruprodukter så som webb- och mobilapplikationer. Det är ett växande och utvecklande koncept men blir ofta utelämnat i utveckling på grund av tidsbrist och kostnader. Utvecklare är också förväntade att hantera användbarhetstesterna men får minimal träning för att göra det.

Syftet med den här studien är att undersöka hur Tieto använder och utför användbarhetstester när de utvecklar sina mjukvaruprodukter. Studien handlar om att identifiera eventuella problem i praktiken och jämföra den med teori och presentera förändringsförslag som Tieto kan ta hänsyn till.

Metoden för det här arbetet är fallstudie med kvalitativ datainsamling genom observation och semistrukturerade intervjuer, för att kunna svara på syftet och få en djupare kunskap av vad några av Tietos anställda tycker om användbarhetstester.

Genom att jämföra resultatet av studien med den presenterade teorin kan slutsatsen dras att praktiken av består av några brister men att huvudproblemet är organisationskulturen kring användbarhetstester.
ACKNOWLEDGMENTS

This is a bachelor essay for the division of Data- and System Sciences by the Department of Systems and Space Engineering at Luleå University of Technology.

I would like to express my gratitude towards Jonas Karlsson at Tieto who made this thesis possible and helped me to get started. I would also like to thank the people who agreed to being observed and interviewed or was somehow affected by this work. Big thanks to my supervisor Jörgen Nilsson, who has guided me and provided tips throughout the process.

Finally, I would like to thank my family and friends for their support, help and encouragement during the work of this bachelor essay.

Skellefteå 23 Maj 2018

Alice Hällkvist
# TABLE OF CONTENTS

Abstract ......................................................................................................................... I
Sammanfattning ............................................................................................................. II
Acknowledgments ........................................................................................................ III

1. Introduction ........................................................................................................... 1
   1.1 Background and Problem Area ................................................................. 1
   1.2 Limitations ................................................................................................. 1
   1.3 Purpose ...................................................................................................... 1
   1.4 Research Questions .................................................................................. 1
   1.5 Concepts .................................................................................................... 2
      1.5.1 Usability ........................................................................................... 2
      1.5.2 Effectiveness ..................................................................................... 2
      1.5.3 Efficiency ......................................................................................... 2
      1.5.4 Satisfaction ...................................................................................... 2
      1.5.5 Learnable .......................................................................................... 2
      1.5.6 Accessible ........................................................................................ 2
      1.5.7 Usability Testing ............................................................................... 2
      1.5.8 Usability Testing vs User Testing ..................................................... 3
      1.5.9 Human-Centered Design & User-Centered Design ....................... 3
   1.6 Structure of Thesis ...................................................................................... 3

2. Theory .................................................................................................................. 4
   2.1 Literature Review ...................................................................................... 4
   2.2 When Should You Test? ........................................................................... 4
   2.3 Different Types of Usability Tests ........................................................... 5
   2.4 The Process for Conducting a Test ............................................................ 6
   2.5 Who to Test ............................................................................................... 6
   2.6 What to Test .............................................................................................. 7
   2.7 Where to Test ............................................................................................ 7
   2.8 Why Usability Testing ............................................................................ 8
   2.9 Attitude Towards Testing ....................................................................... 8

3. Method .................................................................................................................. 9
   3.1 Research Approach .................................................................................. 9
   3.2 Data Sources .............................................................................................. 9
      3.2.1 Field Observation ............................................................................ 9
      3.2.2 Semi-Structured Interviews ........................................................... 10
   3.3 Qualitative Data Analysis ........................................................................ 11
   3.4 Method Discussion ................................................................................... 11
### 3.5 Trustworthiness

- 3.5.1 Credibility
- 3.5.2 Transferability
- 3.5.3 Dependability
- 3.5.4 Confirmability

### 4. Result

- 4.1 The Case
- 4.1.2 Product of Test
- 4.2 Observation
- 4.3 Interviews
  - 4.3.1 Lack of Internal Education
  - 4.3.2 Performance of Usability tests
  - 4.3.3 User Feedback
  - 4.3.4 Building the Right Things
  - 4.3.5 Organizational Culture
  - 4.3.6 Lack of Testing

### 5. Analysis and Discussion

### 6. Conclusion

- 6.1 Suggestions of Change
- 6.2 Proposals for Further Studies

### 7. References

### Appendix

- Interview Questions
1. INTRODUCTION

In the introduction the background to the problem area will be introduced, as well as the limitations to the study, the purpose, research questions, explanation of important concepts and product description. Lastly, there is a short summary about the structure of the remaining thesis.

1.1 BACKGROUND AND PROBLEM AREA

Usability testing is a concept that continues to grow and evolve (Dumas & Redish, 1999), and is a common tool used to evaluate the usability of a mobile application in a development process (Kaikkonen et al, 2005). Usability testing of software applications is a growing research area with different challenges due to unique features of mobile devices (Zhang & Boonlit, 2005). With the fast advances of mobile technology and applications is effective usability even more important for the design, development, and deployment of successful applications (Zhang & Adipat, 2015).

Even though user centered design is gaining in popularity is usability barely considered as one of the primary goals in many design scenarios (Raza & Capretz, 2009). A very common problem according to Raza & Caprez, (2015), with programmers and software developers, is that while designing they think that if a design is good enough for them to be used it will be for the targeted user as well.

1.2 LIMITATIONS

The focus of this study is on how usability testing is conducted during development of software products, such as web and mobile applications.

1.3 PURPOSE

The purpose of the study is to observe and interview people about the use and performance of usability testing at one organization. And through analyzing and comparing the gathered data with theories and practices in literature, give suggestions on how to improve the current usage of usability testing.

1.4 RESEARCH QUESTIONS

From the problem area and the purpose of the study the following research questions can be formed:

“What are the issues in the practice of usability testing of software products?”

“How can the practice of usability testing improve?”
1.5 CONCEPTS
This is a short section of word concepts encountered in this study with their definition to clarify their meaning.

