ABSTRACT

Achieving product and process quality are among the central themes of software engineering, and quality is an important factor in the marketplace. Usability and user experience (UX) are two important quality aspects, particularly for interactive products. To achieve usability means producing products that let users do things in a satisfactory, efficient and effective way. To develop products with good UX, means going beyond usability, in ways that are still not clear to us.

Achieving good usability and UX is hard. This thesis is concerned with organizations working towards these goals. This research has revolved around understanding and improving the processes by which technology is designed and developed, and understanding the demands and expectations users have. It is about how companies can and do develop products with good usability and UX, and what stops them from working towards this as efficiently as they could. The usability and UX challenge has been viewed from the viewpoints of quality, organizations, and institutions, with a focus on participatory design, user-centred design and wicked problems.

This research can be characterised as empirical research performed over a period of seven years, in close cooperation with industrial partners. The research was performed using multiple data collection methods to create constructs and shape theory. The field methods have ranged from being a participant observer, to performing interviews and holding workshops with members of the participating organisations. A case study approach was initially used, but focus soon moved from case study methodology to a closer focus on grounded theory, and finally the focus shifted to constructivist grounded theory.

This thesis contributes to the field of software engineering in several ways. Usability has a long history within software engineering, human computer interaction, and design science, but the different discourses within the fields have meant that communication between the fields was problematic. The research in this thesis has moved between the different fields, contributing to bridging the gap between the areas.

The thesis provides an illustration of how usability work actually takes place in different types of companies, from a developer of operating systems for smartphones, to a global engineering company, which knows that it must find ways of working with, and measuring, usability and user experience. It gives concrete knowledge about the way in which companies can work with usability testing, and how they can provide information to satisfy the information needs of different stakeholders.

This thesis also provides a discussion of the state of UX today, taking up the problems that stop industry making use of the definitions and theories of UX that exist.

Thus, it gives an illustration of the different factors in product design, development and sales, from dealing with organizational factors to satisfying user needs, that all make usability work such a rocky road to navigate.
The Rocky Road: Why Usability Work is so Difficult

Jeff Winter
The Rocky Road:
Why Usability Work is so Difficult

Jeff Winter

Doctoral Dissertation in
Software Engineering

School of Computing
Blekinge Institute of Technology
SWEDEN
Dedicated to my loving, patient, and understanding family.
Achieving product and process quality are among the central themes of software engineering, and quality is an important factor in the marketplace. Usability and user experience (UX) are two important quality aspects, particularly for interactive products. To achieve usability means producing products that let users do things in a satisfactory, efficient and effective way. To develop products with good UX, means going beyond usability, in ways that are still not clear to us.

Achieving good usability and UX is hard. This thesis is concerned with organizations working towards these goals. This research has revolved around understanding and improving the processes by which technology is designed and developed, and understanding the demands and expectations users have. It is about how companies can and do develop products with good usability and UX, and what stops them from working towards this as efficiently as they could. The usability and UX challenge has been viewed from the viewpoints of quality, organizations, and institutions, with a focus on participatory design, user-centred design and wicked problems.

This research can be characterised as empirical research performed over a period of seven years, in close cooperation with industrial partners. The research was performed using multiple data collection methods to create constructs and shape theory. The field methods have ranged from being a participant observer, to performing interviews and holding workshops with members of the participating organisations. A case study approach was initially used, but focus soon moved from case study methodology to a closer focus on grounded theory, and finally the focus shifted to constructivist grounded theory.

The thesis contributes to the field of software engineering in several ways. Usability has a long history within software engineering, human computer interaction, and design science, but the different discourses within the fields have meant that communication between the fields was problematic. The research in this thesis has moved between the different fields, contributing to bridging the gap between the areas.

The thesis provides an illustration of how usability work actually takes place in different types of companies, from a developer of operating systems for smartphones, to a global engineering company, which knows that it must find ways of working with, and measuring, usability and user experience. It gives concrete knowledge about the way in which companies can work with usability testing, and how they can provide information to satisfy the information needs of different stakeholders.

This thesis also provides a discussion of the state of UX today, taking up the problems that stop industry making use of the definitions and theories of UX that exist.

Thus, it gives an illustration of the different factors in product design, development and sales, from dealing with organizational factors to satisfying user needs, that all make usability work such a rocky road to navigate.
Acknowledgements

“In the merry month of May, from my home I started…”

That is the first line of the traditional Irish song “The Rocky Road to Dublin”, which has long been a favourite in pubs and music sessions wherever Celtic music is played. The song has inspired the title of this thesis. It’s also a pretty good description of life.

It is sometimes hard to know when and where journeys begin. I don’t know if this one began in the merry month of May, but “The Rocky Road” symbolises the journey that I have taken to get here, where I am today. In many ways it has seen me following a rocky road. As I have made my way along that road, sometimes singing a song, sometimes humming a dirge, there are many people who have helped and encouraged me. I would like to thank them all.

Particular thanks to Docent Kari Rönkkö and Dr. Jeanette Eriksson, my main supervisor and supervisor, who have patiently kept me on track, guided me past the worst bumps, helped me out of the worst potholes along the road, and lent me a compass when I have stood at the many crossroads along the way.

Thanks to Sara Eriksén, Annelie Ekelin, Pirjo Elovaara, and Yvonne Dittrich, and the others in the e-government gang, for first showing me where the road started. Thanks also to the past and present members of my research groups, first U-ODD and then GSIL. Thanks to Professor Bo Helgeson for helping me at a crucial moment, who got me back on track again when I was stuck and wondering how I was possibly going to explain how I had worked with all this material. Thanks to Bo-Krister Vesterlund and Christina Hansson, for letting me take responsibility for their captain’s bell. Thanks to the 9:07 coffee team, who have answered the call of the bell, and who have brightened up my mornings with talk high and low. Thanks to Jan Björkman, for nice cups of tea, lunch at SG5, and plenty of refreshing discussions over Skype.

I am sure to have missed some worthy people who deserve to be named here. If I have, I hope that you will forgive me. This long journey, bumping down the rocky road, has at times shaken the sense from my head…
List of articles in the thesis

Articles included in the thesis in their entirety. In chronological order, according to publication or submission date:

1. **The Concept of UX Depends on the Intellectual Landscape of HCI Paradigms**
   Submitted to International Journal of Human-Computer Studies (IJHCS), Special Issue on the Interplay between User Experience Evaluation and Software Development.
   K. Rönkkö, J. Winter

2. **Identifying organizational barriers – a case study of usability work when developing software in the automation industry.**
   Under review at Journal of Systems and Software (JSS).
   J. Winter, K. Rönkkö, M. Rissanen

3. **Examining Correlations in Usability Data to Effectivize Usability Testing**
   J. Winter, M. Hinely

4. **SPI success factors within product usability evaluation**
   J. Winter, K. Rönkkö

5. **Satisfying Stakeholders’ Needs – Balancing Agile and Formal Usability Test Results**
   J. Winter, K. Rönkkö

Articles referred to in the thesis as background material: In chronological order, according to publication date:

6. **The Success Factors Powering Industry-Academia Collaboration**

7. **Inside Information 2 – usability and user research: Eight years of research and method development cooperation**
   Technical report, School of computing, Blekinge Institute of Technology, 2009.
   K. Rönkkö, J. Winter, M. Hellman

8. **Reporting User Experience through Usability within the Telecommunications Industry**
   CHASE Workshop, International Conference on Software Engineering (ICSE 2008), Leipzig, Germany. K. Rönkkö, J. Winter, M. Hellman

9. **Developing Quality through Measuring Usability – The UTUM Test Package**
   5th Workshop on Software Quality (WoSQ), ICSE 2007, Minneapolis, MA.
   J. Winter, K. Rönkkö, M. Ahlberg, M. Hinely, M. Hellman
List of published articles not included in thesis

Articles related to, but not included in the thesis:

10. Reporting Usability Metrics Experiences
CHASE Workshop, International Conference on Software Engineering (ICSE 2009), Vancouver, Canada. Jeff Winter, Kari Rönkkö, Mats Hellman

11. Meeting Organisational Needs and Quality Assurance through Balancing Agile & Formal Usability Testing Results
Central and Eastern European Conference on Software Engineering Techniques (CEE-SET 2008), Brno. J. Winter, K. Rönkkö, M. Ahlberg, J. Hotchkiss

Articles not related to the thesis:

12. Adapting to Sweden – Chinese students and their academic experiences at BTH
Lärarlärdom 2012, Karlskrona, Sweden
J. Winter, H. Kyhlbäck

13. Law, Knowledge and Mobility in Local Planning
Jeff Winter

14. What have we learned from the TANGO Arena for Regional Cooperation around E-Government in Southern Sweden?
S. Eriksén, A. Ekelin, P. Elovaara, Y. Dittrich, C. Hansson, J. Winter

15. Educational programs in e-government - An active, practice- and design-oriented network?

16. Agile processes for small public projects and PD outside of office hours
Information Systems Research Seminar in Scandinavia 2004 (IRIS'27), C. Hansson, J. Winter

17. KomInDu – A Small Project About Big Issues
8th Biennial Participatory Design Conference (PDC2004), Toronto, Canada
A. Ekelin, P. Elovaara, S. Eriksén, Y. Dittrich, C. Hansson, S. Landén, A. Larsson, I. Olén, J. Winter

18. Utbildningsprogram i e-government - början till ett aktivt skandinaviskt kompetensnätverk?
S. Eriksén, A. Ekelin, P. Elovaara, C. Hansson, J. Winter

19. What have we learned from the TANGO Arena for Regional Cooperation around E-Government in Southern Sweden?
S. Eriksén, A. Ekelin, P. Elovaara, Y. Dittrich, C. Hansson, J. Winter

# Table of contents

## Contents

### Chapter One

1. Introduction ................................................................. 1
   1.1 The research environment ........................................... 3
   1.2 The research – the big picture .................................... 4
   1.3 Research Challenges ................................................. 6
   1.4 Lenses and Focuses .................................................... 10
   1.5 Research Methodology ............................................... 35
   1.6 Results ........................................................................ 58
   1.7 Contribution to Quality, Organizations and Institutions ....... 65
   1.8 Conclusion .............................................................. 70
   1.9 Future work ............................................................ 70

### Chapter Two

Examining correlations in Usability Data to Effectivise Usability Testing ...... 75
   2.1 Introduction ............................................................... 75
   2.2 Research Questions .................................................... 76
   2.3 Usability and the UTUM Test ........................................ 77
   2.4 Research Method ........................................................ 81
   2.5 Subjects and Context ................................................... 82
   2.6 Validity ........................................................................ 83
   2.7 Data Sets, Possible Correlations and Analysis Steps .......... 84
   2.8 Discussion ............................................................... 89
   Acknowledgements .......................................................... 90

### Chapter Three

The concept of UX depends on the intellectual landscape of HCI paradigms 93
   3.1 Introduction ............................................................... 93
   3.2 Measurement in Software Engineering ............................ 98
   3.3 Definitions of UX in HCI ............................................. 100
   3.4 Paradigms Understood as Generative Metaphors .......... 104
   3.5 Centers and Margins in the three paradigms of HCI ...... 106

:: V ::
Chapter One

Introduction
Chapter One

Introduction

One of the most important questions when designing and developing products, including software and products that include software, is how to achieve product quality. The overall goal of achieving quality has been one of the central themes of software engineering. Given the fierce competition that exists in the marketplace, the margin between success and failure in the marketplace is small. Quality, particularly from the point of view of end-users, and the quality aspects that they experience, is vital to sales, and thereby vital to success.

The overriding theme of the research leading to this thesis has been how to achieve quality when developing a software product that is marketed to a diverse group of end users. It has been concerned with how companies achieve product quality, particularly through software engineering processes. The specific aspect of quality studied is product usability, which is also related to user experience (UX). These two quality aspects are particularly important for end-users, and much effort is being made to ensure that they can be achieved.

The research presented here is an account of empirical research, based on real-world industrial problems. The work has taken place over a period of seven years, in close collaboration with two industrial partners, UIQ Technology AB and The ABB Group. Thus, this thesis gives an illustration of everyday work and accounts of this in different industrial contexts. The research into usability and UX turned out to be a rocky road, passing through the areas of participatory design, user-centred design, institutionalism and wicked problems. The research began by looking at the problems associated with measuring usability and introducing usability metrics in organisations, and progressed to looking at some of the aspects of the developing field of UX. This led to a study of the needs of different stakeholders both within and outside organizations. Finally the study turned to the challenges that occur within complex organisations, from an institutional point of view, and why it is so difficult to apply usability knowledge in these organisations.

Usability has long been, and UX is emerging as, an important subject within software engineering, but historically they both have strong connections to the fields of Human Computer Interaction (HCI) and design science research. The research that has led to this thesis is based on a wide spectrum of fields, thereby contributing to closing an unfortunate gap that was found between the software engineering community and the HCI and design communities (Bass 2003; Harning and Vanderdonckt 2003; Kazman and Bass 2003; Kazman, Bass et al. 2003; John, Bass et al. 2004). To contribute to bridging this gap, this research has moved between these fields, showing how theories and work taken from the different fields can be combined to enrich several fields. For example, in order to gain a better understanding of the roles played by institutionalized organizations, the research
presented here has borrowed knowledge from the fields of management studies and information systems. This knowledge is thereby brought to the field of software engineering.

This research has led to a number of results, from theoretical knowledge to knowledge that can be used in practice. One part of the research involved the design and implementation of a usability testing framework called UTUM (UIQ Technology Usability Metrics), primarily used to measure the usability of mobile phones. An analysis of the data collected during UTUM testing showed how it was possible to simplify the discovery of usability problems in a user interface. On an organizational level, the research showed the existence of different types of stakeholders in the development organization, and the ways in which results could be presented to best satisfy the needs of these stakeholders. The research shows similarities between usability testing and other process improvement methods, and showed how the results could enrich the area of process improvement. It shows how UX is an area with many definitions. It discusses how these definitions are grounded, and why industry finds it hard to make use of UX theories and definitions. It also shows what happens when organizations become institutions, how this can lead to “wicked problems”, and how this affects the way in which usability and UX work get to be done.

<table>
<thead>
<tr>
<th>CASE</th>
<th>ISSUES</th>
<th>LENSES</th>
<th>FOCUSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>UIQ</td>
<td>COMPLEX PRODUCTS FOR A MASS MARKET</td>
<td>QUALITY</td>
<td>Usability and User Experience</td>
</tr>
<tr>
<td></td>
<td>GETTING THE RIGHT INFORMATION IN THE RIGHT PRESENTATION</td>
<td>ORGANISATIONS</td>
<td>Participatory Design</td>
</tr>
<tr>
<td></td>
<td>AGILITY - PLAN-DRIVEN THE RIGHT INFORMATION IN THE RIGHT FORMAT</td>
<td></td>
<td>User-Centred Design</td>
</tr>
<tr>
<td>ABB</td>
<td>INTENTIONS VS. POSSIBILITIES AS ORGANISATIONS BECOME INSTITUTIONS</td>
<td>INSTITUTIONS</td>
<td>Wicked Problems</td>
</tr>
</tbody>
</table>

Figure 1.1 The metaphors of cases, issues, lenses and focuses in this thesis

Figure 1.1 above illustrates the overall structure of the research that has been performed. The diagram shows: the cases that have been part of the study, or more specifically, the

:: 2 ::
companies involved; the high-level issues that have been studied; the lenses that have been used to examine the thematic areas; particular focuses that have been used when analysing the issues. The issues, and the details of how they were approached, are detailed in the following chapters of this thesis. The lenses and focuses are discussed in section 1.4 Lenses and Focuses.

The terminology used above, which refers to “lenses” and “focuses”, is illustrated in figure 1.2. These terms refer to the specific areas that the research has led to, and the way in which they have been studied. To explain the metaphor: when performing the research, I have worn different pairs of “analytical spectacles” in different phases of the research. The spectacle frames represent the area of software engineering, which has framed the whole research effort. The issues that have arisen have led me to look at the area through a particular lens. This lens has been bifocal, so that, although I have studied the overall area through a particular lens, I have also had a particular focus when looking more closely at the details of the study.