1.5.1 USABILITY
Dumas & Redish (1999) writes about usability testing and their definition of usability is as follows: “Usability means that people who use the product can do so quickly and easily to accomplish their own tasks”. This definition comes from four points:

- Usability means focusing on users
- People use products to be productive
- Users are busy people trying to accomplish tasks
- Users decide when a product is easy to use

Another definition is the international standard, ISO 9241-11. The ISO standard provides guidance on usability and define it as: “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.” (Iso, 2018).

1.5.2 EFFECTIVENESS
Accuracy and completeness with which users achieve specified goals (Iso, 2018).

1.5.3 EFFICIENCY
Resources expended in relation to the accuracy and completeness with which users achieve goals (Iso, 2018).

1.5.4 SATISFACTION
The extent to which the user's physical, cognitive and emotional responses that result from the use of a system, product or service meet the user’s needs and expectations (Iso, 2018).

1.5.5 LEARNABLE
Learnability is a part of effectiveness and has to do with the user’s ability to operate the system to some defined level of competence after some predetermined amount and period of training (which may be no time at all). It can also refer to the ability of infrequent users to relearn the system after periods of inactivity (Rubin & Chisnell, 2008).

1.5.6 ACCESSIBLE
Accessibility is in broader terms about having access to the products needed to accomplish a goal. It can also be about what makes products usable by people who have disabilities and how to make the product usable for them (Rubin & Chisnell, 2008).

1.5.7 USABILITY TESTING
Rubin & Chisnell (2008) refer usability testing to a process that employs people as testing participants who are representative of the target audience to evaluate the degree to which a product meets specific usability criteria.
1.5.8 Usability Testing vs User Testing

Usability testing means to test the usability of a product and user testing refers to test the idea of a new product (Mishra, 2016).

1.5.9 Human-Centered Design & User-Centered Design

Human-centered design is defined by ISO 9241-210 as an approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system, applying human factors/ergonomics, usability knowledge, and techniques. This approach enhances effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability; and counteracts possible adverse effects of use on human health, safety, and performance.

The term “human-centered design” is used rather than “user-centered design” in order to emphasize that this part of ISO 9241-210 also addresses impacts on a number of stakeholders, not just those typically considered as users. However, in practice, these terms are often used synonymously. Usable systems can provide a number of benefits, including improved productivity, enhanced user well-being, avoidance of stress, increased accessibility and reduced risk of harm (Iso, 2010).

1.6 Structure of Thesis

The remaining structure of the thesis is followed by a theory section, method description, result of observation and interviews, analysis and discussion section and conclusion with suggestions of change. Lastly, a reference list of used books and electronic resources, and appendix with the interview questions.
2. THEORY

This section of the study is the literature review of the theory for the study. It describes best practices of usability testing, how to perform and conduct testing, when to use different types of tests and why testing is important. It also consists of a short section about the attitude towards testing.

2.1 LITERATURE REVIEW

There is a lot of theories behind usability testing, on how to conduct the testing and why some projects succeed and why some do not. Testing has always been an important part in any development project but is not an exact science. There is still little known about software development (Myers, Sandler & Badgett, 2011). The history of the profession of usability started as early as the 1980s. Many methods come from earlier fields of Ergonomics and Human Factors that began near the beginning of the 20th century and had a strong influence through World War II. Books and papers of the subject has continuously been released since (Measuring U, 2013).

Usability continues to evolve and new methods are arising. For example, eye tracking has developed and more studies regarding usability testing are conducted. Usability testing is seen as a key element in user-centered design and studies has shown that it increases the chance of developing usable software (Thompson, Rozanski & Haake, 2004).

Usability testing of software products such as web applications and mobile apps during the development process is important. Even though the high importance of testing to make sure the software is usable by the targeted user, it is not always a high priority (Rubin & Chisnell, 2008). It is important to conduct usability testing throughout the whole designing process and the product life cycle. Rubin & Chisnell (2008) describes why testing is important and when you should conduct testing. Their book provides step-by-step guidelines to help testing the usability of a product. It provides help on how to plan, design and conduct effective tests. According to them, for a product or service to be “usable” it must be useful, efficient, effective, satisfying, learnable, and accessible.

2.2 WHEN SHOULD YOU TEST?

It is important to involve the user early in the project. A direct contact between users and the design team throughout the whole development life cycle is useful. Thus, people who handles these interactions must be trained to do so. The demand for usable products is higher than the number of available and trained professionals who can conduct the testing. Therefore, many product developers, engineers, system designers, technical communicators, and marketing and training specialists have had to take the roll with primary responsibility for usability within their organization. Since they lack training in usability engineering or user-centered design they might be asked to do tasks they are not prepared for (Rubin & Chisnell, 2008).

Since testing takes time is it important to know when to test, not test too much or too little, and when to test. Rubin & Chisnell, (2008) describes four different types of tests: exploratory (formative), assessment (summative), validation (verification) and comparison test. These four tests are suitable for general phases that any product development cycle goes through. Figure 1 illustrates the usability testing throughout the product lifecycle.
Exploratory testing is used early in the development cycle and is followed by assessment testing. Validation is done after in the final testing and product launch phase. Exploratory testing on features and follow-up assessment tests is in needs analysis and requirements gathering phase. The comparative tests can be conducted together with the other tests during the whole lifecycle.

![Figure 1. Usability testing throughout the product lifecycle.](image)

Rubin & Chisnell (2008) also talks about the importance of early testing when critical design decisions are about to be made and lay ground for following development. A bad start and wrong assumptions will probably lead to usability problems later.

### 2.3 Different Types of Usability Tests

There are many different types of usability testing. As mentioned earlier Rubin & Chisnell (2008) describes four basic tests. Exploratory, assessment, verification tests and comparison. The exploratory test is used to examine the effectiveness of early design concepts, such as how well it supports user’s tasks within a goal, communicates intended workflow, and explore different kinds of issues. This kind of testing tries to answer questions like; “how easily and successfully can users navigate?”, “what do users conceive and think about the product?”, etcetera. Assessment tests are used to expand the findings from the previous exploratory test. It is done by evaluating the usability of low-level operations and aspects of the product. The objective of verification tests is to evaluate how the product compares to some predetermined usability standard or benchmark. The standard could be project-related performance standard, internal company or historical standard (Rubin & Chisnell, 2008).
Lastly, comparison tests can be used together with any of the previous mentioned tests. It is used to compare two or more designs. It can be two different interfaces styles, or the design of a manual with a proposed new design, or to compare a product with a competitor’s product. This kind of test is often used to decide which design that is easier to use or learn, or to understand the advantages or disadvantages of different designs.

2.4 THE PROCESS FOR CONDUCTING A TEST

The process for conducting usability tests is long and consists of many steps and things to think about before, during and after the test. Rubin & Chisnell (2008) describes the process in eight different steps:

1. Develop a test plan
2. Set up a testing environment
3. Find and select participants
4. Prepare test materials
5. Conduct the test sessions
6. Debrief the participant and observers
7. Analyze data and observations
8. Report findings and recommendations

These overall steps over the process consists of multiple steps themselves. They provide checklists for some of the important steps.

A few important steps to use when conducting the tests are listed below:

- Script with introduction
- Offer refreshments
- Explain why they are there
- Explain any unusual requirements
- Mention that it is okay to ask questions at any time
- Inform that he or she is not being tested
- Ask for any questions

Dumas & Redish (1999) provide similar steps as Rubin & Chisnell (2008). They start off with “planning and preparing for a usability test” which consists of 11 different steps. Another part they mention is “conducting and using the result of a usability test”, which describes how to conduct the test and how to analyze and communicate the results.

2.5 WHO TO TEST

One of the steps Rubin & Chisnell (2008) describes is “find and select participants”. When conducting usability tests, is it essential to select participants with background and abilities that are representative of your product’s intended user. If wrong people are tested the results will be questionable and of limited value, even if you put much effort into rest of the test preparation. To select the participants is done by identifying and describing the relevant behavior, skills, and knowledge of the persons who will use the product. A user profile is used as description of the targeted users and should be developed early in the product development.
After the user profile is determined it is important to ascertain the most effective way of choosing people from the targeted audience, within limits of time, budget, resources and so on.

It is also necessary to determine the number of participants to test. How many that are needed is based on many different factors, such as:

- The degree of confidence required
- The number of available resources to set up and conduct the test
- The availability of participants
- The duration of the test session

2.6 WHAT TO TEST

What is it that is going to be tested? A paper prototype or a real functional product? An advantage of conducting usability tests of prototypes, is that it is in early stage and is not invested in much time or other resources for a completed product yet. In iterative testing is each prototype tested and improved, which leads to better and more robust versions (Rubin & Chisnell 2008).

Paper prototypes or low-fidelity prototypes are hand drawn or hand formed prototypes with a few or no working parts. The planned elements of the user interface are sketched on separated pieces of papers so that they can be moved around, added to, or changed during the test. These kinds of testing are naturally done early in the design process and can even be performed to test concepts to determine usefulness and feasibility of an idea before any further investment (Rubin & Chisnell 2008).

Clickable or usable prototypes are closer to the real product. In these kinds of tests some parts should be clickable or interactive. It could also be static paths which means that only parts of the interface are actually interactive, they look as they are live, but nothing happens when clicked. During a testing with these kinds of prototypes needs a bit more skill compared to a full working prototype since statements to explain the situation to the test person is needed (Rubin & Chisnell 2008).

2.7 WHERE TO TEST

There are two main locations or ways to conduct usability testing. Either a field test or a more controlled environment in a usability test lab. The technology and the use of usability testing has changed over the years. Rubin & Chisnell (2008) provides factors to consider when determining the appropriate location for the usability testing. There are many factors to consider when deciding to conduct the testing in a lab or at the user’s site. For example, is it needed to sit next to the participant to collect qualitative data, need to prompt or give hints, or manage the product being tested?
These two approaches as different advantages and disadvantages. Kaikkonen et al. (2005) did a study comparing field and lab testing. They came up with the conclusion that field testing took more time than the controlled lab environment but that it did not give much difference in the result.

2.8 Why Usability Testing
The reason why usability testing is important for most companies is to improve the profitability of products. Data gathered from users help expose design issues, so they can be remedied and minimize or eliminating frustration for users. By including usability testing throughout an iterative design process, it is possible to make products and services that are useful, usable, and maybe even delightful (Rubin & Chisnell, 2008).

By involving the user during the development through usability testing it increases the user acceptance of the product. According to Davis (1993), by using the TAM model, it increases the chance of user acceptance of the software.

2.9 Attitude Towards Testing
Dumas & Redish (1999) has worked with many organizations over a few years. They have noticed that the perception about the need for usability testing is uneven among people in many organizations. Some people are enthusiastic about having a user-driven approach, while others are more skeptical and indifferent about the value of usability testing. Even though user centered designs are gaining popularity is usability still barely considered as one of the primary goals (Raza & Capretz, 2009). To design a usable system is difficult but many organizations still treat is as a common sense (Rubin & Chisnell, 2008).

What is the reason for the lack of usability testing? Does it cost too much and take a lot of time that usability tests are not performed as much as it should? Dumas and Redish (1999) writes about a few reasons of why it is difficult to achieve quality (usability) in documents and other parts of products. One reason is that managers are forced to focus on headcounts rather than an overall budget, which means that they must assign programmers to write online help or even print documentation instead of hiring technical writers. Developers and writers are also given minimal training and expected to conduct usability testing because there are no spots to hire trained usability staff. Another reason is that usability and documentation are low priorities.
3. METHOD

This section describes the chosen research method for the thesis, as well as how the data collection were done, qualitative data analysis. The section also consists of a method discussion and description for trustworthiness of the research.