![Figure 1.2 An example of the analytical spectacles](image)

1.1 The research environment

The research began within a software engineering research environment at Blekinge Institute of Technology (BTH) (in Swedish, Blekinge Tekniska Högskola). The research environment, called *Blekinge Engineering Software Quality* (BESQ), was launched in 2002. BESQ was initially run as a research programme, and the objective was to create a strong research environment within software engineering, which would continue beyond the initial six year period (BESQ 2004). As the name indicates, the focus within the programme was developing software quality, which has remained a focus throughout the research. The starting point was the fact that all software projects constitute a trade-off between different qualities, and the vision of the project was the achievement of “quality-balanced software”. The overall objective of the research within the project was to provide methods, techniques and tools to engineer the different qualities and the trade-off between them through the software life cycle from requirements to acceptance testing via design and implementation. As a result of this programme, the BESQ Research Center (BESQ RC) was established in 2007 (BESQ 2007). The mission of BESQ RC was to provide support
for quality-balanced software to its partners. Individual projects (usually in close industrial collaboration) were aimed at contributing to the overall mission of the programme.

Thus, the research within BESQ had as its overall goal the achievement of quality in software engineering processes and products. A further focus was that it should be aimed at empirical research together with an industrial partner. Within BESQ I was recruited to the research group User-Oriented Design and Development (U-ODD 2011) headed by Dr Rönkkö, which represented the usability view in the BESQ environment. The goal of this particular research sub-project was to discover how usability testing can be used to improve product quality. The research was performed together with a company working in a strongly market driven field, designing, developing and marketing a mass-market product. The main purpose of the research was to contribute to, and study, the development and use of a method and a tool for measuring usability, quality of use, and user experience. The format and contents of the test were driven by the needs of multiple stakeholders, in the context of market driven requirements (BESQ 2008).

A further phase of the research took place outside BESQ, when research cooperation began together with ABB Corporate Research. Here, focus was still on usability and UX, but the context of the research, when compared to the initial phase, was with a very different type of industrial partner, with a very different type of product, in a very different market environment.

The different phases are described in more detail in the next section, where a more detailed overview is provided of the stages and contexts of the research presented in this thesis.

### 1.2 The research – the big picture

This section gives a more detailed description of the research that has taken place over this period of seven years. There are in essence two projects, where research was carried out at two companies. The research in the first project was based on a number of case-studies that built upon one another, using the results of previous case studies to inform and guide new case studies. The second project is based on the work performed in the first project, and transfers some of the knowledge gained in the first project to a different type of organisation and setting.

The initial research was performed in close cooperation with UIQ Technology AB (UIQ). This research was, in turn, based on earlier research cooperation between UIQ and BTH. UIQ, which was established in 1999, and closed operations in 2009, was a company that developed operating systems (the UIQ Platform) for mobile smart phones. It was one of the early actors within this area, and they were responsible for a number of innovative telephone models designed and produced by Sony Ericsson, Motorola, Benq, Nokia and Arima.

UIQ produced technically advanced and complex mobile phones (smartphones) for a mass consumer market. These phones were aimed towards the mass market, but many of the features were heavily niched towards early adaptors of advanced mobile phone technology. Right from the beginning, the company understood the importance of usability for sales of their handsets, and the collaborative research that was performed involved designing and
developing a metrics based usability test package that eventually became a standard within the company, and has also been used in other environments. The test package was given the name UTUM (UIQ Technology Usability Metrics). The development of this test package, which was developed through a number of releases, was the first concrete result of the research process (see e.g. Winter, Rönkkö et al. 2007; Rönkkö, Winter et al. 2009). Eventually, focus moved beyond usability and was extended towards UX, which was becoming more important as consumer product technology developed, and new products appeared on the market (see e.g. Rönkkö, Winter et al. 2008). Chapter three takes a broad view of the way in which UX definitions are created, and the problems associated with understanding and using them.

After UTUM was developed and introduced, practical requirements led to a further stage in the research process. During the initial stage of the research, the importance of organizational factors when analysing and presenting the results of the usability testing became apparent. Continued research showed that there were a number of different types of stakeholders within the organization, and that these different stakeholders had different information needs, regarding usability knowledge. It became clear that there was a need to tailor the collection and presentation of the results of the usability testing to the different groups. It was found that they were on two sides of a spectrum, regarding the level of detail included in the results, and the time scale that the results concerned.

Further analysis of the results led to deeper studies of the importance of processes when using the results of usability testing. The results of the usability testing were used to create a number of Key Performance Indicators (KPIs) for usability. To use these KPIs in a relevant fashion, it was found important to look at the processes that were in operation, and which affected the formulation and use of the KPIs. This led to a closer look at parallels between the research results in the field of usability, and the area of Software Process Improvement (SPI). Chapters two, four and five are based on the work performed at UIQ.

As a result of turbulence in the mobile phone industry, UIQ ceased operations in 2009 (for some background information, see e.g. Gigacom 2008; The Register 2008; OS News 2009). However, the research continued with a new research partner when usability and UX researchers from the ABB Group (ABB Group 2012), working within the ABB Corporate Research Center – Sweden (ABB CR 2012) made contact. The researchers at ABB had read the results of the usability research work with UIQ, and were interested in cooperating, since they were beginning research on how to introduce usability KPIs in their organization.

However, whereas UIQ was a young company, producing consumer-oriented electronics, ABB is a very different type of organisation, with a long background in the engineering tradition. The specific business unit that was involved in the study specialised in producing industrial robots and the software associated with them. The study at ABB had its roots in an ambition to work with usability and UX, and usability and UX metrics, but although this division of the company was perhaps one of the parts of the company that had progressed most in their usability work, the usability and UX maturity here, and within the organisation as a whole, was not at the same level as it was at UIQ.
Thus, the environment for the final stage of the research, at ABB, was within a company with a very different structure, background and field of operations, and with a different organisational history. The cooperation led to the performance of an interview-based study examining the situation within one of the business units, regarding the possibility of formulating and operationalizing usability KPIs. Since ABB is a radically different type of organization compared to UIQ, with a strong focus on traditional engineering values, this study gained a strong focus on investigating and understanding organizational factors. This fact led to a closer look at theories regarding organisations and the way in which they become institutionalized, and the way in which this affects usability work. The research showed the existence of a wicked problem that deeply affects the way in which usability is discussed and worked with in the organisation. Chapter six is based on the research at ABB.

1.3 Research Challenges
The research involves a total of five research challenges, associated with the three lenses used in the thesis (see Figure 1.3).

<table>
<thead>
<tr>
<th>LENSES</th>
<th>TOOLS</th>
<th>CHALLENGES</th>
<th>CHAPTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUALITY</td>
<td></td>
<td>C1. To find ways improve the analysis and use of usability test results</td>
<td>2</td>
</tr>
<tr>
<td>USABILITY</td>
<td>UTUM</td>
<td>C2. To find ways to make use of today’s scattered concepts and definitions of UX</td>
<td>3</td>
</tr>
<tr>
<td>USER EXPERIENCE</td>
<td></td>
<td>C3. To satisfy the information needs of different stakeholders in organizations</td>
<td>4</td>
</tr>
<tr>
<td>ORGANISATIONS</td>
<td></td>
<td>C4. To find ways to combine usability testing processes with other organizational processes</td>
<td>5</td>
</tr>
<tr>
<td>STAKEHOLDERS</td>
<td></td>
<td>C5. To map the organizational factors that affect how usability work is done</td>
<td>6</td>
</tr>
<tr>
<td>PROCESSES FOR QUALITY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INSTITUTIONS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1.3 Lenses, tools and challenges in the thesis work, connected to publications*
Chapter One
Introduction

The first research challenge (C1) deals with finding appropriate ways of analysing and using the results of usability testing.

The second research challenge (C2) involves dealing with the problems associated with understanding and using the definitions and concepts of UX. It involves examining and discussing the importance of definitions and the way in which UX is defined and treated in the research community, and the problems associated with UX measurement, particularly for the field of software engineering.

The third research challenge (C3) deals with finding ways to use the results of usability and UX testing to satisfy the different information needs of different stakeholders in organisations.

The fourth research challenge (C4) deals with finding ways in which research in usability and UX can contribute to other process improvement methods, and can in turn be improved by studying other process improvement methods.

The fifth and final research challenge (C5) concerns understanding the way in which organizational factors affect how usability work is discussed and performed in a large organisation, and the way in which the history and structure of the organization affect what gets done within the organization.

The order of the following chapters reflects the number of the challenge that is addressed, not the publication date of the article that is included in the chapter. In some cases, the research has returned to challenges, working with ideas and research material that was collected earlier in the process, producing work based on these ideas or material. The challenges are related to the lenses shown in Figure 1.1. In this section, the articles are numbered according to the list of articles that are included in this thesis, found on page iii

1.3.1 Lens: Quality
Chapters two and three in this thesis deal with usability and UX, and take up the first two research challenges.

**Challenge 1 (C1).**
The first research challenge deals with finding appropriate ways of analysing and using the results of usability testing.

This is dealt with in chapter two: “Examining Correlations in Usability Data to Effectivize Usability Testing” (Article 3).

The research in this chapter deals firstly with examining the correlation between measurements of effectiveness, efficiency and satisfaction, which are all different aspects of usability. Furthermore, it examines whether a statistical analysis of task-completion time allows the discovery of problematic use cases. This was done in an attempt to discover whether there are correlations that allow the discovery of usability problems on the basis of a simple metric, which could easily be measured without requiring the presence of a usability tester.
The first part of the research is based on the idea that there may be a sufficiently strong correlation between the three factors of usability that measuring one of them would give a reliable indication of the usability of a mobile phone. The second part of the research is based on the hypothesis that there is an expected distribution of completion times for a given use case and that deviations from goodness of fit indicate user problems.

**Challenge 2 (C2).**

The second research challenge involves dealing with the problems associated with understanding and using the definitions and concepts of UX. It involves examining and discussing the importance of definitions and the way in which UX is defined and treated in the research community, and the problems associated with UX measurement, particularly for the field of software engineering.

This challenge is dealt with in chapter three: “The Concept of UX depends on the intellectual landscape of HCI Paradigms” (Article 1).

The research in this chapter deals with solving the difficulty that industry has in making use of the knowledge and theories of UX that are produced in academia and industry, despite a strong desire to understand and make use of UX to compete. This article contributes by providing one explanation for this difficulty, based on the way in which definitions are grounded. It presents the idea that it is necessary to discuss the origins of UX as a concept in order to characterize UX as a concept, and suggests that it is necessary to look at the origins of the theoretical foundation that UX concepts are built upon. It also discusses the challenge of using measures of UX in software engineering, and the way in which the current state of the field affects these possibilities.

1.3.2 Lens: Organizations

**Challenge 3 (C3).**

The third research challenge involves dealing with finding ways to use the results of usability and UX testing to satisfy the different information needs of different stakeholders in organisations.

This challenge is dealt with in chapter four: "Satisfying Stakeholders’ Needs – Balancing Agile and Formal Usability Test Results" (Article 5)

The research in this chapter deals with the best way to capture usability data through usability testing, and the best ways to represent that data. The research involved finding factors that would allow the identification of the different types of stakeholders present in the organization who had an interest in usability data. Furthermore, it dealt with discovering the types of testing methods and result presentations that were most appropriate for these different groups, which are involved at different levels and in different stages of the product design and development process.

The research was also aimed at studying the importance of the test leader when presenting the results of the testing, to see if test results could be presented in a meaningful way without the test leader being present, or whether presentations could be sufficient in themselves, without the test leader’s presence to explain the results.

:: 8 ::
Challenge 4 (C4).

The fourth research challenge deals with finding ways in which research in usability and UX can contribute to other process improvement methods, and can in turn be improved by studying other process improvement methods.

Chapter five contains: "SPI success factors within product usability evaluation" (Article 4)

The research in this chapter deals with finding out how the results of usability testing can be used to improve usability processes within organizations. Furthermore, it compares usability and SPI, to discover if there are similarities. It asks how the product focus of usability and the process focus of SPI can be combined, and what would be the effect if these efforts were related to usability process improvement efforts within organizations, and how organizations can work more effectively with these processes.

1.3.3 Lens: Institutions

Challenge 5 (C5).

The fifth and final research challenge concerns understanding the way in which organizational factors affect how usability work is discussed and performed in a large organisation, and the way in which the history and structure of the organization affect what gets done within the organization.

Chapter six is the final chapter in the thesis: “Identifying organizational barriers– a case study of usability work when developing software in the automation industry” (Article 2)

The research in this chapter deals with issues that occur when organisations become institutions, and what can be done when faced with a situation that can be characterised as a wicked problem.

1.3.4 An overview of the rest of the chapter

The remainder of this chapter is structured as follows. Firstly, details are given about the contents of the research (1.4 Lenses and focuses), divided into a number of lenses and focuses, according to the structure shown above in Figure 1.1.

- Quality
  - Usability and UX

- Organisations
  - Participatory Design
  - User-Centred Design

- Institutions
  - Wicked Problems

Following the overview of the different lenses and areas of focus, comes a presentation of the methodology that has been used (1.5 Research Methodology), together with an overview of how it has developed over time. Following this, the results from the articles are presented (1.6 Results), and then the contributions that this research makes to the field are given, by discussing the results in relation to the areas that are included (1.6 Contribution to
Quality, Organizations and Institutions). This section of the thesis is completed with a short conclusion (1.8 Conclusion) and ideas for future work (1.9 Future Work).

1.4 Lenses and Focuses

In order to deal with the multiple perspectives that were involved in this work, use has been made of theories from a number of areas, including software engineering, information systems, and management and organizational theories. In the early phases of the research, a case-study approach was used. When using a case-study approach, a key factor is considering a broad range of literature (Eisenhardt 1989). Furthermore, this use of a broad range of literature is important since a grounded theory (GT) approach was also used. According to Urquhart, Lehmann et al. (2010) grounded theory must be set in the context of other theories in the field. They state that one of the advantages of GT for researchers is the obligation to engage with theories from other disciplines. Charmaz (2009), also states that a GT researcher, upon encountering a surprising finding, should consider all theoretical ideas that could account for it, gather more data to put the ideas to the test, and then adopt the most plausible explanation. In addition to the ambition to bridge the gap between different discourses, this explains the use of material from varied research fields.

Returning to the metaphor of viewing the research through different spectacles, a specific lens is inserted in the spectacle frames, to begin to look at the area of quality. Having taken a general view of the area of quality, the lower part of these quality lenses is used, to focus more closely on usability and UX.

1.4.1 Lens: Quality

Thus, the first lenses inserted in the spectacle frames allow a closer look at issues of quality. The overriding theme of the research has been how to achieve quality when developing a software product that is marketed to a diverse group of end users. The overall goal of achieving quality has been one of the central themes of software engineering, and as early as 1996, Osterweil (1996) stated that product quality was becoming the dominant success criterion in the software industry, and that the challenge for research is to provide the industry with the means to deploy quality software, allowing companies to compete effectively. The research in this thesis contributes to addressing this challenge. Changes in user demands are forcing companies to change their view of what constitutes quality. Changes in the way interactive technology is developed and marketed, exemplified by what many would call “the iPhone revolution”, have affected the view of what constitutes the quality of interactive products. This has been particularly relevant for this research, as it began together with a company that produced software for smartphones. This research has focused on the role of usability work in achieving software and product quality.

From a software engineering point of view, it is important to find methods to ensure the quality and utility of products, and there must be a strategy for producing quality software. According to Osterweil (1996), quality is multi-dimensional, and impossible to show through one simple measure, so research should focus on identifying various dimensions of quality and measures appropriate for it, and a more effective collaboration between practitioners and researchers would be of great value. This research contributes to these
efforts by presenting work that has been performed in close cooperation with practitioners over an extended period, aimed at testing and measuring usability and user experience as quality attributes.

Osterweil (1996) maintains that testing as a form of quality assurance is important, and this view was supported by Harrold (2000) in her roadmap for testing. However, most projects involve a number of stakeholders, and each of these has a different definition of quality, and different acceptable limits for it. According to Preece, Rogers et al. (2011), differing views of what constitutes quality lead to disputes when quality attributes are discussed. This points to the need for developing testing procedures that provide different stakeholders with information adapted to their particular needs, which has been one focus area of this research.