3.1 RESEARCH APPROACH

The research approach for this study is single case study done with qualitative data collection through observation and semi-structured interviews. This method helps to study a case in depth, in real-life settings and guides through the work with a variety of data collection methods. The method also The study was done during four months in the winter/spring 2018, and the observation and interviews were done during this period. The observation was performed on team Kafka and their testing of a mobile application view for students.

Yin (2003) writes in his book about case study research and how to implement it in a study. He describes the case study as a linear but iterative process that starts off with a plan of the study, followed by design, prepare, collect, analyze and share. These different steps consist of different tasks or aid to carry through with case study research. He defines the case study research method “as an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used.” According to Yin, case study is useful in studies with questions such as how and why.

3.2 DATA SOURCES

Yin (2003) talks about collecting multiple sources of evidence. Documents, archival records, interviews, direct observation, participant-observation and physical artifacts. These are six sources of evidence that may be used in the case study. In this study, the data is collected through field observation and interviews. Firstly, the observation was made to see which methods, different types of tests they use, and how they work at Tieto and the interviews were important to get a deeper understanding of their thoughts and attitude towards the testing.

3.2.1 FIELD OBSERVATION

Observations are a way to capture the life situation of the people being observed, which makes it view private things and information that people do not want to expose might be published. Therefore, are all the people being observed in the teams anonymous to limit the ethical consequences (Ahrne & Svensson, 2015).

Ahrne & Svensson (2015) talks about two different kinds of observations, open and closed observations. These two techniques have different advantages and disadvantages. Open observation means that the participants know about and has accepted the observation. Closed observation however, means that the participants do not know about the observation taking place. In open observations it is important to accept the group being observed and have their permission and build trust. In this study an open observation was performed, and all the people were informed about the observation and agreed. The strategy was to passively participate and take notes on the computer and phone during the observation.
The focus is on one organization, Tieto, and team Kafka acts as the case. To get as much time with Tieto and the team as possible I had my own place in the team room. It helps getting a better and focused view in the study. It may take time to get to know the location and get connections to the people there (Ahrne & Svensson, 2015).

3.2.2 Semi-Structured Interviews

The purpose of the interviews is to get a deeper understanding and knowledge about people’s attitude and thoughts towards testing and the performance of testing today at Tieto. Qualitative interviews help study feelings, experiences, thoughts, intentions, how power works, and decision making (Ahrne & Svensson, 2015).

An advantage with qualitative interviews is that it can be done in different ways, switch between more structured questions and decide how to complement with more opened questions. It is also possible adapt the questions and to decide which order to ask questions (Ahrne & Svensson, 2015).

The interviews in this study are semi-structured and were recorded using a free recording application on a Mac computer and transcribed by hand using oTranscribe.com. The interviews were conducted in Swedish, thus, it is easier and less limiting for both the informants and myself when doing it in our native language. The entire interviews were transcribed. They were later summarized and translated into English. The interview questions were grouped into different themes, as illustrated in the appendix.

All interviews were conducted in March 2018, through scheduled meetings that took about 30 minutes. They were conducted during working hours at Tieto in empty office rooms with low noise level to be able to record the conversation, with agreement from the informant. The interviews were recorded to be able to transcribe the conversation and to be sure that nothing is left out in the result (Denscombe, 2009).

Five people were interviewed, one UX-designer, two developers from a team that use usability testing often, and one developer and scrum master from one team with less usage of usability testing. The chosen participants are people with different knowledge about the subject or are affected by the observation. The purpose with the selection of informants is to get a view on how developers from different teams with different amount of performance of usability testing and to hear what they think about usability testing. All the informants were informed about the reason for the interview and explained how the results of the interviews will used in the report. They all also agreed to being recorded and were informed that they have the possibility to read the report before being published, to make sure that they are aware of the usage of their answers (Ahrne & Svensson, 2015).
3.3 Qualitative Data Analysis

The analytic strategy in this study is to compare the data from the observation and interviews with the theories presented in the theory section. To analyze qualitative data such as observations and interviews, there are three different main strategies. To sort, reduce and argument. The collected data needs to be sorted into different themes and qualitative data collection often results in large amounts of data, which needs to be reduced to manageable level. The sorted and reduced data is lastly argued for (Ahrne & Svensson, 2015). The steps in this analysis is following:

- To sort and summarize the interviews with analysis of the received answers. The interviews were sorted and compared to find similarities and differences in the answers
- Compare the observation with presented theory
- Argument the analysis

Thematic analysis is used for sorting and reducing the interview data. This is done in 6 different phases. The interviews were in the first phase transcribed entirely word by word. The transcribed text was then read multiple times to get a view of the material before phase two. In phase two were important words and sentences identified, words or sentences that are recurring or different throughout the interviews and these are the first candidates of codes. Potential themes are found in phase three and the codes are thereafter placed within the themes. The themes are then compared to the theoretical starting points the study affects (Cope, 2007).

In the fourth phase is the themes named with informative titles that connects with the codes. In the fifth phase will some of themes be removed and other arise to maintain the structure. The last and sixth phase is the writing of the result. The interviews will be summarized in coherent text under the chosen themes (Braun & Clarke, 2006).

The created themes that will be used in the result are: lack of internal education, performance of usability tests, user feedback, organizational culture, building the right things and lack of testing.

3.4 Method Discussion

It was difficult to follow the chosen method since the study had to adapt to the team’s project cycle. Therefore, I did the theory research before and after the observation, which might have affected the research as less objective to the subject. It is possible that it limited the theory search. The analysis of case study is not well developed either and is a difficult part of case studies. Although, the method gives good guiding in the study and the data collection.

Since observations and interviews are new for me as a researcher some parts might be missing, for example how to conduct and take notes of observations. I also had no deeper knowledge about usability testing before this study. The only experience of usability testing and user-centered design I had prior to this study is an UX course and being participant in one usability test.

It is difficult to apply this study on the whole organization because only one team was studied, which gives a small amount of material to generalize with. Teams within Tieto also
uses different methods and might not look the same and give the same issues regarding the usability testing.