Thus, to ensure quality, there must first of all be an understanding of what is meant by quality, and it is clear that different perspectives influence what the term quality means (Pfleeger and Atlee 2006). The idea of what constitutes quality differs, depending on the standpoint of the viewer. Preece, Rogers et al. (2011) state that when we use an artefact, we have a view of its quality. Whether or not this view of quality is expressed formally or informally is unimportant. What is important is that the view exists, and that we use it to evaluate the products that we use. Thus, there are many views of what constitutes quality.

Pfleeger and Atlee, citing (Kitchenham and Pfleeger 1996), discuss the fact that context is important in deciding which aspects of quality are important, and they look at three different aspects of quality: the quality of the product; the quality of the process; and the quality of the product in the context of the business process where the product will be used (Pfleeger and Atlee 2006). As will be seen in the contents of this thesis, this research has emphasised product quality and process quality, and has focused less on quality in the context of the business environment, which according to the discussion in an article by Pfleeger and Atlee (2006) has so far been mostly concerned with return on investment. Focus in this research has been on end-users and their use of software and products, and has not been placed on the ways in which investments in technology repay the investments made.

As previously stated, the importance of different characteristics depends on the stakeholder, and concerning product quality, Pfleeger and Atlee (2006) discuss the way in which quality and functionality are intertwined. They find that users often judge software to be of high quality if it does what the user wants it to do in a way that is easy to learn and easy to use. However, if software is difficult to learn or use, it can still be judged as having high quality if the functionality is worth the trouble. This is discussed further in the section on user experience (UX), when discussing the way in which the attractive features that a product offers can actually outweigh the lack of what might otherwise be seen as essential functionality.

The starting point for this research was on the measurability of usability and UX as quality attributes. Measurement is an important tool, and software engineering has a strong focus on what is measurable. Producing better products, by way of improved processes, resources and methods, is a driving force in software engineering research, and software
measurement is seen as a key aspect of good software management (Pfleeger and Atlee 2006). The type of measurement varies according to what is being measured, why it is being measured, who is interested in the measurement, and on the environment where the measurement is performed, but software measurement is a way of applying measures to software engineering objects to achieve predefined goals (Basili, Caldiera et al. 1994). However, even though measurement is seen as vital for improving software processes, the complexity and uncertainty involved mean that companies have found it difficult to implement metrics programmes as part of their process improvement efforts (Iversen and Kautz 2001). These difficulties were particularly apparent in the latter phase of the research, together with ABB.

When looking at the range of measurements that are detailed by Pfleeger and Atlee (2006), it can be seen that the majority of software engineering measurement models that are discussed are internal measures, such as general sets of measurement for object-oriented systems, measures of coherence, cohesion and coupling for object-oriented systems, internal attributes that affect maintenance characteristics. Some external measurements, such as those regarding maintainability, are also discussed. However these are mainly discussed in terms of the impact on internal processes, rather than as a direct attempt to increase product quality from a user point of view. As a roadmap for future work, Pfleeger and Atlee (2006, p. 163) discuss Wasserman’s Steps to Maturity, and they state that in the future it will become increasingly important to go to “other factors” beyond faults and failures, and to measure quality on a broader framework, including customer satisfaction and business needs. The research in this thesis, with its focus on organisations and institutions, usability and UX, participatory design and user-centred design, and wicked problems, has focused on the “other factors” mentioned above.

Focus: Usability and User Experience

The quality lens (Figure 1.4) used in the previous section is now used more specifically to focus on the fields of usability and user experience, which have been central areas for this research.

![Figure 1.4 The Quality lens, focusing on Usability and UX](image)

The most commonly accepted definition of usability, which was used as the foundation of this research, is that usability is 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. In this definition, \textit{effectiveness} is related to the accuracy and completeness with which users achieve specified goals. \textit{Efficiency} concerns the resources expended in relation to the accuracy and completeness with which users achieve goals. In the context of usability, efficiency is related to ‘productivity’ rather than to its meaning in the context of software efficiency. \textit{Satisfaction} concerns freedom from discomfort, and positive attitudes towards the use of the product (ISO 9241-11 1998). This ISO standard, although it has been in existence many years, is still influential in the way in which usability is discussed and applied.

According to Dumas and Redish (1999), good usability means that people who use a product can do so quickly and easily to accomplish their own tasks. They state that usability rests on four points, which are: focusing on the users; the fact that people use products to be productive; that users try to accomplish tasks, and; that users decide themselves if a product is easy to use. They also state that usability must be built in to a product from the beginning, as it is affected by all design and development decisions. In this research, it was found that these factors pointed out by Dumas and Redish (1999) have been a concise summary of many of the factors that have been found to be important, with a focus on users and their tasks, on the importance of discovering the actual users’ view of usability, and the importance of using this to work with usability at all stages of the design and development processes.

As discussed under the subject of quality, testing and measurement are important, and a primary focus of this research has been on usability testing and measurement, and even on the possibility of testing and measuring user experience.

In (Pfleeger and Atlee 2006), usability testing is seen as being a form of performance testing. It is referred to as human factors testing, and it is said to deal with investigating requirements dealing with the user-interface, examining screens, messages, report formats and other aspects related to ease of use. Operator and user procedures are also checked to see whether they conform to ease of use requirements. Usability work has typically been concerned with measuring efficiency, effectiveness and satisfaction, in accordance with the ISO standards, e.g. (ISO 9241-11 1998). There is a general agreement from standards boards on these dimensions, even though there is less agreement on which metrics should be used to quantify the dimensions (Sauro and Kindlund 2005). An important factor in this research has been to develop metrics that can be used to measure usability as a quality attribute, to generate appropriate measures, and to use these measurements to improve products and processes.

The aim of usability testing has been the measurement of aspects of usability, and by defining the abstract concept “usability” in measurable components, the ambition is to achieve an engineering discipline where usability is “systematically approached, improved and evaluated (possibly measured)” (Nielsen 1993). Usability engineering is an iterative design and development process which begins with identifying users, analyzing tasks, and setting usability specifications, and continues with developing and testing prototypes, through iterative development and test cycles (Dumas and Redish 1999). One important characteristic of usability engineering is writing down verifiable usability criteria, which
involves specifying and documenting quantifiable measures of product performance, to be used when assessing a product (Preece, Rogers et al. 2011). This was one of the focus areas in the usability test that was developed together with UIQ, which is discussed in more detail in chapters two, four and five in this thesis.

However, although there are clear areas of agreement between the areas of software engineering and usability, with their common focus on measurement and metrics, there have traditionally been gaps between the field of software engineering and, for example, Human Computer Interaction (HCI), where much of the work within usability has been performed. These gaps have been found to be difficult to bridge, and this has been discussed in a number of publications (see e.g. Bass 2003; Gulliksen 2003; Harning and Vanderdonckt 2003; Kazman and Bass 2003; Kazman, Bass et al. 2003; John, Bass et al. 2004). The fact that there do not appear to be recent publications on this subject may either mean that the gaps have been bridged, or, as my research experience within the software engineering field leads me to believe, that some gaps remain to be closed.

Although there are probably many different reasons for this, it has been suggested that one of the reasons for difficulties working with usability in the software engineering field is the fact that usability is an area which has traditionally been a moving object, and the current meaning of it in the research discourse does not necessarily correspond to the current quality models that are in use in the software engineering community, making it difficult to match the two areas (Sjoberg, Dyba et al. 2007). This pace of change may have slowed in the area of usability, but given the state within the area of user experience, where change is still the norm, and a multitude of definitions exist, these difficulties are still in existence.

Focus is now changed, to look at the area of user experience (UX) and its relationship to usability. In recent years, discussions of UX have become more frequent in the research community, and UX is now an important topic in HCI, in contrast to other areas such as software engineering, where the topic is still relatively untouched. Early in this research, based on the research cooperation with UIQ, and the developments and needs that were seen there, it was clear that UX was an important and interesting subject for research. In 2006, UX was described as an expanding area of research and practice, but it was regarded as being immature, both theoretically and methodologically, with no definition that could inform the research community how to design for and evaluate it (Lai-Chong Law, Hvannberg et al. 2006). This lack of a coherent definition of UX, which has been pointed out a number of times over the years (see e.g. Lai-Chong Law, Hvannberg et al. 2006; CHI2008 2007; COST294 2008; UXEM 2008), has been seen as a problem within the field. This issue is treated in chapter 3, which presents present an analysis of the situation regarding the presence of so many different definitions of UX, and how these definitions may be viewed in order to make use of them in practice.

There are obvious connections between usability and UX, but it has long been unclear exactly how UX differs from the traditional usability perspective (UXEM 2008). This is a fact that still appears to be true today. There are of course differences that have been pointed out, but it is also clear that for UX to gain acceptance as a field in its own right, it
must differentiate itself from usability, and add something new to the traditional view of interactive product quality (Hassenzahl, Lai-Chong Law et al. 2006).

Traditionally, usability has been equated with task related quality, based on efficiency, effectiveness and satisfaction, where good product usability means that a user performs tasks in intended ways, in a way that does not lead to disturbance or annoyance (ISO 9241-11 1998). This focus can be illustrated by relating it to the Kano model, which is a tool for gaining an understanding of customer satisfaction (see e.g. CQM 1993; Sauerwein, Bailom et al. 1996). The factors that have been seen as central to usability can be related to the “must be” requirements in the Kano model (see figure 1.5) which are seen as implied, self-evident, non-expressed, and obvious. These features are taken for granted, and although improving them will not increase customer satisfaction, a lack of them will cause dissatisfaction. Usability factors can also include the “one-dimensional” requirements in the Kano model, which are usually demanded by customers, and customer satisfaction is proportional to the degree in which these requirements are fulfilled.

Usability has a long history of concentrating on these factors that lead to dissatisfaction, but do not necessarily lead to satisfaction, and by doing so has made much progress towards satisfying industrial needs. The focus given to these areas has formed the way in which usability work has been performed. By concentrating on barriers, frustration and failures, usability testing (in much the same way as other software testing) has concentrated on the negative, and identifying and removing the negative is and will remain an important part of product design and verification. In general, usability work, via observation and surveys, has become successful at identifying “hygiene factors” (Tietjen and Myers 1998).
However, when compared to usability, which is characterised as being pragmatic, objective and negative, UX is seen as being holistic, subjective and positive, (Hassenzahl, Lai-Chong Law et al. 2006). Examples of the influential UX concepts that have shaped the understanding of UX include pleasure (Jordan 2002), joy (Hassenzahl, Beu et al. 2001) and hedonic, a widely used term in the UX context, which also means related to, or characterized by, pleasure (Hassenzahl 2001).

The positive is a further dimension of quality, and high levels of satisfaction result from the presence of positive qualities, rather than the absence of negative qualities. The focus given by usability to the negative is not seen as unimportant, but emphasis is given to the fact that the positive is not simply the absence of the negative. This concentration on these types of aspects can also be related to the Kano model, where they can be equated with the “Attractive” requirements, which delight the customers if they are present (see figure 1.5). These requirements are not expressed by the customers; in fact they may not even be aware of them. Therefore, failing to fulfil them does not lead to dissatisfaction. However, fulfilling them can lead to more than proportional satisfaction.

Lai-Chong Lay, Hvannberg et al. (2006) discuss the fact that both usability and UX can concentrate on efficiency, effectiveness and satisfaction. However, they state that UX attempts to reduce satisfaction to elements such as fun, pride, pleasure and joy, and attempts to understand, define and quantify these elements. The holistic approach means to find a balance between pragmatic aspects and hedonic aspects – non-task related aspects of product use and possession, such as beauty, challenge and joy. This means that UX is subjective, taking an interest in how people experience the products they use. This focus is important, because the subjective experiences of users colour their impressions of owning and using a product, thereby affecting their future behaviour and buying patterns. Furthermore, these experiences affect how people describe to others their product-owning experiences. This is related to what sells well in the marketplace, and has become particularly important for market driven product development, where the experiences of others, spread via a multitude of channels, affect what gets to be sold.

When looking at the description of UX above, given the broad focus on what can be seen as imprecise concepts, and the relatively short time that the concept has been in use, it is perhaps easy to understand the difficulty of reaching an agreement on what exactly UX is and how it should be measured.

In connection with the differences between subjectivity and objectivity, Bevan (2009), addresses the differences between usability and UX methods. He points out that the concept of UX broadens both the range of human responses to be measured and the circumstances in which measurement occur, and sees this as being tricky. This is supported by the fact that researchers have argued that there are substantial difference between objective task-related measures of usability and subjective measurements of users feelings related to UX (Roto, Obrist et al. 2009). Bevan (2009) maintained that finding appropriate measures for UX would be simplified if the different perspectives on UX were identified and distinguished.
As mentioned at the beginning of this section on UX, there are a multitude of definitions of UX, which was noted by Bevan (2008), who stated that the interpretations of UX were much more diverse than those of usability. This situation that does not appear to have changed since he wrote this, and actually appears to have become more prevalent (for more details regarding this, see chapter three in this thesis, where various definitions of UX that have been published are discussed).

Context is one of the factors that the research in this thesis have shown to be of great importance, and the results of this research into usability and UX in the context of industrial development have led to an understanding of the role of different stakeholders, not least in the context the industrial organisation. The importance of context is also supported by the view of Avgerou (2001), who, when discussing research and practice, states that it is of crucial importance to associate technology innovation with the context where it is embedded. Innovations should be considered in relation to socio-organizational change, analysis should consider not only local but also national and international context, and the analysis should consider both the technical/rational decisions and actions and the social and cognitive forces of the process.

The next part of this introduction begins by looking at some of the theories related to organisations and organizational change, before providing an overview of some theories within the fields of participatory design, and user-centred Design, which means that new lenses must be inserted in the spectacles, beginning by taking a general view of theories concerning organizations.

1.4.2 Lens: Organisations

An issue that has dominated management and organizational thinking for more than two decades has been that of change and the need for it, and an increasing proportion of interventions designed to implement change have involved the introduction of Information and Communication Technology (ICT) (Barrett, Grant et al. 2006). The consequences of this introduction often involve significant, widespread and unanticipated change, and there are increasingly close interconnections between ICTs and institutional change (ibid.). The impact of IT on jobs, skills, authority and the managerial labour process is conditioned by the complex interplay of the technology, the social and political conditions of its implementation, and its interaction with other change initiatives (Harley, Wright et al. 2006).

Orlikowski discusses how “[t]he culture of the workplace, managerial ideology, and existing bases of expertise and power significantly influence what technologies are deployed, how they are understood, and in which ways they are used” (Orlikowski 1992, pp. 421-422). Here, Orlikowski points out that there is little agreement on the definition and measurement of technology, and no compelling evidence on the role of technology in organizational affairs. The divergent definitions and opposing perspectives of technological research have limited our understanding of how technology interacts with organizations. According to Arias, Eden et al. (2000), the major changes that have occurred in HCI have been at the levels of design, systems, technology and media, where actions and changes can take months or even years or decades. When working at these levels, Arias, Eden et al.
Chapter One
Introduction

(2000) state that the relevant theories would probably be grounded in social and organizational themes.

This points to the importance of considering the role that organizations play in the processes of developing technology and technological systems, and it is important to study the organizational context where this work takes place. The research in this thesis, with its focus on practice at many levels of the organization, is an example of this kind of research. According to Avgerou (2001), much research effort has been done focusing on developing general theories for the implementation of ICT, but less effort has been put into systematically considering variations of the organizational or the broader context where the ICT is embedded. There is a danger in ignoring context, since it entails the risk of misleading and frustrating efforts to make sense of and appropriate technology, and the significance of considering context cannot be over-emphasized (Avgerou 2001). Having performed research in two very different types of organization, it was apparent how the context and the organizational culture affected the way in which, in this research, usability and user experience work was discussed and performed.

Barrett, Grant et al.(2006) discuss the way in which perspectives on organizational change have shifted from a mechanistic description of the organization, and they present change processes as contingent, embedded and complex processes. They claim that this view of change can give insights into the factors that affect ICT related organizational change. However, they also see the importance of regarding the material nature of technology, since there is a risk that this may be lost in the increasing focus on the social constructivist approach to technology. They found that research had focused on the social shaping of technology and its emergent nature, and has attempted to develop frameworks that see technology as both material and social objects at the same time, and claim that there is much to be gained from a more complex view of the technology.

According to Meyer and Rowan (1977), theories assume that organizations become successful through coordination and control of activity, and this is tied to the assumption that organisations function according to their blueprints, that rules and procedures are followed, and that activities conform to the formal structure. However, much research casts doubt upon these assumptions. It has been shown that many formal organizations are loosely coupled, that rules are often violated, and that decisions are often un-implemented, or lead to uncertain consequences. This fact points to the importance of studying the way individuals in organizations actually perform their work, when designing, developing and using technology in their everyday practice. However, when discussing the development of information systems, Avgerou (2001) shows that the main function has been to develop technologies, to manage and exploit them effectively, whereas the organizational and social change that unfolds in interaction with the technical innovation has not been the focus of studies. So far, there has been a tendency to reduce technology to an abstract and material determinant of work and organizational structure, and to downplay the role of human agency (Barrett, Grant et al. 2006).

Avgerou (2001) discusses the importance of seeing the way in which innovations are embedded in and inseparable from processes of social change, and states that it is important
to research how ICT supports the social activities of organizations. IT innovation is rarely the result of free choice and action. Instead, it is decided by events, trends, pressures opportunities or restrictions in the national or international arena, and is a combination of technical/rational and institutional action. Therefore, analysis should be extended to address the socio-technical processes of technological innovation across layers of context (Avgerou 2001). This is also in accordance with Suchman’s view, when discussing practice-based design of information systems (Suchman 2002). Here, her concern is with work practices, and the particular forms of work that make up the production and use of technology wherever they occur. According to this, one of the central issues is how to understand the relation between existing practices and projects of new technology and organizational design. She claims that the prevailing discourse has been characterised by a situation where anonymous and un-locatable designers, guided by their training, problematize the world. They design and deliver technology to de-contextualized and un-locatable users. These processes of “design from no-where” are tied to the goal of constructing technical systems as commodities that can be exported to the sites of their use.

Harley, Wright et al (2006) discuss the way in which organizational change perspectives tend to see change as imposed by managers on employees. They discuss how middle management resistance to change is often caused by organizational constraints, such as a lack of management support or a lack of time to negotiate, settle differences of opinion, and communicate change messages to teams. These were the kinds of factors that were found in the study at ABB (Winter, Rönkkö et al. 2012) where the study participants pointed out many instances of these kinds of issues.

As has just been shown, one of the issues in the design and development of technology is that there is often a gap between design processes and users of technology. Having looked at the discourse regarding some of the organizational factors that are in operation when designing and developing technology, focus is now placed on who it is who participates in design and development activities, beginning by examining the importance and problems of including representative users, and the ways in which participation can take place.

Focus: Participatory Design
The lens used here is still aimed at looking at organizations, but the area of focus is now moved to participatory design (PD), and the issues of how organizations can ensure that the right people participate in the design and development of the technology. The discussion here has already focused on usability and UX, where it is clear that fulfilling users’ expectations is an important part of producing the right product. Participatory design centres on the way in which users of technology can and should be involved in the design and development of technology, not only from the point of view of designing the right features, but also from the point of view of their right to participate and make decisions.
Researchers in the field of PD are concerned with the conditions and methods for user participation in the design and introduction of computer based systems in the workplace. Traditionally, PD has a range of techniques that are used in practice, and a number of guiding principles. The principles include: the acknowledgement of diversity; ensuring that the participants’ viewpoints have a genuine impact in the design process; mutual learning at all levels; separation of interest groups, and; protection of participants. Methods include visiting workplaces, creating work scenarios, future workshops and organisational games, mock-ups and prototyping (Törpel, Wulf et al. 2002).

A traditional PD approach has two important characteristics. The first is of a political nature, and deals with questions of workplace democracy and control. PD has traditionally had an explicit agenda for organizational and political change (Kensing and Blomberg 1998). In particular, the Scandinavian approach to system design is strongly tied to thoughts and principles regarding democracy within the workplace. “This Scandinavian approach might be called a work-oriented design approach. Democratic participation and skill enhancement – not only productivity and product quality – themselves considered ends for the design” (Ehn 1993, p. 41). The second characteristic is of a technical nature, stating that the participation of skilled users in a development process contributes to successful design and high quality products. Here, PD is seen as a learning process where the user and the system developer, or the designer, learn from one another (Ehn 1993).

Central to the PD approach towards computer system design, is an emphasis on the fact that those who are expected to use the system should be actively involved and play a critical role in designing it. It emphasises the inclusion of all stakeholders in the design process, and it demands shared responsibility, active participation, and a partnership between users and implementers (Schuler and Namioka 1993). There is always the need to move marginalized voices and ways of knowing into the centre of commercial, governmental and public discussion of information and IT (Suchman 2002). According to Sanoff (2007), PD actually covers a broad spectrum, and there is no shared understanding of what PD is, due to the diversity of the background and perspectives of its proponents. However, there is still a shared view of the central principles of PD. This view states that participants are experts at what they do, and maintains that the voices of all participants must be heard. It is recognized that good ideas arise when participants with different backgrounds collaborate, and finds that this should preferably take place in the users’ environment, rather than performing tests in laboratories. There is a political dimension to PD, where the purpose of
participation is not only to reach agreement, but is also aimed at empowering individuals and groups, allowing group participation in decision making, and engaging users in the adaptation and change of their environment.

The concepts of the workplace and the user have, however, changed with the spread of interactive technology, and modern PD also concerns the design and use of technology in other environments that those that can be regarded as traditional workplaces, with other users. In the following discussion, when the workplace is mentioned, it can be interpreted in a wider meaning as relating to all situations where users make use of technology. In the first phase of the research at hand, with UIQ, users were individuals in a mass market, using a product that was aimed at satisfying a huge spectrum of users with one and the same product. In the final stage of the research, at ABB, users were also a diverse group, with different backgrounds and needs, and the product was also aimed at satisfying all of these diverse users. This places new demands on the conditions and methods available, and who participates. A discussion of this reappears later in this section.

To a large extent, the tools and methods used in PD are based on early projects concerned with developing methods for “participative analysis of relations between work and technology and strategies for union influence over technology projects run by managers and IT professionals” (Kensing and Blomberg 1998, p. 174). The goal that is common for tools and methods used in PD is to provide users and designers with methods for connecting current and future work practices with visions of a future technology, by allowing access to the experiences of both users and designers (Kensing and Blomberg 1998). When designers participate in PD activities, it gives them new and better ways of gaining an understanding of the user’s everyday work practices (Schuler and Namioka 1993). Once again, as in the research at hand, the concept of work practices must be expanded to include other activities where people use technology to perform tasks (or for example entertain themselves).

According to Kensing and Blomberg (1998), the three main issues that have dominated research within PD are: the politics of design; the nature of participation, and; methods, tools and techniques for use in design projects. The political aspects of the introduction of computer-based systems have always been explicit in PD, which began in reaction to the ways that computers were introduced in workplaces and the negative effects that this had on workers. Researchers argued that computers were becoming a management tool to exercise control over workers, and the neglect of workers’ interests was at the centre of critique. In order to strengthen the bargaining power of workers, researchers attempted to help build proficiency amongst workers. Some of the assumptions of PD research changed as changes occurred in workplace politics, and focus moved to the rationales of participation and how different actors could influence the way in which technology was implemented and designed.

There are three basic requirements for participation, and these are: access to relevant information; the possibility to take a position on problems, and; the possibility to participate in decision making. Within PD, it is seen as a precondition for good design that the users participate, and taking into account the skills and experience of users increases the
likelihood of systems being well-integrated into work-practices. To design and develop good systems requires that designers have knowledge of technological options, and this knowledge is best developed if users (workers and others) and designers cooperate when design takes place. A change has taken place in who participates in PD activities, and various actors in the organisation, including representatives of management, are now included in PD activities. In addition, as is discussed further below, it is not always possible for all those who are affected by the design effort to participate, and they are often represented by usability specialists, who work as proxies for the users.

According to Suchman (2002), there has long been a simplistic division between users and designers that has obscured the realities of system design. Although the boundary between the groups is highly shifting, it has traditionally been regarded as unambiguous, thereby obscuring the way in which professional designers are themselves amongst the most intensive technology users, and neglecting the multiple forms of design-in-use that actually take place. The simplistic view of a duality between designers and users hides the subtle differences that do actually divide the groups, and obstructs the replacement of the simplistic duality with a rich picture of identities and relations that could help us clarify the actual positions. Therefore, Suchman and others turned to Scandinavian PD, where there is no distinct boundary between technology design and use. This lack of boundaries is important, since if professional designers are to design systems of any integrity, they must develop them in relation to specific use-settings. Furthermore, if technologies are to be made useful, then other practitioners must effectively become designers. This view rejects the view of discrete phases in a system life-cycle. Here integration, local configuration, customization, maintenance and redesign become phases of articulation work, without boundaries between them. This points towards the problem that the boundaries that currently define design and use are realized through institutionalized arrangements that are organised systematically to reproduce them. Later in this chapter, the role played by institutions is returned to.

The need for PD is not only grounded in the politics of design, but also reflects the practice of development as it actually takes place. The Scandinavian school of PD which Suchman turned to had its foundation in the political issues of the time. However, the importance of political issues has declined within the PD tradition, and the current emphasis is more on active user participation, although this has been found to be problematic in practice. Many systems are so large and complex that even if all of the prospective users were known, it would still be impossible for them to participate directly. Most systems are developed in a product development context, with millions of potential users, and these users are often not known at the time the product is developed. In these cases, usability specialists are often called in as proxies for the users, but these experts have not been chosen by the users as representatives, and neither are they typical users (Iivari and Iivari 2006). This was typical of the situation at UIQ. PD projects often take place in a setting that is protected from what Kensing & Blomberg (1998, p. 175), refer to as the harsh realities of organisational life, where resources are allocated to the project, and participants are given time off from their daily work, to be able to experiment with new approaches. However, Kensing and Blomberg (1998) maintain that PD must find forms to survive in the everyday world, with
limited resources, conflicts and time restraints. Törpel, Wulf et al. (2002) found that PD measures are time consuming, and require the allocation of resources. They discuss the fact that when PD takes place within organizations, participants will only be motivated if they receive compensation for their work, and if their duties are performed whilst they are absent. In addition to this, as in the research that has led to this thesis, there are problems of finding ways to find and motivate users to participate in PD activities, particularly when the product is aimed at the whole of the consumer market.

The processes of PD are in many ways connected to the role of the designer, the developer and the user within the organization. In many cases, where development takes place in-house, all of these actors are within the same organization. However, this is far from the case, and in the environments that have been part of this research, the actual users may be general users of a mass market product, or users of an advanced GUI that steers an industrial robot. As early as 2002, Törpel, Wulf et al. (2002) found that the traditional methods of PD, which were developed in connection with projects in large companies, in cooperation with trade unions, are dependent on a number of factors, which are not always present in other kinds of organisations, especially small ones. As Hansson and Winter (2004) found in a study of how to engage parents of school-age children in the development of e-services, it is difficult to find resources for engaging occasional users, external to an organisation, in the work of designing systems, and it was also difficult to find appropriate forms for their participation. Measures to deal with situations outside one organisational context are unusual. Dittrich, Ekelin et al. (2003) found that PD has left the stage where the goals are those of the traditional workplace, and is entering a stage where it must include the goals of users outside of what is traditionally considered to be the workplace. They see that there is a need for a flexible technical infrastructure to accommodate changing requirements regarding services and their provisions, and that there must be continuity, a long-term and stable relationship between developers, local tailors and users, to support the practices necessary to accommodate the changes that are taking place.

Thus, PD is a complex issue, affecting the roles of designers and users, whether they are inside or outside the organization. However, as Suchman (2002) discusses, technologies designed at a distance are usually characterized by a design/use gap that requires substantial reworking of the technology. If the prospective users have sufficient power, it can even lead to rejection of the technology. This strengthens the argument regarding the importance of user participation in design, and in the next section, a new set of lenses is used, that focuses more closely on the mechanisms of designing, with focus on the importance of user-centred design.

**Focus: User Centred Design**

Focus now moves to the field of user-centred design (UCD). Although it is similar to PD, in that users are seen as central figures, and has thereby been influenced by PD, there are differences. The political background to PD is not explicit in UCD. In PD, a central tenet is that practitioners should become designers, but this is not explicit in UCD. Although UCD is built upon the fact that users should be actively involved in design and development, it
does not necessarily call for deeper levels of participation. Thus, if PD is seen as a philosophical approach, UCD can be regarded more as a practical methodology for how to take into account users and their requirements. It could be used as a part of the PD toolbox, but this is not necessarily the case.

As recently as 2010, the user has been characterised as the great unknown in systems development (Iivari, Isomäki et al. 2010). In information systems research, user participation has been a research topic for decades, and according to Iivari, Isomäki et al. (2010), this is understandable, since user participation is related to the distinctive activities of developing information systems, which are: mutual alignment of IT artefacts and the organizational and social context where the artefact is to be used; identifying and specifying the needs of the people who are expected to use the system; the organizational implementation; the evaluation/assessment of the artefacts and the related changes. They maintain that it is hard to see that any of these activities could be carried out without user participation or involvement.

However, user involvement has not been a particular research topic in the software engineering (SE) tradition, even though systems development has long been a subject of intensive research within SE, human-computer interaction (HCI) and information systems. For example, the most recent version of the SE body of knowledge makes no mention of user participation or involvement (Iivari, Isomäki et al. 2010). Thus, despite an acknowledgement in the SE literature that user participation is crucial to the success of the requirements engineering process, it is considered only at the beginning of the development cycle, in requirements engineering or construction. This is in spite of the fact that deficient requirements are seen as the greatest single cause of the failure of software projects (ibid.)

UCD can be regarded as a problematic concept. According to Iivari and Livari (2006), despite the length of time that it has been in existence and despite all of the attention that it has gained over the last few years, the concept of UCD is still unclear. They state that the phrase “user-centred” emerged in the HCI field, but has adopted a variety of ideas from different sources, and has thus adopted a number of meanings from different sources. They found that the ISO 13407 Standard (1999) defines human-centred design in terms of four principles, but the principles are not clearly formulated, and do not define UCD unambiguously. Furthermore, they have found that there are ambiguities regarding the overall goals of UCD. ISO 13047 has been criticised for providing very ambiguous
guidance for designing usability, even though it was defined as “an approach to interactive systems development focusing specifically on making systems usable”.

Gulliksen, Göransson et al. (2003) maintain that applying UCD requires a profound shift of attitudes towards systems development. They agree with livari and livari above, that a clear definition of UCD has been lacking, with the result that it has become a concept without meaning. To remedy this, they have formulated a number of principles for UCD, which summarize their research results and experiences. They are based on principles that are specified elsewhere, and take into account movements such as the Scandinavian tradition of PD. These principles are: User focus; Active user involvement; Evolutionary systems development; Simple design representations; Prototyping; Evaluate use in context; Explicit and conscious design activities; A professional attitude: Usability champions; Holistic design; Process customization; A user-centred attitude should always be established. They found that it was necessary to formulate a new set of principles for UCD, as they found that the existing ones were not used. Even with a focus on UCD, projects still lose sight of usability, and there appear to be obstacles to usability and user involvement in the actual development processes. Their concern has been to address shortcomings in the processes that de-rail the focus on usability and user involvement. In their work, they found, amongst other things, that there was a need for usability champions, who should have the mandate to decide on usability matters, as there was otherwise a clear risk that the skilful and experienced work of the usability designers was ignored in later phases of the development project (Gulliksen, Göransson et al. 2003). However, they found that this was not the case, which reflects a situation that we also found in the research at ABB.

It is not only the concept of UCD that is at question. According to Millerand and Baker (2010), even the concept of the user, in common with a number of other key terms and phrases, is underdeveloped in theory. Common understandings of complex systems, standards and users often rely on simplifying black-box models. Millerand and Baker (2010) claim that roles change depending on the system development method at hand, that users and use-contexts are hard to identify, and that they usually are “discovered” during the system development processes.

However, understanding the concept of the user is important. livari and Isomäki et al. (2010) maintain that companies are forced to understand users’ needs and wishes in order to construct appropriate products, and this necessitates a thorough understanding of the user, or more specifically, an understanding of the developers’ or designers’ interpretation of the users. Thus it is important to understand all stakeholders interacting with the user. However, they also found that, despite many studies on understanding the user and capturing requirements, the user still remains unknown in the majority of the studies. They found that the user is usually regarded as an insubstantial technology user, not as an actor in an organizational setting, and that, even when the user is recognized, the role of the user is often reduced to that of a static entity. This was apparent in the research at ABB.