To schedule some of the interviews took time because of waiting for email answers, which made the interview process take longer than anticipated. There is also a risk of the “interview effect”, which means that informants answers questions depending on what identity, such as sex, age and as a student, of the interviewer. It is limited by a giving a neutrality and try to make the informant comfortable and feel that they can give honest answers (Denscombe 2016).

3.5 TRUSTWORTHINESS

Guba & Lincoln (1985) was early with establishing the concepts credibility, transferability, dependability and confirmability for trustworthiness of research. The positivists concepts of reliability and validity cannot be addressed in the same way in naturalistic work. They describe trustworthiness as to how an inquirer can persuade the audience that the findings are worth paying attention to, worth taking account of, and also what arguments can be used, what criteria invoked and what questions to ask to be persuasive. To establish trustworthiness in this study the four concepts are taken into consideration.

3.5.1 CREDIBILITY

Guba & Lincoln (1985) describes five different activities to improve the probability that the findings and interpretation to be found credible. In this study is prolonged period of engagement and triangulation used. The period of prolonged engagement is intended to provide the investigator the opportunity to build trust. To build trust with the people being observed and interviewed I have worked closely to them and they have had the opportunity to read through the collected data. Triangulation can be established through multiple and different sources, methods, investigators, and theories. The use of multiple sources allows a broader range of issues such as, historical, attitudinal and behavioral. The data collection is in form of semi-structured interviews and observation. The data is also documented in multiple ways.

3.5.2 TRANSFERABILITY

Transferability can be established by providing the readers with evidence that the findings in the study could be applicable to other situations, contexts, times and populations. This is done by describing the study in detail and presenting the data.

3.5.3 DEPENDABILITY

Dependability is used to establish the research study’s findings as consistent and repeatable. This study is read by others to ensure the accuracy of the findings and that they are supported by the data collected.

3.5.4 CONFIRMABILITY

Lastly to establish trustworthiness in this study is the use of confirmability, which means the level of confidence that the research study’s findings are based on the participants’ words and not potential researcher biases. The technique audit trail is used to establish confirmability by detailing the process of data collection, data analysis and interpretation of the data.
4. **RESULT**

In this section is the result of the study presented. The observation and interviews will be described in summary.

4.1 **THE CASE**

Tieto is an IT software and services company providing IT and product engineering services. They strive to be the customer's first choice for business renewal, and the leading software and service company in the Nordic region. Tieto is located in 20 different countries, a few of them are; Sweden, India, Czech Republic, Latvia, and Finland. The headquarter is located in Esbo, Finland. Tieto consists of several branches; Education, Health and Welfare, Energy, Telecom and so on (www.tieto.com).

In Skellefteå Tieto has five branches; Education, Telecom, Welfare, PDS and Value Networks. Tieto Education has 11 different development teams. The names of a couple of them are Kafka, Sherlock, Garfield and Albert. The office in Skellefteå has about 260 employees, in Sweden there are 1800. The total employment is 14 000 (Tieto intranet).

4.1.2 **PRODUCT OF TEST**

Tieto has developed a mobile application for parents and guardians to keep track of their children’s activities in preschool and school. Its purpose is for guardians to easily find information and keep updated on important events daily. The Tieto Education is successively implementing new easy handled services to the application. Tieto Edu wants to provide services for teachers, parents, and students, that will create a seamless communication between school and home. Tieto Edu is available for preschools and schools using the Tieto Educations learning platform, and has activated access through mobile application, and also a parent with an account on the platform. The application is available for both Android and iOS, on mobile and tablets.

The next service that is under development is a view for students. The students are planned to be able to view their schedule with information such as subject, classroom, and time. They will also be able to view and send messages to teachers, view notifications, and view their study plan. The service went through a design sprint to come up with a design fast and a usability test was performed on the design. The new application view is still under development and more testing will be conducted.

4.2 **OBSERVATION**

Two days before the scheduled usability testing occasion at Balder gymnasium in Skellefteå, team Kafka started to plan and write scenarios. Three different scenarios were written with a couple of complementing questions, to test wanted functionality. The test was going to be done on two devices (Android and iOS), with two different versions (A and B). The A version has navigation of services with icons at the front page and version B has navigation at the top right corner under a navigation tab. One purpose of the usability test was to see which version the test persons, in this case the students, prefer. A few problems during the small meeting was discovered and was planned to be fixed during the day and the following. How many test persons and testers, and the overall test was discussed. I got the feeling that they were not ready for the test yet.
One day before the test they finished the last preparations of the test by fixing the noticed faults from the day before and conducting a pilot test in-house at Tieto. The pilot test is conducted to catch any problems that might occur during the testing. Two developers performed the testing, one explained the reason of the test and read the scenarios, and the other took notes. I sat quiet and observed. Two employees at Tieto got to participate, they got to try both versions on two devices and answer the complementing questions. The tests went well, and no big issues was discovered. A few buttons were difficult to press, but it was decided to not change anything for the big test the next day. However, it was found that a short introduction about why the test is conducted and where they come from is needed. It was also discussed how we should sit to not make the test person nervous or feel pressured. The test will be run through VPN to be able to use the application at Balder. And the application is partly functional, some of the functions is clickable but not consisting of further functionality.

The day of the test at Balder. I met with the testers at the school around 9 am. They had brought the equipment needed and refreshments in form of a soda, for the test persons to receive after conducted test. The location was chosen in the school’s cafeteria. A table for four people, the Tieto developers, the test person and I. The volume from people and music in the background was high. At the location it was discovered that a button to view full schedule was not working correctly. Later on, during the test the button started working again. They took turns in finding students to do the usability test on. It was a bit difficult when the students had classes or started going on lunch break. One was responsible for introducing and explaining the reason of the test, that he/she is not being tested, if anything is unclear it’s not their fault but the application, for the chosen test person and the other takes notes. The tests were documented with pen and paper during simultaneously during the testing. The successfulness of the scenarios was documented with a percentage scale over the test persons’ completion of the task. 100% illustrates that the person managed to follow the task by themselves without any issues, 50% illustrates that they managed to complete the task with some help or issues, and 0% illustrates that they did not complete the task. The answers to the complementary questions were also written down and if the user makes any certain move in the application.