User groups are becoming heterogeneous, complex and diverse, making it difficult to know who users are, or to identify, contact, involve and represent them (livari, Isomäki et al. 2010). However, technology is not successful if it is not used by its intended users, so it is
still important to understand users when developing new services and applications. Companies that sell products or services to individual users struggle with these diverse groups of target users, where conceptualizing the user is a challenge for systems developers. According to Millerand and Baker (2010) the term “user” has an intuitive simplicity that is used as part of a two-category set of groups, users and developers, that suggests distinct and separated stages of work. First the system is developed, and then it is used. This perspective leads to the emergence of an emphasis on “the” user, even when the user group is not homogeneous. If this concept is challenged, a variety of user roles emerge, and these users are involved in multiple relations, towards objects, actions and settings. This variety of roles was very much apparent in the research presented in this thesis.

Keinonen (2008) found that socially relevant design must be related to the needs of people. He discusses how UCD arises from a need to transform complicated technologies into real-world applications. Previously, UCD was concerned with optimising the design of HCI. Now, it is concerned with a wider range of interests, aimed at matching interactive technologies with human-human communications, organizational requirements, and social and emotional perspectives. This is strongly connected to gaining an understanding of the user, and furthermore is connected to the importance of working towards usability and UX, to support these interests, which is what the research in this thesis has been concerned with.

As previously mentioned, there are connections between PD and UCD. In UCD, the role of the human, who was once reduced to an operator, part of an information processing system, is replaced by a more active and holistic contributor (Keinonen 2008). However, when looking at the dimension of user participation, Iivari and Iivari (2006) found that, despite the fact that the Scandinavian PD approach has had a profound influence on UCD, and that there is a potential for conflicts between management and workers, developers and users, and different organizational groups, power and politics have been surprisingly absent in the UCD literature. They maintain that political milieu where development takes place must be taken into account, including the prevailing power relations in organizations, which make it unrealistic to assume that users alone have the right to decide about changes in the work domain, including IT-systems. Users are only one stakeholder group in systems development, especially when developing systems for work context. However, as will be discussed further in the section on institutions and wicked problems, it is far from clear that these problems are solved.

Furthermore, Iivari and Iivari (2006) point to a problem regarding the diversity of users. The difficulty of involving all types of users in the design process makes it hard to design the system to fit each individual user. This reflects the cases that were studied in this research, which dealt with products that were aimed at diverse user groups in many different contexts. As discussed by Millerand and Baker (Millerand and Baker 2010) there is a movement away from a monolithic view of the user as a group, to a more complex and fragmented one, where users and developers are not stable entities. They see that users and developers adopt multiple roles that evolve throughout the system development process. They argue for the concept of the user in terms of a dynamic web of users, and for
recognition that the boundaries between users and developers are fuzzy and permeable. They see that the recognition of the dynamic nature of the active relationship between the users and developers is a critical issue for design and development. However, achieving this, as also became apparent in this research, has been problematic.

In their study of the varieties of user-centeredness, Iivari and Iivari (2006) found four different dimensions in the UCD literature. They were user-centeredness as: user focus; work-centredness; user participation and; system personalization. In the *user focus* view, the ideal is to take into consideration every individual user’s capabilities and to satisfy each user’s needs. However, it is hard to achieve the ideals of satisfying every user’s needs. Many systems are platforms that are used by a number of users, as was the case in this research, both at UIQ and ABB. Here, not every user can expect that their personal needs can be fully satisfied. Also, systems can have a huge number of globally distributed users, who cannot be accessed when designing the system. This was also the case in this research. Furthermore, involving “real” users is difficult, as previously discussed here, and it has sometimes been seen as unnecessary, where the use of surrogates of users, e.g. personas, has been suggested as a replacement (Iivari and Iivari 2006).

According to Gulliksen, Göransson et al. (2003) even though, as far back as 1995, active user involvement was judged to be the primary criterion for success in IT development, many organisations still do not understand the benefits of involving users. They found how, even with an explicit commitment to UCD and usability, there was still a risk that usability fell by the wayside in development processes. This shows that there are obstacles to usability and user involvement inherent in the development processes. They (Gulliksen, Göransson et al. 2003) found that although user-centred methods have been shown to increase the utility and usability of systems, the degree to which organizations adopt UCD methods varies significantly. They also found that the most common way of performing usability activities has been through the use of informal UCD methods, and that this informal use risks that usability work is cut out when deadlines get tight, which was what was found in the study of usability and UX work at ABB.

The importance of the participation of technology users, and the importance of taking of context into account, becomes obvious. In their study of the key principles for UCD, Gulliksen, Göransson et al. (2003) state that computer systems (in particular in a work context) must not only support the official rules and versions of the work, but must also support the particularities of each situation, and it is only the users themselves who can provide the requisite knowledge to achieve this. However, they found that, although many organizations claim that they are willing to work with UCD and usability, few know how to achieve it. Although they found a growing realization amongst developers about the importance of product usability, there is little knowledge of what to do about it. Despite the fact that active user involvement has been found to be one of the key factors of how to be successful in IT-development projects, organizations still do not recognize the benefits of including users in development processes. There are a number of obstacles and Gulliksen, Göransson et al. (2003) cite (Rosenbaum, Rohn et al. 2000) who list a number of these: Resistance to UCD/usability; Lack of understanding/knowledge what usability is; better
ways to communicate impact of work and results; Lack of trained usability/HCI engineers; Lack of early involvement; Resource constraints; No economic need – customers not asking for usability. Gulliksen, Göransson et al. (2003) think all of these factors are related to a lack of knowledge of how to apply UCD methods, and their potential benefits.

Orlikowski (2000) discusses how designers and developers create technology with a set of properties. However, the way in which the properties are used is not pre-determined, but depends on what people actually do with them in particular instances. Users actually make use of technologies in ways which are situated and emergent; that are unanticipated or unintended by the designers and developers. Despite this, use of technology is not open to all possibilities, since there are physical attributes that bound its use. According to Orlikowski (1992), despite these boundaries, even the most rigid forms of technology have to be apprehended and activated by human users, and in such interaction, the users shape the technology and its effects. Operators of technologies often deviate from formal, rule-bound practices to deal with complex interdependencies, unanticipated effects and local conditions. Suchman (2002) claims that the official accounts of ICT development and implementation tend to neglect local, everyday accounts of innovation. She maintains that it is the job of the researchers to bring this “invisible work” into the light, even though invisibility affords some spaces for action that might otherwise be threatened with closure if it becomes apparent. The research presented in this thesis has been aimed at achieving this.

According to Iivari and Iivari (2006) the human-factors approach has been criticized, as neglecting issues such as the underlying values of the people involved and their motivation in the work setting, so in their work-centeredness view, an alternative approach has been suggested. This involves understanding the relationship between people, technology, work requirements and organizational constraints in work settings, where people are actors in situations, with skills and shared practices. One of the central focus areas of HCI research, and of UCD, has become understanding users’ work activities. In this area, the field has been influenced by socio-technical thinking, and for example, Computer Supported Cooperative Work (CSCW) has focused more on the detailed work practices of users in cooperative settings. In this research, particularly at ABB, it was apparent that there were difficulties when capturing, understanding and making use of knowledge of the users’ work activities.

According to Avgerou (2001), IT innovation and the social context where it takes place are so intertwined that it is impossible to distinguish between technology as content and society as context. Doing so would be a simplification, obscuring the processes where the technology and the human actors together form socio-technical entities. The change associated with technological innovation can be described as a heterogeneous network, conveying the perception of “information systems as social systems”. Technical artefacts carry with them engineers with their conventions, industries that sell, install and support them, users who interpret and use the systems according to their circumstances, and consultants who convert them to competitive advantage. However, Millerand and Baker (2010), show that collaborative elements introduce complexities in the terms of ICT design.
and user involvement, and this raises new challenges. To understand the intertwined social, organizational, technical and political dimensions needs interdisciplinary scholarship and collaborative care, as well as attention to collaborative techniques and development processes.

Orlikowski and Iacono (2001), when commenting upon different conceptualizations of technology, grouped them in five different categories: the tool view; the proxy view; the ensemble view; the computational view; and the nominal view. In their terms, the research discussed in this thesis views technology as a development project, which is part of the ensemble view. In this view, technology is conceptualized as an artefact in formation, and the focus is on the social processes of designing, developing and implementing technical artefacts, often in specific organizational contexts. The roles of key stakeholders are central, together with how these roles engender conflict, power-moves, and the influence of methodologies on development processes. In studies where technology is seen as a development project, there is often a socio-technical perspective, and these studies have helped to deepen the understanding of systems development as a complex process, and how these processes take place in organizations. That has been one of the ambitions of the research in this thesis.

Although the institutional context where the ICTs are introduced may play a role in shaping their effects on organisations, the changes associated with them are not likely to be homogeneous. Their impacts are likely to be shaped by the interaction of institutional patterns and the interests of individuals and groups (Barrett, Grant et al. 2006). The intentions within organisations, such as concentrating on usability, are sometimes waylaid by organizational issues that are difficult to do anything about, because the organization has become an institution. In complex cases, we are dealing with wicked problems, which was found to be the case at ABB.

In the following section, the next lens is inserted in the spectacle frames, in order to look more closely at the role institutions play in the design and development of technology, and finally, focus is placed more closely on wicked problems, and the effect that their presence has on the way in which organizations deal with design and development activities.

1.4.3 Lens: Institutions and institutionalism

Having looked at organisations in general in section 1.4.2, the next lens is used to look more closely at a particular phenomenon that concerns what happens when organizations become institutions. This state became apparent when studying the situation at ABB. Closely related to theories of the way in which organisational change takes place are theories of institutions and institutionalism. According to these theories, actions and structures of organizations cannot be adequately accounted for by the technical and rational norms of competition and strategic behaviour (Agerou 2001). Institutional theory states that patterns of action become established as standardized practices in organizations. Over time, such practices become institutionalized, forming structural properties of organizations. These properties are drawn upon by humans in their interactions, and such use also reinforces the institutionalized properties (Orlikowski 1992). Meyer and Rowan (1977) discuss institutions in terms of what they call the myths that create and reinforce
them. Elements of formal structure are ingrained in a widely spread understanding of social reality, and the elements of social structure are manifestations of institutional rules that function as powerful myths that are binding on organizations. These myths identify social purposes as technical ones, and through rules, they specify appropriate means to pursue technical purposes. They are furthermore highly institutionalized and are thus beyond the control of individual participants or even organizations.

Avgerou (2001) also discusses the role of myths and institutions, and maintains that it is impossible to explain what is happening in organizations by considering only rational acts of managers and technology experts. Avgerou states that the core of institutionalist theory is that formal organizational structures and processes are not maintained because they are the most efficient ways of carrying out complex activities, but because they are sustained by powerful myths: meaning-laden public knowledge about how organizations should function. Institutional forces can be social, cultural or cognitive in their nature, and the sources of these forces may be other organizations, formal legislation, professionals or the collective values of society. They determine not only the legitimacy of organizational forms, actions or missions, but they also set the factors that must be taken into account in the choices of technical rational action.

In 2001, Orlikowski and Barley (2001) stated that IT researchers had yet to ask how institutions influence the design, use, and consequences of technologies, either within or across organisations. They explain that studies have taken place at a “lower” level of abstraction, and have not looked at how regulative processes, normative systems, and cultural frameworks shape the design and use of technical systems. However, they maintain that understanding and guiding techno-social developments requires not only knowledge of technological systems, but also of social processes and their interactions. There is a need for research that concerns both technology and the social dynamics of organising, which in addition to understanding the constraints and affordances of technologies also embraces the importance of understanding the role of the human agency as embedded in institutional contexts. The impression given by the research in this thesis is that this need is still in existence today.

According to Barrett, Grant et al. (2006), in the field of organizational studies, the growth of awareness concerning institutionalism is a rediscovery of the importance of institutions. Institutions are seen to be a product of past decisions that engender a particular way of doing things and they develop a taken-for-granted character. Past decisions structure and affect future decisions. Meyer and Rowan (1977) discuss how formal organizations are traditionally seen as systems of coordinated and controlled activities that arise when work is embedded in networks of technical relations and exchanges that cross boundaries. However, they also discuss how the formal structures of many organizations reflect the myths of their industrial environments, rather than the demands of their work activities, and in this situation, rules become institutionalized. They state that institutionalized rules are distinguished from social behaviours. They may have an effect on organizational structures and their implementation in technical work is very different from the effects caused by networks of social behaviour.
However, it is not only organizations that become institutionalized. IT is also seen as an institution, with its own set of norms and patterns, which interacts with but is not determined by the institutional characteristics of organizational practices (Barrett, Grant et al. 2006). Technology is not an objective force or a socially constructed product, instead, it is a product of human action, and it assumes structural properties. Once developed, however, technology tends to become institutionalized, losing its contact with the humans who gave it meaning, and becomes part of the objective and structural parts of the organization (Orlikowski 1992). Furthermore, institutions have a central impact on the way in which technology is constructed. According to Orlikowski (1992) studying the design and development of technology shows how designers, influenced by the institutional properties of their organization, fashion the technology to meet managerial goals. Studying the use of technology in an office or factory shows how users are influenced by the technology itself, and how the technology affects the institutional properties of the organization.

The fact that people work within institutions necessarily affects the way in which they use technology. People use the properties of the technology, but draw on their skills, power, knowledge and assumptions about the technology and its use. Furthermore, they draw on their knowledge of the institutional contexts in which they live and work, and the conventions of acting in these contexts. Thus, their use of technology is structured by all of these factors, which leads to a set of rules and resources that structure their future use as they interact with the technology (Orlikowski 2000).

However, Orlikowski (2000) discusses the way in which even institutionalized technologies are never fully stabilized, and can be changed through human actions, which is how all social structures are changed. As their knowledge or needs change, users can enact different technologies-in-practice. Technologies can be looked at through a practice lens, examining the situated use people make of technology, focusing on the structures that emerge as people interact with the properties of the technology at hand. According to Feldman and Orlikowski (2011) the practice lens is concerned with the notion that social life is an on-going production, and that it emerges through the recurrent actions of people, and this lens can be used to analyse social, technical and organizational phenomena. This is valuable, since contemporary organizing is increasingly seen to be complex, distributed, dynamic and transient, and requires approaches that can help us theorize these kinds of indeterminate and emerging phenomena. The practice lens, with its focus on dynamics, relations and enactment, offers powerful analytical tools to help here. Orlikowski (2000) says that the use of the practice lens shows that, even though technologies-in-practice can become institutionalized over time, this is only a temporary stabilization, and there is always the possibility to enact different structures. This means that the practice lens can focus on human agency and the set of emergent structures that can be enacted through the use of technology. As discussed further in the section on the research methodology that has been used in this research, the research approach used when viewing the organizations, in the studies at the two participating companies, contains elements of this approach.
Having looked at the particular case of organizations that become institutions, focus now moves to the situation that can arise when technology is designed and developed in a complex context, for use in a complex context, where institutional structures bound what actually gets to be done, and what is left undone. This can be seen as a wicked problem.

**Wicked problems**
The lens used here is still adapted to looking at institutions, but is now focused upon a specific type of problem that can be found in many complex areas, from social planning, to large software development projects. This section deals with wicked problems.

![Figure 1.8 The Institutions lens, focusing on Wicked Problems](image)

The concept of wicked problems originates in the work of Rittel and Webber (1973) who discussed dilemmas in societal planning. They pointed to the fact that the professional’s job was traditionally regarded as finding solutions to problems that are definable, understandable and consensual. These types of problems were characterised as being “tame”. However, they maintain that a shift occurred in the types of problems that professionals must address, where societal problems are characterised as being ill-defined; reliant on political judgement for solution. The term wicked is used for these problems, “in a meaning akin to that of ‘malignant’ (in contrast to ‘benign’) or ‘vicious’ (like a circle) or ‘tricky’ (like a leprechaun) or ‘aggressive’ (like a lion, in contrast to the docility of a lamb).” (Rittel and Webber 1973, p. 160). They discuss the difficulties society has when dealing with wicked problems in a planful way. They maintain that this is difficult because there is a plurality of publics with diverse valuative bases, pursuing a diversity of goals, making it hard to set goals. They claim that the traditional approach has been to entrust-decision making to professional experts and politicians. However, they maintain that, even if this is ethically tolerant, such a tactic does not solve the problem, since there are no value-free, true-false answers to any of the wicked problems that government must deal with. Although Rittel and Webber discussed societal planning, the research presented in this thesis has found clear parallels when studying descriptions of the way in which usability work took place at ABB.

According to Rittel and Webber (1973), there are ten distinguishing properties that characterise a wicked problem. These are:
1. There is no definitive formulation of a wicked problem.
2. Wicked problems have no stopping rule.
3. Solutions to wicked problems are not true-or-false, but good-or-bad.
4. There is no immediate and no ultimate test of a solution to a wicked problem.
5. Every solution to a wicked problem is a “one-shot operation” because there is no opportunity to learn by trial and error, every attempt counts significantly.
6. Wicked solutions do not have an enumerable (or exhaustively describable) set of potential solutions, nor is there a well-described set of permissible operations that may be incorporated into the plan.
7. Every wicked problem is essentially unique.
8. Every wicked problem can be considered to be a symptom of another problem.
9. The existence of a discrepancy regarding a wicked problem can be explained in numerous ways. The choice of explanation determines the nature of the problem’s resolution.
10. The planner has no right to be wrong.

It is suggested that the reader goes to the original text for more detailed information regarding these properties.

So, according to Rittel and Webber (1973), societal problems are in contrast to the problems addressed by natural sciences, in that they are not definable and separable, and have no solutions that are findable. However, although the original discussion of wicked problems deals with the problems of government and societal planning, parallel types of situations and problems have been found in many different fields. The change in the types of problems that must be addressed, and the presence of ill-defined and complex problems, is not unique to the area of societal problems: in the research presented in this thesis, many clear parallels have been found to the situation discussed by Rittel and Webber.

The concept of wicked problems has been referred to in many cases within the area of the design and development of technology. Keating, Rogers et al. (2003) discuss the ways in which systems engineering approaches have successfully addressed well-bounded problems, and how engineering disciplines have traditionally been directed at solving clearly defined problems, with clear goals, that can be attacked from an established body of theory and knowledge. However, they find these conditions are no longer the rule, and increasing information richness, and higher degrees of complexity lead to a need to address a situation where complex systems problems are emerging. These environments are characterised by ill-defined and potentially tacit goals that are value-laden and that may be impossible to define explicitly.

Curtis, Krasner et al. (1988) found that the problems that are experienced in large system development projects were not experienced in smaller well-understood applications with stable and complete specifications, and they characterise these problems as wicked problems. Hughes, King et al. (1994) states that design often means dealing with wicked problems and is, at best, a ‘satisficing’ activity. This involves being governed by, and influenced by, the interplay of political, moral as well as technological considerations. Arias, Eden et al. (2000) also characterise design problems as wicked, and state that the
design context is characterised by change, conflict and multiple stakeholders. They state that communication breakdowns often occur, since different stakeholders use different norms, symbols and representations. According to Camillus (2008), designing systems means dealing with a wicked problem, and it is thereby impossible for companies to develop models of the complicated environment where they work. It is the social complexity of wicked problems that make them hard to manage. Denning (2007) describes wicked problems as messes, which are seen as being worse than problems. Problems are difficult situations where there is an expectation that solving some difficulty will end the problem, whereas messes are complex, and there is often no agreement on whether the mess consists of a single problem, or the convergence of many problems. According to Denning & Yaholkovsky (2008), messes are often found to be non-technical in origin.

Farbey, Land et al. (1999), when discussing the problems of evaluating IT projects, found a number of important factors that make the evaluation of IT projects into a wicked problem. These are: Change, Multiple stakeholders, Multifactorial problems and Emotion. This is interesting in a wider perspective, particularly when they discuss a number of learning themes that should be taken into account. These are: Evaluation theory; Stakeholder theory; Evaluation as part of decision making; Project dynamics and; Management learning. Stakeholder theory has been particularly relevant for this the research in this thesis. They found that systems development and project management methodologies have given little attention to the role of multiple external stakeholders, and few attempts have been made to analyse this phenomenon in detail. Relationships between stakeholders were found to be multi-stranded, and stakeholders were unaware of, or not able to fathom the larger picture. Their findings related to management learning were also interesting, and they found it to be important to challenge common assumptions about IT projects. Strategic thinking was not maintained in projects, or even present from the outset, and the original strategic thinking is lost. Project boards perceive projects differently to users, who may well be peripheral, and one of the roots of failure may be differences between groups. Projects are usually seen as “fixed” from the moment of go-ahead, and focus is placed on implementation in time and on cost, but projects exist in fluid environments and it would be better to focus on emerging strategy and benefits. All of these factors are instrumental in the way wicked problems occur and must be approached and dealt with. In the case of the research at ABB, there were also problems involving multiple external stakeholders, and problems in gaining a clear view of the larger picture. There were clear differences in the views of users and project boards, and important issues were left unattended because of fixed views of what was important to include in the project budget.

Keating, Rogers et al. (2003) discuss the ways in which systems engineering approaches have successfully addressed well-bounded problems, and how engineering disciplines have traditionally been directed at solving clearly defined problems, with clear goals, that can be attacked from an established body of theory and knowledge. However, these conditions are no longer the rule, and increasing information richness, and higher degrees of complexity lead to a need to address a situation where complex systems problems are emerging. These environments are characterised by ill-defined and potentially tacit goals that are value-laden and that may be impossible to define explicitly. Once again, this view is supported by
the situation that this research uncovered at ABB, which is moving from a systems engineering context to a situation where new and unaccustomed problems are arising.

As mentioned previously, design is, at best, a ‘satisficing’ activity, often dealing with ‘wicked’ problems, and that this involves being governed by, and influenced by the interplay of political, moral as well as technological considerations. Satisficing is the process whereby humans make decisions and solve problems, even when faced by many alternative possibilities. Decision-makers set feasible goals and use decision methods that look for good, or satisfactory solutions, instead of trying to find optimal ones, as rationality might suggest (Atwood, McCain et al. 2002). Thus, despite the problems, we do find solutions to problems, one way or another.

Camillus (2008) states that designing systems is difficult because there is no consensus on what the problems are, let alone how to solve them. He claims that dealing with a wicked problem involves experimenting with a number of strategies, and is a science of muddling through. One strategy is to involve stakeholders, document opinions and communicate. In other words, solutions are reached, despite the problems involved. Rohde, Stevens et al. (2009) also agree that IT design is a wicked problem, and they discuss what they call a symmetry of ignorance, where communication breakdowns between users and designers are often experienced because stakeholders belong to different cultures. Thus, it is difficult to articulate precisely what the real benefits of a design solution for a certain practice are. They state that developing IT artefacts is a creative process, and that the effects of the development cannot be anticipated, since the intention of the design and its actual use are interwoven. They maintain that design is a social practice, and that the action of designing artefacts and introducing them in organizational settings, thereby develops other social practices. Their suggested solution consists of an analysis of social factors, activities, artefacts, design processes and organisational change (Rohde, Stevens et al. 2009).

This final paragraph summarises well many of the issues that arose in this research, particularly in the latter stages of the research. The next section describes the research methodology and methods that were used when investigating these types of complex problems, and shows why the combination of approaches that has been used has been appropriate to answer the research questions that arose.

1.5 Research Methodology

This section presents the research methodology that has been used during the course of the research, discusses how and why the chosen methods have been used, how and why some of the methods and approaches changed during the course of the research, and how the methods were applied within the different phases of the research projects.

Figure 1.9 shows the different approaches that have been used during the research process. The way in which they were used, and the methodological focus changed over time. As can be seen, Cooperative Method Development with its foundation in action research was influential through most of the research, but was less influential at ABB, when focus moved to institutions and the study became interview based. A case study approach was initially used, but focus moved from case study methodology to a closer focus on grounded
theory, and the focus there gradually shifted to constructivist grounded theory as the research progressed. This is discussed in greater detail in the following sections of this chapter.

![Diagram of research methodologies and approaches used in different phases]

**Figure 1.9 The research methodologies and approaches used in different phases**

This research has been concerned with studying, and finding ways to improve, the processes by which technology is designed and developed within organizations, and the demands and expectations users have when they make use of technology to perform actions. Orlikowski and Iacono (2001) discuss the way in which technology can be conceptualized, and they divided it into five broad meta-categories that carry with them a set of assumptions about IT in research. In the tool view, technology is an artefact that does what the designers intend it to do, and this view is said to “black-box” technology. The proxy view focuses on elements that stand for properties of the IT, and assumes that critical elements of IT can be captured through quantitative measures. The ensemble view sees technology as one element in a package, where variants of the view (technology as a development project, as a production network, as an embedded system, or as a structure)
focus on the dynamic interaction between people and technology during the design, construction or use of the technology. In the computational view of technology, research is concentrated expressly on the computational power of IT. In the nominal view, technology is regarded as being absent, and technology is invoked in name, but not in fact. IT artefacts are not described, conceptualized, or theorized.

This research has viewed technology from the ensemble view, and focus has been on technology as a development project, and as a structure. When regarding technology as a development project, it is conceptualized as an artefact in formation, and focus is placed on the social processes of designing, developing and implementing technical artefacts, often in specific organizational contexts. The roles of key stakeholders are central, together with how these roles engender conflict, power-moves, and the influence of methodologies on development processes. Studies of technology as a development project often have a socio-technical perspective, which helps to deepen the understanding of systems development as a complex process, and how these processes take place in organizations. When seeing Technology as a Structure, focus is placed on the ways in which technology is enmeshed in the conditions of its use. The technology is seen to embody social structures which have been built into the technology by designers and developers during its development, which are then appropriated by users as they interact with the technology (Orlikowski and Iacono 2001).

The research in this thesis has always considered the importance of taking into account practice and context in the organizational environment where the work is performed. This is supported by Orlikowski and Barley (2001) who maintain that, in order to understand the transformations that occur in the nature of work and organising, it is necessary to consider both the technological changes and the institutional contexts that are reshaping economic and organisational activity. Furthermore, Avgerou (2001) maintains that it is of crucial importance that research and practice associate innovations in technology with the context where they are embedded. Innovations must be considered in relation to socio-organizational change, analysis should consider both national and international context, and should consider not only technical/rational decisions and actions, but also the social and cognitive forces of the process. However, Avgerou saw that little research of technology development systematically considered the organizational or the broader context where the innovation was embedded, and maintains that ignoring context entails the risk of misguiding and frustrating efforts to make sense of and appropriate technology. Thus, the significance of considering context cannot be over-emphasized (Avgerou 2001), and the importance of context has been of particular importance in the research presented in this thesis. To study context places particular demands on the research approach that is chosen, and the approach taken in this research has been specifically aimed at enabling a study of the complexity exhibited in the software design and development context.

Since most engineering requirements can be met with multiple designs, all technologies represent a set of decisions made by specific designers. Technologies are both social and physical artefacts, and therefore, neither a constructionist nor a materialist stance alone is adequate for studying technologies in the workplace. Therefore, according to Orlikowski
and Barley (2001), accounts of technological change need hybrid explanations that weave human actions and choice with the functions and features of technology and the contexts of technology use, in a way that attends to the micro-dynamics of situated practice.

The approach that has been used in this research, when studying the development and use of technological artefacts, can be equated with what Orlikowski calls the practice lens (Orlikowski 2000). This approach implies a focus on knowledgeable human action and the way in which technology use constitutes and reconstitutes emergent structures of using the technology, seen as technologies-in-practice. The practice lens assumes that people are knowledgeable, purposive and inventive agents, who engage with technology in multiple ways to accomplish their various ends. We know that technologies are frequently used in ways that are not intended or designed for. A focus on practice lets researchers examine the processes by which this happens, and to what extent users realize the designers’ intentions for technology. Looking at technologies-in-practice rather than technologies as artefacts enables an exploration of the ways in which these processes take place.

The first phase of this research, together with UIQ, provided data for a study of the factors that are important in for success in industry-academia collaboration (Wohlin, Aurum et al. 2012). The study showed that industry does not generally view the research environment as having the same importance as academia does. The participants in industry stressed the need for buy-in and support from company management much more than academics in the study did, illustrating how important it is that academia has an understanding of the need for management support in collaborative projects. In the research presented in this thesis, the fact that the company was a partner in the large-scale BESQ research environment (BESQ 2004) shows that the management at UIQ were committed the research. Furthermore, it was fortunate that this research could be built upon previously established research contacts within the company. A central factor in the research was to engage and spread ideas and results, to many different stakeholders on several organizational levels. This ensured that management, amongst others, were kept informed and could see benefits of the research. Another important factor pointed out in Wohlin, Aurum et al. (2012) is that there must be a research champion at the company; not only someone who is assigned responsibility for the research, but who actually ensures that it is encouraged and progresses. Here also, it was possible to build upon established contacts, where the head of the interaction design department functioned as a strong champion, ensuring that access was given to key operators in the company, and opening many doors that otherwise may have remained closed. One more interesting factor found by Wohlin, Aurum et al. (2012) was that social skills were regarded as being particularly important in a long-term collaboration, and this fits in well with the idea that researchers in social research fields need particular skills. Robson (2011, pp. 133-134) maintains that researchers doing flexible research (Robson’s description of qualitative research: see section 1.5.6. in this thesis regarding this) need a number of personal skills over and above being experienced and well-trained in research. He states that personal qualities are needed, such as having an open and enquiring mind, being a good listener, having general sensitivity, and responsiveness to contradictory evidence. Beyond this, he states that there are a number of
general skills that are important, such as being good at asking questions and listening, being adaptive and flexible, interpreting the situation to get a grasp of the issues, and lacking bias.

In the second phase of the research, the cooperation was with the corporate research team at ABB, not with the company as a whole, and the research could not be regarded as being equally demanding of company resources, or possibly intrusive, as the research at UIQ. These success factors were therefore not fully as important. However, even here, the fact that the research programme manager instigated the study, and that the researchers from corporate research were so committed to the study, meant that the preconditions for successful research cooperation were in place.

To describe the research in the broadest possible way, it can be characterised as empirical research performed over a period of seven years, in close cooperation with industrial partners. A mix of methods and approaches has been used, in line with Eisenhardt (1989) who states that that triangulation through the use of multiple data collection methods is important, as it provides substantiation of constructs and emerging theory. Eisenhardt also sees that the combination of different types of data, which can be both qualitative and quantitative, can be synergistic, and this has also been a feature of this research.

The research has been based on what is usually described as a qualitative approach. According to Denzin and Lincoln (2000) the term qualitative implies that entities, processes and meanings are not experimentally examined or measured – if they are measured at all – in terms of quantity, amount or frequency. Instead, emphasis is placed on the qualities of these factors, and stress is placed on the socially constructed nature of reality, the intimate relationship between the researcher and that which is studied, and the situational constraints that frame the research. Qualitative research tries to examine how social experience is created and given meaning, and emphasises the value laden nature of inquiry. Quantitative studies are, on the other hand, based on an assumption that they are performed within a value-free framework, and emphasise the measurement and analysis of causal relationships between variables. Thus, even though statistical methods have been used, for example statistical analysis of field data, the data have not been collected or used in a manner that could be called quantitative research.

Thus, the research has been concerned with building theory, and has been performed with a qualitative, or interpretive, research approach. According to Charmaz (2010), the most prevalent definition of theory derives from positivism, where theory is seen as a statement of relationships between abstract concepts that cover a range of empirical observations. Here, theoretical concepts are viewed as variables, and definitions are constructed for hypothesis testing through empirical measurement. Theory is constructed in order to explain and predict. The interpretive theoretical approach emphasizes practices and actions, and calls for an imaginative understanding of the situation, where we interpret the participants’ meanings and actions, whilst they at the same time interpret ours. The process of analysis is shaped by our actions, and we create explications, organizations and presentations of the data, rather than discovering order within the data. This view of theory emphasizes understanding, rather than explanation, and theoretical understanding is seen as abstract and interpretive. Interpretive theories do not seek causality or depend upon linear
reasoning, but prioritize showing patterns and connections. The concept and role of theory are taken up again in the more extensive discussion of grounded theory that takes place in following sections below.