During the tests was the VPN connection lost a couple of times and some of the buttons were difficult to press for a few of the users. When the test person got stuck or felt a bit lost he or she was given a while to think for himself before the person conducting the test helped by asking questions or give small guiding hints. If the participant was still unsure, they continued with the next scenario. Some of the tests participants were distracted by friends or nervousness. One test participant watched when another conducted the test which might have influenced the answers.

The test was performed on a total of thirteen students, six guys and seven girls. The first six started with one of the versions and the remaining six got to start off with the other version. The last test was an exploratory test. The test person got to look at the application without the scenarios and come up with proposals of functions to add to the app.

The analysis of the test lead to the decision of choosing one of the two versions. The students preferred version A, with navigation icons at the first page. The development team decided to follow the feedback from the test and use version A.
4.3 Interviews

Through transcribing of the interviews and analyzing, the relevant information has been chosen and generated six different themes. These themes are lack of internal education, performance of usability tests, user feedback, organizational culture, building the right things and lack of testing. Informant 1 represents the UX-designer, system developers from team Kafka as Informant 2 and informant 3. Developer from team Cortanza as informant 4, and lastly, scrum master from Cortanza as informant 5. The informants are all working at Tieto Education in Skellefteå, with working experience at Tieto within a range from two to 17 years. Informant 4 has no real experience of usability testing.

4.3.1 Lack of Internal Education

None of the informants has received any actual internal education from Tieto regarding usability testing. There have been a few workshops, but no further education or training provided by Tieto. The informants have gained knowledge mostly through self-studies. Informant 2, informant 3 and informant 4 has gained their knowledge through self-studies and by receiving help and guidance from the UX-designer. They have no other experience prior to working at Tieto either. Informant 2 however, was lightly introduced to usability testing when working on a bachelor thesis at Tieto, were the informant got to perform a few usability tests.

Informant 1 believes that it is more important to try and educate people about the need of usability testing and to raise interest rather than educating everyone at once about usability testing. Informant 2 argues that it might be good with some sort of education or training for people that are new to usability testing and needs some sort of start with proposals on how to. The informant also informs that they help and try to inspire people to start testing. Informant 3 describes their own experiences within the team and that they did not really know where to start when testing was new to them. They learned by doing and received help from the UX-designer when needed.

According to informant 3 and informant 4 there is a budget for Tieto employees to use for training, but it is up to the teams and developers themselves to decide if they want to take an internal course and informant 3 believes that it is not used very often. Informant 4 mentions that there have been education opportunities but that there has been no time to participate.

4.3.2 Performance of Usability Tests

The informants think that the current performance methods are working well and that there is no need for them to change at the moment. Informant 1 believes that the usability tests are performed at the right times and are improving but believes that there is still a long way to go. Informant 1 said following regarding improvements in the way they test:

“There are probably many faults, but that we perform tests and practice on it makes us better, therefore it is important that we actually do it.” – Informant 1

Informant 2 thinks that the test sessions are well performed with well described scenarios and good tools to handle them. However, the informant would like to make the developed products available for the users as a complement to the usability tests. The methods used today is that they first pilot test the test itself at Tieto. Sometimes they end there but otherwise the next step is to test it on real users, according to informant 1. Informant 2 informs that they follow scenarios and documents the results in a scale from how the users can finish a task or scenario. Everything is printed down afterwards, and decisions are made based on the results.
They often take the opportunity to ask for other feedback, if the participants have any comments of what they would like to have in the application. Informant 2 also talks about the observed test occasion and mentions that unstable test environments might affect the test result because that the user might lose focus and the testers finds it irritating.

One flaw according to the informant 2 is the human factors, meaning that it is not easy to know why a user reacts a certain way and it is difficult to get 100% accurate data. Informant 3 thinks that they can be better at evaluate the test and talk to the users and to evaluate the used method.

4.3.3 USER FEEDBACK

The informants talk about the importance of user feedback when developing software products. Continuous feedback from users during development limits the risk of building the wrong things according to informant 2 and 3. By letting the user test the product frequently it is easier to verify the predetermined assumptions and build what the user wants. Since informant 4’s team does not conduct usability testing their feedback process is a bit different. When they release something new they let a few customers be pilot customers and try it out, and later come with feedback. Some feedback comes in through the support regarding functions that the customers thinks are missing or needs to change.

4.3.4 BUILDING THE RIGHT THINGS

Informant 2 describes as usability testing as one of the most optimal way to find out that right thing is being done.

“Well, it is one of the most optimal way to know that you are doing the right thing instead of just sitting and guessing. Yes, but maybe they want this, it is when you have chosen, and they try it as you actually get the answer. Is this something you actually need?” – Informant 2

Informant 3 also talks about the importance of continuous feedback to decrease the risk of building the wrong things for a long time. By limiting the risk of building wrong things and perhaps building too much the development process can be more effective and spend less time on unnecessary things. It would also decrease costs and extra work.

“Say we are building the wrong thing, about half of things that we should not have built at all. We will still have to maintain those things, we still have someone using them, and if there is a bug we have to fix it, and it may be complex because you have these things. Everything that we do extra, which is not really needed, is an enormous cost and that is where the speed comes from. […] So, you can already understand when you build things that will not really be needed. So, that is where, in usability testing or UX at all, it’s true where the big money is to earn. As I interpret it.” – Informant 3
4.3.5 Organizational Culture

All the informants have positive attitude and finds usability testing as an important part of application development and they all agree that the testing is something that helps and improve products.

Tieto as an organization is positive towards usability testing but could put even more into it and make sure it actually happens according to informant 3. There is possible to set aside time for people that can coach and work with teams. It might cost money in the beginning, but it is needed to make that investment. Informant 5 is on the same track and believes that Tieto finds usability testing as important but does not follow it the whole way through. It is not always that Tieto provide those people who can help and push that development either. Tieto should provide necessary training and spend more time on the subject.