Orlikowski and Yates (2006), summarising some of the common themes found in research on ICT and organizational change, found that most of the articles emphasise how important it is to attend to what people do with technologies in practice. Field studies are important when studying what people actually do, as they can help to account for the messy, dynamic, negotiated, improvised and multi-level character of ICTs. They maintain that a situated and on-going view of technology-mediated organizational change is served by a focus on the everyday practices of key actors, and a focus on everyday actions entails examining the micro-level activities that shape and are shaped by the technological and institutional conditions and consequences. The field work in this research, and the methodologies that have informed it, were designed to take into account this complexity, with a focus on how everyday work is performed, on many levels in the organisations, in order to gain an understanding of the complexity involved.

Having given a brief summary of the research methodology as a whole, the following gives a brief description of the projects that have been part of this research.

1.5.1 The research environment

The research took place within two companies; firstly together with UIQ Technology (UIQ) and finally together with The ABB Group (ABB Group 2012). At UIQ, the research focus was on usability and user experience, and research efforts centred upon the development and use of a usability test framework. This research involved day-to-day cooperation with the members of the interaction design team. I had a workplace at the company, and had access to company documentation, and spent a lot of time in the field, getting to know the company and its operations. I interviewed members of staff, observed and participated in meetings and working procedures, and studied the development methodologies and project models in use at the company. This took place over a period of four years, between 2005 and 2009, until the company ceased operations. After the closure of the company, the research process continued as work progressed with the large amount of research data that had been collected during the initial stages. The results of this research are presented in chapter two to chapter five.

The research process continued with researchers from ABB Corporate Research – Sweden (ABB), who had read and were interested in the research with UIQ, and the work within usability testing. Thus, the research became part of a cooperation with ABB, and together, an interview-based study was performed, aimed at clarifying some of the problems ABB had experienced when working in the area of usability, based in a desire to operationalize usability Key Performance Indicators (KPIs) in the organization. The results of this research are presented in chapter six.

1.5.2 Cooperative Method Development

The research process has been influenced by the research and method development methodology Cooperative Method Development (CMD), (see Dittrich 2002; Dittrich,
Rönkkö et al. 2005; Rönkkö 2005, Chapter 8; Dittrich, Rönkkö et al. 2007) for further details).

The approach taken in CMD is to combine qualitative social science fieldwork methods together with problem-oriented method, technique and process improvement (Dittrich, Rönkkö et al. 2007). CMD research is founded in existing practice in industrial settings. It is motivated by an interest in use-oriented design and development of software. However, it is not specific for these methods, tools and processes.

CMD is built upon a number of guidelines, which are:

- Ethnomethodological and ethnographically inspired empirical research, combined with other methods if suitable,
- focusing on shop floor software development practices,
- taking the practitioners’ perspective when evaluating the empirical research and deliberating improvements,
- involving the practitioners in the improvements (Dittrich, Rönkkö et al. 2007)

The research process in CMD consists of the following three phases:

**Understanding Practice.** The research begins with empirical investigations, to build an understanding and explanation of existing practices and designs from a practitioner’s point of view. This is based on the practitioners’ historical and situational context, and is in order to identify aspects that are problematic from the practitioner’s point of view.

**Deliberating Improvements.** The results of the first phase are used in a cooperative manner by the researchers together with the practitioners. These results are used as input for the design of possible improvements. The outcome of this phase is the deliberation of measures.
that address some of the identified problems. These measures are expected to improve the situation at hand.

**Implement and Observe Improvements.** Improvements are implemented, and the researchers follow these improvements as participant observers. The results are evaluated together with the practitioners. During the evaluation process, the concrete results are summarised for the companies involved. The results build a base for the researchers to evaluate the proposed improvements.

The intention of CMD is that it should be a framework that can be further detailed in relation to specific projects, and can even be limited to the first phase in the model if no improvement is decided upon (Dittrich, Rönkkö et al. 2007, p. 233).

The research performed at UIQ involved the development and implementation of a usability test framework, which came to be called UTUM (UIQ Technology Usability Metrics). UTUM evolved through a number of versions, became a standard tool in the development processes, and was also used by other organizations. The research work involved studying the research field of usability and UX, whilst also studying the operations and structure of the organization. This knowledge was then used to inform the development, introduction and use of the test framework and the results that it led to. The study also examined the way that stakeholders in the organization made use of the results of the usability testing. Thus, at the same time as the framework was developed, the research process was influencing the development, studying the process of how it was developed, studying the results of the testing and how these were used, and helping to spread the results, both as research publications and within the company.

This can be equated with an action research approach. Action research is a “vague concept but has been defined as research that involves practical problem solving which has theoretical relevance” (Mumford 2001, p. 12). In a similar fashion to CMD, action research often takes the form of a spiral process, involving observation, planning a change, acting and observing what happens following the change, reflecting on the processes and consequences, and then planning further action and repeating the cycle (Robson 2011, p. 190). Thus, the purpose of action research is to influence or change some aspect of whatever the research has as its focus, and improvement and involvement are central to it (Robson 2011, p. 188). It involves understanding a problem, and generating and spreading practical improvement ideas. Furthermore, it involves the application of the ideas in a real world situation and the spreading of theoretical conclusions within academia (Mumford 2001). Action research is important when studying complex, multivariate phenomena in the real world, which cannot be studied with more positivist approaches, and it is especially important in situations where organizational change and participation are needed (Baskerville and Pries-Heje 1999). To introduce change leading to improvement, to understand problems and spread practical improvement ideas, to apply the improvements in an actual work setting, and to spread the theoretical ideas in academia were all important facets of the research performed at UIQ. Furthermore, there was also a focus on participation and organizational change.
The terms participatory research and participatory action research are sometimes used as synonyms for action research, since a central aspect of action research is collaboration between researchers and those who are the focus of the research, and their participation in the process (Robson 2011, p. 188). The importance of participation is one of the underlying themes that have informed this research. A focus on participation is one of the key guidelines of CMD, as listed above, but it is not only from the AR point of view that participation has been seen as important. The focus given to participation is also connected to the importance given to the principles of Participatory Design (PD). In a PD approach to computer system design, those who are expected to use a system are actively involved and have a central role in the process of designing it, and it demands shared responsibility, active involvement and partnership between users and implementers (Schuler and Namioka 1993). Since an overview of PD has already been given when looking at the focuses that have been central to this work, this will not be presented in any further detail here; it is sufficient to say that the principles of PD have informed this work, and have been yet another element in finding and representing the members’ point of view.

The concept of the members’ point of view is an integral part of ethnographic research. At UIQ, the research was performed as the continuation of previously established research cooperation, with a network of contacts, and positive experience of performing research together. This eased the introduction into the research process, and allowed close cooperation in the everyday operations of the company. This meant that, from an early stage, it was possible to have a focus on ethnographically inspired research methods.

Ethnography has its foundations in anthropology and is a research strategy taken from sociology. It is a method that relies upon the first-hand experience of a field worker who is directly involved in the setting that is under investigation (Rönkkö 2010). Ethnographic practice is based on a number of basic principles: Natural settings; Holistic; Descriptive; Members’ point of view (Blomberg, Burrell et al. 2002). Since people have a limited ability to express what they do or how they do it, activities should be studied in their natural settings. Activities must also be understood in the context where they occur, and not taking a holistic view of activities can lead to an incomplete understanding of what is taking place. Ethnographic accounts should be descriptive rather than prescriptive, even though accounts can be used to point out problems or possibilities for change. The ethnographer should see the world from the point of view of the people who are studied, and describe the world in a way that is relevant and understandable for the participants. This is what is referred to as the members’ point of view. An ethnography should “present a portrait of life as seen and understood by those who live and work within the domain concerned” (Hughes, King et al. 1994, p. 430). This focus upon the members’ point of view is also reflected in the Grounded Theory (GT) approach, which was made use of, and which is discussed in more detail below. The origins of GT in fieldwork, and the fact that social interaction is a central part of it, make it a good fit for ethnographic research (Timmermans and Tavory 2007).

The original focus of ethnography was on the study of small scale societies, but the use made of ethnography in this research context is in line with changes that have occurred in the focus of ethnography. Ethnography has now moved to the study of specific settings in industrialized societies. It has become an approach that is found in most of the social
Chapter One
Introduction

sciences, and in interdisciplinary fields such as HCI and Human Factors Engineering (Rönkkö 2010). In the context of the study at UIQ, the history of close cooperation meant there was access to the organisation and the research process was accepted as a natural part of the cooperation between the university and the company. Thus, it was possible to gain a holistic view of operations and study the natural setting in which the work is performed. The openness that was a part of the climate of cooperation meant that it was possible to observe and participate in the meetings, discussions and everyday work, even on visits to an external industrial partner in England.

There is no single method in use in ethnographic research. Instead, there are a number of methods that are in use within the field. These have been developed to enable the achievement of a “descriptive and holistic view of activities as they occur in their everyday setting from the point of view of study participants” (Blomberg, Burrell et al. 2002, p. 967). The methods used include observation, interviewing, self-reporting techniques, remote data collection, artefact analysis and record keeping (Blomberg, Burrell et al. 2002; Rönkkö 2010). In these studies, the main methods have been participation, observations and interviews, combined with literature studies of the different areas of interest. Despite the fact that ethnographic methods have been used, concerned with understanding the events in the everyday practice of the practitioners in the study, the process has also gone further than the descriptive side that is characteristic of ethnography. This is in accordance with the action research agenda, where there is an intention and readiness to introduce change.

The next section of this chapter looks more closely at the way in which case study methodology has been used in this research. Walsham (1995), when discussing the nature and methods of case studies in information systems (IS) research, takes his starting point in the ethnographic research tradition. Although this was written with its starting point in IS systems, the description is equally relevant for software engineering. He states that the ethnographer is faced with multiple complex structures, superimposed on one another, which must first be grasped and then made understandable to others, and that this is equally true for a researcher entering an organization, who is in the same way faced with multiple complex and intertwined structures that are difficult to grasp and make intelligible. Therefore, “thick descriptions” are needed when trying to understand what is happening in connection with computer-based information systems (Walsham 1995).

The research material has developed into a thick description, in the research log book that was maintained, and in other material collected and associated with the study. However, a stringent ethnographic study has not been performed. This, combined with the fact that the work has been performed in the field of software engineering, where ethnography is not always regarded as one of the standard tools in the research community, affects the way in which the results are presented. They are not presented as a traditional ethnography, as thick descriptions. Instead, the work is presented in a form that is adapted to the software engineering discourse. This is also a typical feature of the way in which ethnographic field work is presented in GT research, discussed further below, where priority is given to the processes or phenomena involved, rather than describing the setting, and these types of studies are often more analytical than descriptive (Timmermans and Tavory 2007)
1.5.3 Case Studies

The first stages of the research were based on a case study approach. A case study is “an empirical enquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” (Yin 2003, p. 13). This fits well with the focus on real life and the importance of studying context, which has been a part of this research. There are a number of important features of a case study (Robson 2011, p. 136). Firstly, it is a strategy or an approach, rather than a method. It is empirical, meaning that it relies on the collection of evidence about what is happening in a particular setting, and it is concerned with research in a broad sense, including evaluation. It uses multiple methods of evidence or data collection, to study a phenomenon in its context, and is a study of the particulars of a specific case. One particular concern is how the case can lead to generalizable results, and how this can be achieved. Thus, it focuses on a particular case, and its context, and it usually involves multiple methods of data collection, where qualitative data are almost always collected, but can also involve the collection of quantitative data (Robson 2011, p. 135). However, as previously noted, the collection of quantitative data does not mean that the study has been framed in a quantitative research approach. Compared to the description of this research in the previous section, detailing the real-life focus, the focus on setting and context, the collection of many types of data using multiple methods, the connection is clear to the case study approach.

In case studies, it is important to overlap data analysis and collection, thereby giving the researcher a head-start in analysis and allowing a process of flexible data collection, which allows taking into account emergent themes or special opportunities for collecting data (Eisenhardt 1989). Adding different data collection methods is permitted since the purpose is to understand cases in as much depth as possible, and adding methods can help to ground the theory or provide new theoretical insight (ibid.). This has been relevant in this research where the data has been collected through interviews, observation and participant observation, workshops, and documentary methods.

In case study research, interviews are often used as primary data, and they are an efficient way of gaining rich empirical data (Eisenhardt and Graebner 2007). A key approach here is using numerous, knowledgeable participants who view the focal phenomena from different perspectives, where the actors can be from different organizational levels, groups and places (ibid.) This is an approach that has been used in all phases of the research, and is an example of the way that theoretical sampling has been performed in the research. Selection of cases is important, and the population defines the set of entities from which the sample can be taken. However, when building theory, the sampling of cases is purposive, and the goal of theoretical sampling is to choose cases which are likely to extend or replicate the emergent theory (Eisenhardt 1989). Since this research has concerned different aspects of two separate cases, the term “case” is interpreted to even include different aspects of one case, seeing these studies as separate, although they are influenced and build upon one another. The principle of theoretical sampling is also interpreted to include the way in which the participants have been chosen in the different phases of the research. Theoretical sampling is important, since the purpose of the research is to discover theories, not test
existing theories. Sampling means that cases (in this research, even participants) are chosen because they are suitable for illuminating relationships and logic among constructs (Eisenhardt and Graebner 2007).

In this research, as described in the previous section on CMD, the initial intention was to develop theory in the areas of research interest, whilst at the same time helping the company to improve their knowledge and practice in the area of usability and user experience. According to Eisenhardt (1989), a broad initial definition of the research question is important when building theory from case studies, but theory building research should begin as close as possible to the ideal of no theory under consideration, and with no hypotheses to test. In this research, the broad focus, which was given beforehand, was to perform empirical research concerning usability, but the details of the study were not given, and were allowed to develop over time, as the research progressed and the issues and questions crystallized.

A central idea in case study research is using cases as the basis from which to inductively develop theories. The theory is emergent, since it is situated in and developed by recognizing patterns of relationships among constructs within cases, and their underlying arguments (Eisenhardt and Graebner 2007). Building theory via case studies is applicable when there is little knowledge of a phenomenon, because in that case, theory building is not dependent on previous literature or empirical evidence, and can also lead to theory that is desirable when extant theory seems to be inadequate (Eisenhardt 1989). This has been relevant in this research because, for usability and the way it can be used in software engineering, and particularly in the case of UX, there remains much work to be done in formulating theory. This focus on building theory pointed the way to the area of grounded theory, and different approaches to grounded theory are discussed in more detail in the next sections of this chapter.

1.5.4 Grounded Theory

The focus on building theory led to a research methodology that is explicit in its ambition to build theory, with well-defined methods and guidelines for how to proceed. Early in the analysis process, inspiration was taken from Grounded Theory (GT) (see e.g. Glaser and Strauss 1967; Strauss and Corbin 1998). The use of GT continued throughout the research, but as discussed in the continuation of this section, the view taken of the analytical focus of GT changed during the research process, moving to a focus on Constructivist Grounded Theory (CGT). GT is a systematic approach to constructing theory, and is designed to encourage researchers’ persistent interaction with their data, while constantly engaging in their emerging analyses, so that data collection and analysis inform and affect one another, making the collected data more focused and the analysis more theoretical (Bryant and Charmaz 2007).

When describing the processes of GT, Charmaz (2010) writes that GT entails studying empirical events and experiences, and that a GT researcher follows up hunches and potential analytic ideas about them. Data is constructed through observations, interactions and through gathering material about the setting or topic at hand. The researcher studies the data at an early stage in the process, and then separates, sorts and synthesises the data.
through qualitative coding. This entails attaching labels to data to depict what the data is about, allowing comparison with other segments of data. Memos are written, which are preliminary analytic notes about the codes and comparisons, and ideas that occur regarding the data. By studying the data, comparing them, and writing memos, the researcher defines ideas that fit and interpret the data as tentative analytic categories. The researcher purposefully seeks data to close gaps and answer questions that inevitably appear when the data is analysed. Categories become more theoretical, as the analysis proceeds. Levels of abstraction are built directly from the data, and additional data is gathered to check and refine emerging categories. The work concludes in a “grounded theory” or abstract understanding of the case at hand.