At the moment it is up to the teams and developers themselves to adapt and use a user-centered design. According to informant 1 is the current working approach is more develop-centered design:

"But if you start from the beginning, then we do not really work with user-centered design, but perhaps, develop-centered design. Thus, working from your perspective and assuming something that you produce yourself or know that it will work with the end user, that is a risk if you see it that way. So, whatever you design, you have to test it on the user to see if it works or not, it is just as simply as that." – Informant 1

Some teams have to take the initiative and succeed in convincing people that it is a good thing says informant 2. Informant 3 is saying that many thinks it is a waste of time and that they do not have the time to do it. Informant 1 emphasizes that people have different interests and respects that.
4.3.6 Lack of Testing

All the informants strongly indicate that the amount of performed usability tests needs to increase. Most of them argues that through continuously performing usability testing will improve the practice. Informant 1 mentions the importance of all teams performing usability testing and to learn more about the users through both usability testing and other kinds of research. Informant 3 provides similar thoughts:

"I think that even if one has not been used to it, I think that just putting the software in the hands on someone that it is useful, even if you are not a super expert on testing, I think that one still learns just by watching someone else uses things. I think that only that is worth something, no matter what. But yes, it is perhaps possible to improve the testing, but it still has a value in itself." – Informant 3

Informant 3 is also saying that smaller things can be tested on anyone, for example someone in the building that has not worked with the developed product. It is less time consuming and should be done more often according to the informant.

Far from all things developed is being tested according to all the informants. Informant 2 describes that they try to perform tests as much as possible, especially when they are feeling unsure about something or have a few different solutions. It can also be that they feel certain of what to do and therefore skips the usability testing. When they are short on time or when testing is not calculated in the budget is usability testing often left out according to informant 2 and 3. Informant 2 says that the validation step can be left out sometimes, but they try to do test as often as possible. It is easy to prioritize away because it is not something that shows or punishes you right away and has a slight delay, according to informant 3.

It is also up to the teams themselves to decide if they want to conduct usability tests during development. Informant 3 talks about people being interested and good at different things and that you cannot force people to conduct testing if they do not want to. Informant 5 argues the opposite. The informant thinks it might be necessary in some extent to force people to be able to get them started with usability testing.
5. Analysis and Discussion

In this section the result will be analyzed and discussed with the presented theory in the study.

Comparing the presented theory materials with the collected data from the observation and the analysis of interviews it shows that the observed test at Balderskolan consists of most of the steps presented by Rubin & Chisnell (2008). They started with a meeting to plan and write the scenarios before the test occasion and talked about who to test. They also noticed a few faults in the application and planned to get them fixed. The intended users in this case was students since the developed application is a student view for Tieto Edu. They knew what group of participants to use in their testing but had not decided beforehand exactly who to select. The participants were randomly selected at the location, but it was evenly distributed between the sexes.

The test was conducted as a field test, which means it was at the location of the participants. The chosen spot had high noise level and other students sometimes came to watch and perhaps influenced each other, which might have affected the test result. However, the environment is where the users would most likely use this application. The application was partly functional with clickable functions, which Rubin & Chisnell (2008) describes as a prototype closer to the real product. Of the four different test types presented by Rubin & Chisnell (2008) did they use exploratory and comparison test. The developers wanted to test how well the intended user could finish a given task and explore different kinds of issues at the same time compare two designs. After the test they analyzed the data and made a decision to use one of the tested versions.

They also followed most of the presented steps with the performance of the testing. They had a script with an introduction where they informed the participants about the purpose of the test. They also explained to the participants that he or she is not being tested and if anything is unclear it is the developers fault. To motivate students to participate they offered refreshment in form of a soda.

Dumas & Redish (1999) talked about the perception about the need for usability being uneven among people in many organizations and that developers are given minimal training and expects to conduct usability testing. From the interviews it shows that the lack of internal education is recurring and the believe to conduct testing even more. The interviews showed clear similarities between the answers of the informants. Possibly, they did not feel like they could give honest answers and provided answers that they believed were “expected”, and therefore giving similar answers.

Perhaps would a different selection of the informants have given a different result and show other thoughts and ideas. The spread of knowledge about usability testing was uneven among the informants and it was difficult to adapt the questions to their knowledge. The questions I used for the ones with knowledge and conducted tests did not quite suit as well to the informant with less knowledge and experience. However, it was at the same time interesting and very fruitful to get a different view on the subject, which was the idea of the selection of the informants.
Since only one observation was made it is difficult to cover everything in a usability test and maybe it would also have given a different result. More observations would have given me more to compare with and a wider data collection to analyze. During the work of this bachelor thesis there was only one usability test occasion unfortunately. Nevertheless, this observation covered different steps and it provided useful information. It was interesting to see the whole process from planning of the test to the test itself. In the early stages of the observation I got the feeling that they were not ready for the test but as the preparations went on that feeling went away. The test occasion was also done quite early in the work of this thesis and I had little time to prepare myself on how to perform an observation. It would have been good with more knowledge before observing, but it gave me something to work with during the whole essay.
6. CONCLUSION

This section sums up the study and presents suggestions of change and ends with proposals to further studies.

What are the issues with the practice of usability testing of software products? This study indicates that there are no big problems in the way of performing usability testing of software products. Best practices and important steps is followed, and the developers gain experience through learning by doing. However, this study clearly shows that the teams using usability testing as a verification tool finds usability testing positive and is necessary in development of software products. But since it is up to the teams and developers themselves it is not used by all development teams. The informants thought that usability testing is the most optimal way of validating the products and the testing will improve if they continue to practice and conduct tests.

6.1 SUGGESTIONS OF CHANGE

What can be done to improve these issues? Organization needs to push and show teams and developers that it is important and necessary with usability testing and try to get it as a standard procedure in the software development. It is also important to adapt it as a culture within the organization and strive to work with a user-centered design instead of “develop-centered” design.

Internal education is a way to introduce teams and developers to usability testing and help them get an understanding of the importance of these kinds of tests. Education is also important and helpful to aid guidance on how to start and tips on how to conduct usability tests. Furthermore, it might lower insecurities and worries towards usability testing and thus increase the usage of usability testing amongst teams that currently does not perform tests.

By doing usability tests often and get it as a routine and as a standard in development the performance of the test will improve. It will lower the risk of building the wrong things and lowering costs when unnecessary things can be ruled out. During the tests it is important to provide stable test environments to avoid issues while testing.

6.2 PROPOSALS FOR FURTHER STUDIES

This study was focused on usability testing on one department at Tieto in Skellefteå. It would be interesting to observe and study other Tieto offices and departments, to examine if this is a common phenomenon and perhaps study other IT-companies as well. To find similarities or differences in the practice and organizationally.

Other studies at Tieto in Skellefteå could be to try and introduce a culture that favors usability testing. What would happen if the teams received education and help to start with usability testing? And see if the usage of usability testing would increase within development teams after internal education.
7. REFERENCES

Books


Electronic


Figures

APPENDIX
In the appendix is the interview questions presented, in both Swedish and English.

INTERVIEW QUESTIONS

Övergripande

Vad är din yrkesroll här på Tieto?

Hur länge har du jobbat här?

Har du fått någon utbildning genom Tieto kring användbarhetstester? Eller har du några andra tidigare erfarenheter kring användbarhetstester?

Tycker du att det behövs mer utbildning eller någon form av dokumentation typ lathund eller guide kring användbarhetstestningen?

Utförandet

Utför ni användbarhetstester på allting som ni utvecklar?
- Om inte, varför?
- Om ja, är det nödvändigt och viktigt?

Utför ni alla i teamet användbarhetstester eller är det vissa som brukar ta hand om det?

Brukar du eller ni i ditt team utföra användbarhetstester?
- Om nej, varför gör ni det inte oftare?

Har du utfört användbarhetstester någon gång?

Sker testning mot användare tillräckligt ofta i utvecklingsprocessen enligt dig?
- Om inte, varför sker inte testning oftare?

Skulle du vilja att ni genomförde användbarhetstester oftare?

Är det alltid nödvändigt med användbarhetstester? Eller kan det uteslutas ibland?

Vad för typ av användbarhetstester utför ni? Är det alltid på plats hos användaren eller sker tester på Tieto ibland?

Följer ni några modeller eller metoder, eller andra viktiga steg när ni genomför testerna?
- Om ja, vilka?
- Om nej, varför inte?

Tycker du att den nuvarande testmetoden, (testmetoderna) fungerar bra?

Tycker du att det finns brister i sättet ni testar på?
- Vilka brister?

Hur gör ni med resultatet eller den insamlade data från testerna?
Använda

Använder ni er av communityt som finns för att lära er av varandra?

Hur såg det ut tidigare med testning? Skedde det oftare, mer sällan? Genomfördes dom av utvecklarna?

**Användarna**

Tror du användarna uppskattar tester?

Hur förberedda är användarna inför ett test? Behöver de vara förberedda? Kan de vara mer förberedda? Skulle det kunna påverka resultatet?

Tror du att testerna minskar risken för missnöjda kunder eller användare?
- Om ja, på vilket sätt då?

**Inställning**

Hur är din inställning till användbarhetstester? /Hur ser du på användbarhetstester?

Vad skulle göra så att du eller ni i ditt team blev intresserade av att börja göra mer användbarhetstester?

Är användbarhetstestning en viktig del i utveckling av applikationer, enligt dig?

Känner du att testningen är något som hjälper och förbättrar produkter?

Tycker du att Tieto lägger tillräckligt stor vikt i användbarhetstester?

Hur tror du andra ser på testning? Andra utvecklare på Tieto eller kunden och användarna?

**Förbättringar**

Utifrån dina erfarenheter, är det nödvändigt att förbättra testningen enligt dig?
- Vad skulle kunna förändras eller utvecklas för att förbättra testningen i så fall?

Har du någon idé kring hur man kan utföra användbarhetstesterna på ett annat, mer innovativt sätt?
**General**

What is your profession here at Tieto?

For how long have you worked here?

Have you received any education through Tieto about usability tests? Or do you have any other previous experiences about usability tests?

Do you think it is necessary with more education or some kind of documentation like a guide or quick reference about usability testing?

**Performance**

Do you perform usability tests on everything you develop?
- If not, why?
- If yes, is it necessary and important?

Does everyone in the team perform usability tests or are a few of you usually taking care of it?

Do you or you in your team perform usability tests?
- If no, why is it not done more often?

Have you performed usability sometime?

Does testing against users occur enough in the development process according to you?
- If not, why does testing do not happen more often?

Would you like you to perform usability tests more often?

Is it always necessary with usability tests? Or can it be ruled out sometimes?

What kind of usability tests do you do? Is it always in place with the user or is Tieto doing tests sometimes?

Do you follow any models or methods, or other important steps when you perform the tests?
- If yes, which?
- If no, why not?

Do you think the current test method (test methods) works well?

Do you think there are faults in the way you are testing?
- What faults?

How do you use the result or the collected data from the tests?

Do you use community to learn from each other?
How did it look earlier with testing? Did it happen more often? Were they performed by the developers?

**Users**

Do you think users appreciate tests?

How prepared are users for a test? Do they need to be prepared? Can they be more prepared? Could it affect the result?

Do you think the tests reduce the risk of dissatisfied customers or users?
- If yes, in what way?

**Attitude**

How is your attitude towards usability tests? / How do you look at usability tests?

What would make you or you in your team interested in starting to make more usability tests?

Is usability testing an important part in developing applications, according to you?
- If yes, in what way?

Do you feel that testing is something that helps and improves products?

Do you think Tieto puts sufficient importance to usability tests?

How do you think others are thinking of testing? Other developers at Tieto or the customer and users?

**Improvements**

Based on your experience, is it necessary to improve the testing according to you?
- What could change or develop in order to improve the testing?

Do you have any idea on how to perform usability tests in another, more innovative way?