Urquhart, Lehmann et al. (2010) discuss the distinctive characteristics of GT. First, the main purpose of the GT method is to build theory. Second, the researcher must ensure that prior knowledge of a field does not lead to them pre-formulating hypotheses that they then try to verify, as this would hinder the emergence of the theory. However, this stance is open to discussion, and this is discussed further in this section, which looks at the different approaches that are used in GT methodology. Third, the core process is concurrent data collection and comparison, which are the basis of analysis and conceptualisation. Finally, through a process of theoretical sampling, the researcher decides where to take the next sample. The fact that the research in this thesis has been influenced by action research can seem to be in conflict with the use of GT. For example, Baskerville and Pries-Heje (1999) discuss the situation whereby action research typically commences with a practical problem that suggests predefined concepts and categories, and the fact that it is too goal-directed to permit the full use of a comparative method such as grounded theory (GT), particularly as there is an interventionist strategy in AR. However, they maintain that, despite this, the rigour stipulated by GT is compatible with the AR approach.

GT arose as an attempt to move qualitative inquiry beyond simple descriptive studies, to a situation where explanatory theoretical frameworks could be built (Charmaz 2009). GT had its roots in ethnography, and although the connection has become weaker, and GT has become more of a systematic qualitative data analysis approach, ethnography is still suited to understand interaction; and interaction is central for most social science research (Timmermans and Tavory 2007). According to Gasson (2003) GT involves generating theory from data collected in investigations of “real-life” situations relevant to the research problems, and it is referred to as being grounded because contributions to knowledge are grounded in the data collected from one or more empirical studies, rather than being generated from existing theory. There was an early split between two of the founders of the tradition, and there are a number of different forms of GT, but all researchers can use the GT guidelines that exist, such as coding, memo-writing and sampling, and the actual comparative methods themselves are essentially neutral (Charmaz 2010). GT methods consist of systematic, flexible guidelines for the collection and analysis of qualitative data, and these guidelines are a set of principles and heuristics, rather than rules. The data is used to construct theories “grounded” in the data themselves (Charmaz 2010). GT can be used even in cases where the analysis process does not continue as far as developing theory, as it can be, and often is, used for the development of concepts (Urquhart 2007). Most of the
grounded theories that are built are substantive theories, which means that they address delimited problems in specific substantive areas. However, GT can reach across substantive areas and be used to generate formal theory (Charmaz 2010).

I have previously described the research approach as being qualitative. It has been a subject of debate whether GT can be described as positivist, interpretive or critical: however, GT is primarily a qualitative method for gathering and analysing data, and it is actually independent of the underlying epistemology; it is the researcher’s own ontological and epistemological position that decides how the data is coded and analysed, and how the theory is created (Urquhart, Lehmann et al. 2010). GT represents a number of methods and approaches, but they all begin with inductive logic, subject data to rigorous analysis, aim to develop theories, and value GT studies for informing policy and practice (Charmaz 2009). However, although creating theory is one of the primary aims of GT, Urquhart (2007) argues that it could be also be used to generate concepts about technology, even if it did not lead to the development of theory.

The characteristics and nature of the method actually means there is a risk of following GT guidelines slavishly, and Charmaz (2010) claims that for example following an axial coding matrix, or a set of theoretical codes can force the data into preconceived frameworks, thereby misdirecting the focus that is called for, and making the analysis process directed and prescriptive, undermining the value of the researcher’s own analysis. Rather than feeling that we must adhere to a “pure” version of GT, Urquhart (2007) suggests that we should instead make use of the inductive nature of GT, and use it in the way it has been used in, for example, information systems, where it has been adapted to the needs of research performed in that field. There have also been other cases where GT has been combined successfully with other methods, for example in the case of Baskerville and Pries-Heje, who combined GT concepts with AR (Baskerville and Pries-Heje 1999). This fits well with the pattern of this research, which has combined CMD and action research, case studies, and different approaches to GT, and has adapted the methodologies in order to deal with the varying types of issues and environments that have been part of the research.

This way of combining and mixing approaches has been discussed in connection with other studies. Baskerville and Pries-Heje, for example, maintain that GT methods can be suitable in AR since they allow the theory to emerge naturally in the unpredictable research environment, and GT defines units of analysis that can provide a suitable theory development technique for AR. GT methods are based on a reciprocal relation between data collection, analysis and theory formulation, which ties in well with the cyclical nature of AR. In AR studies, grounded theories are inductively discovered through the careful collection and analysis of data (Baskerville and Pries-Heje 1999). The use of GT methods in case studies is also well in accordance with the view of Eisenhardt and Graebner (2007) who, when discussing the way in which theories can be built from cases, state that theory building takes place in a process of recursive cycling between case data, emerging theory and extant literature, and that the data is central. The close connection to ethnography is also central. Timmermans and Tavory (2007) maintain that the ability to theorize is gained when researchers participate as social actors in the setting they work in. They state that the
iterative movement between data gathering and analysis that is typical of GT fits the socialisation process of ethnography better than any other research method.

According to Gasson (2003) an interpretive method such as GT needs methods that deal with context and process, and to gain a holistic view of a situation requires multiple approaches. The use of multiple approaches in GT can lead to much deeper insights than a mechanistic use of inductive coding, and the use of multiple methods can enrich the insights that are reached in data analysis. Thus, the mix of methods that has been used in this research work has been based on a pragmatic approach, which has combined the methods that have been appropriate to solve the problem at hand. This is also in accordance with the approach formulated by Baskerville and Pries-Heje (1999) who attempted to improve action research methods to improve their ability to contribute to IS research by merging some of the techniques of GT with the theory formulating steps in AR, giving a “theory rigorous” and greatly improved AR method.

After a long period of research together with UIQ, with its focus on CMD, the research environment, and also the style of the research, shifted. In the final stage in this research, together with ABB, an interview-based process was used. This research is described in detail in chapter six in this thesis. The main reason for the shift in style was that the research was within a new organization and was concerned with mapping a new area of cooperation. Therefore, there was not the same sort of access or acceptance as there was at UIQ. The ethnographic approach used at UIQ was not a relevant option in this early stage of the research, and was not appropriate for the kind of study that was planned. However, the research was still performed in close cooperation with the industrial partner and its representatives. I visited the site several times, together with my main PhD supervisor. We met the corporate research team, and discussed the operational focus of the research, and I was given a workplace and access to documents that described the company and its operations.

The work leading up to the change in research environment and style, together with deeper studies of the area of GT led me to the Constructivist Grounded Theory (CGT) approach (see e.g. Charmaz 2009), which was used in the continuation of my research. The next section of this introduction gives more details of CGT, and continues the discussion of the relationship between GT and CGT.

1.5.5 Constructivist Grounded Theory

The research with ABB was still based on issues regarding the use of usability and UX, but the research was not intended to progress as far as formulating any agenda for change. The intention was to study the situation, and based on that possibly formulate ideas for interventions that could remedy some of the problems that were found. However, it did not progress beyond an initial study of the situation at hand, and the processes that are in action in the organisation. Even here, GT was applicable, as it is particularly good for studying processes, and in the kind of research that concerns contexts where technological artefacts and people are juxtaposed, there are all kinds of possibilities for the interaction that takes place, and a real need for theories to be generated in this area (Urquhart 2007).
Thus, the study came to be interview-based. The six main sources of case-study data are documents, archival records, interviews, direct observations, participant observations and physical artefacts, but it is arguable that with respect to studies as an outside observer, interviews are the primary source of data. Interviews can be seen as an important data source, since they allow the researcher to step back, to examine in more detail the interpretations of the participants (Walsham 1995). In this research, my role was that of an outside observer, and this meant that the study could be performed as an interview based GT study.

Here I worked according to the Constructivist Grounded Theory (CGT) approach. This approach was particularly relevant here, since CGT is a return to earlier GT emphasis on examining processes, whereby the study of action is central, as is the creation of interpretive understandings of the data. The concept of process is defined as “unfolding temporal processes that may have identifiable markers with clear beginnings and endings and benchmarks in between. The temporal sequences are linked in a process and lead to change. Thus, single events become linked as part of a larger whole.” (Charmaz 2010, p. 10). In the research, interviews were performed to investigate different actors’ impressions of the action involved when working with usability within the organization. If it had been possible to study the actual processes and actions, another approach would have been taken, relying on another type of fieldwork, but in this situation, an interview-based study was the most appropriate approach.

CGT takes a different view of theory compared to what Charmaz defines as Objectivist GT (OGT). The concept of theory in GT contains both positivist and interpretive elements, where objectivist GT is part of the positivist tradition, and CGT is part of an interpretive tradition (Charmaz 2010). Charmaz (2009) describes the way in which CGT has its roots in pragmatism and relativism, and is based on an assumption that there are multiple realities and multiple perspectives on these realities. Data are mutually constructed through interaction, and are not separate from the viewer or the viewed. The representation of data, and its analysis, is thereby seen as problematic, relativistic, situational and partial. OGT, on the other hand, has its base in positivism, and assumes that data exists in the external world, and that this data can be discovered by a neutral but expert observer. Conceptualizations about the data arise from the data themselves, data are separated from the observer, and as such should be viewed without preconception. Given the discussion above, outlining the qualitative and interpretive stance taken in the research, and the focus on cooperation and participation, it is clear that the CGT approach is relevant in this research.

Constructivist GT is described as a revision of classic GT as formulated by Glaser and Strauss (1967). Charmaz (2009) states that CGT sees knowledge as socially produced, acknowledges multiple standpoints, and takes a reflexive standpoint towards actions, situations and participants in the field, and the way these are analysed. Constructivist GT claims that we produce knowledge by wrestling with empirical problems, and that the knowledge we have rests on social constructions. Constructivists enter participants’ worlds of meaning and action in a way that traditional grounded theorists do not, and constructivist GT entails a realization that conducting and writing research are not neutral acts (Charmaz
This correlates well to the CMD approach and focus on ethnography that has informed this research. Intervention was not an option in this study, nor was ethnography possible, but the focus on entering the world of the participants, and understanding the members’ point of view has influenced even this stage of the research.

The standpoint taken in CGT also correlates well with the description of qualitative research given by Denzin and Lincoln (2000), who maintain that qualitative research involves an interpretive, naturalistic approach to the world, where researchers study things in their natural settings, whilst attempting to make sense of phenomena in terms of the meaning that people bring to them. They describe it as a situated activity that locates the researcher in the world, consisting of a set of interpretive practices that make the world visible, whilst at the same time transforming the world into a series of representations, such as field notes, interviews, conversations, recordings and memos. According to Charmaz (2010), the constructivist approach means learning in what way the studied experience is part of larger and sometimes hidden networks, situations and relationships. This helps to make visible differences and distinctions between people, and hierarchies of power and communication that perpetuate these structures. Once again, although it was not possible to study the actions in their natural settings, the interviews have been performed with the intention of uncovering as much as possible about how the actions take place, and the structures and relationships that affected usability work in the organization.

When discussing the role of theory in interpretive studies, Walsham (1995) discusses Eisenhardt’s view of the role of theory in the context of organizational research, which maintains that there are three uses of theory: to guide the initial design and data collection; as part of the continuous process of data collection and analysis; as a final product of the research. Walsham (1995) compares this to the approach formulated by Glaser and Strauss (which Charmaz refers to as OGT) where they argued that the researcher should be primarily concerned with the discovery of theory from field data, and warn against the use of theory as an initial guide to designing the study and beginning data collection. Walsham’s (1995) description of the role of theory is closer to the view taken in CGT, where it is accepted that the researcher comes to the field with previous knowledge, but that by admitting and dealing with reflexivity, the problem is taken into account.

Data analysis also differs in the two approaches, and where OGT sees that objective data analysis is achieved through making generalizations more abstract through comparative analysis, thereby neutralizing researcher bias, CGT acknowledges that subjectivity not only enters data collection, but also data analysis, and argues for explication of how standpoints, situations and interactions may have influenced the results of analysis (Charmaz 2009). The overall process model in GT is reflexive, because it is centred upon making implicit the influences and inductive processes of the researcher (Gasson 2003). Reflexivity is a particularly important factor to take into account in the CGT paradigm. Walsham (1995), when discussing the nature and method of interpretive case studies, states that it is important that researchers have a view of their own role in the complex human process, and neither of the roles that the researcher may take on, the outside observer, or the participant or action researcher, can be viewed as that of an objective reporter. Charmaz (2009) also
points out the importance of taking reflexivity into explicit and continuous account, and distinguishes her mode of GT from earlier forms of social constructionism that viewed the research participants’ actions as constructed, but not the actions and situations of the researcher.

CGT adopts the inductive, comparative and open-ended approaches of earlier GT approaches, but also includes the abductive approach when checking and refining the developments of categories (Charmaz 2009). According to Reichertz, (2007) later approaches to GT exhibit two operations, in addition to coding and the development of new middle or long range theories. They also include the finding of similarity, or coding within known codes, and the development of the new, or creating new codes. This has its parallel in the distinction between qualitative induction and abduction, where the finding of things in common between the already known codes is induction, whereas the next step, adding something new to the data, that it does not already contain, that does not already exist, is abduction (Reichertz 2007). Abduction is a type of reasoning where, after the examination of the observed data, all possible explanations for these data are entertained, whereby hypotheses are formed and tested until the researcher reaches the most plausible interpretation of the observed data (Bryant and Charmaz 2007). This process is illustrated in this research by the process that occurred when analysing the interview results from ABB, where the field of research was thrown wide open, leading to expand the field of interest from software engineering, and the use of theories from HCI, to the inclusion of theories from other fields such as information systems, organizational studies, and political science, where, for example, institutions and institutionalism are discussed.

Abduction is a process of “assembling or discovering, on the basis of an interpretation of collected data, such combinations of features for which there is no appropriate explanation or rule in the store of knowledge that already exists” (Reichertz 2007, p. 219). This discovery causes a surprise, or even a shock, and necessitates the search for new explanations. Since no suitable types can be found, a new one must be invented or discovered by a mental process, and to do this requires leaving the conventional view of things. Thus, with the aid of intellectual effort, new orders or rules are discovered or invented. The process can take the form of something unintelligible being found in the data. On the basis of the mental design of a new rule, the underlying rule is discovered or invented, and it immediately becomes clear what the case is. The logical form of this process is abduction, where the creation of a new type is a creative outcome that engenders a new idea (Reichertz 2007). This is illustrated by the way in which the concept of the wicked problem was suddenly found, to understand and explain many of the factors that affected the situation within ABB. The concept of the wicked problem is not in itself new, but in this research, it became a clarifying concept, that had to be brought in from outside the normal discourses in the software engineering paradigm. Discovering this concept, and using it to focus upon the situation that was described in the interviews, opened up new models for explaining and understanding the situation, and for finding possible solutions.

Charmaz (2009) states that abduction underlies the iterative process of moving backwards and forwards between the data and conceptualization, which is what actually characterizes
all GT. According to this, abductive reasoning arises from experience, leads to creative but logical inferences, and creates hypotheses for testing these inferences, to arrive at plausible theoretical explanations. Charmaz (2010) places the CGT approach in contrast to the way that Glaser and Strauss, in early work on GT, talk about discovering theory. In their approach there is an assumption that the theory emerges from data separate from the observer, whereas CGT assumes that neither data nor theories are discovered. Instead, since we are all a part of the world we study and the data we collect, our theories are constructed through our past and present interactions with people, practices and perspectives. Thus, when I as a researcher have interacted with the other participants in the research, whether it be in an action research approach, or simply as an interviewer, I must constantly be conscious of the fact that I am not only affecting the situation by the very fact of my presence, but that the results that are arrived at are also a result of my own experiences, knowledge and perspectives.

The mechanisms of reflexivity necessarily means that the role played by the researcher has a profound effect on the results of the analysis, and this must be taken into account when considering the results. This leads to a discussion regarding what has traditionally been seen as checks on validity and reliability.

1.5.6 Judging the quality of qualitative research

Questions of scientific quality are important, and what is regarded as quality differs from field to field. The sections above have already discussed differences between positivist and interpretive approaches, and discussed which research paradigm GT can be found within. GT is often criticized for a lack of rigour, and this is because positivist concepts of rigour are applied to research that derives from an interpretive worldview (Gasson 2003). Charmaz (2009) discusses the fact that, if we accept the notion of a multiplicity of perspectives and multiple realities, we are forced to construct layered analyses and must attend to varied ways by which we and our participants construct meaning. Accepting this necessitates accepting the fact that the traditional objectivist views of what constitutes good research cannot be applied here. There are differences in different approaches, and the interpretive approach, which was followed here, can be distinguished from the positivist approach. Although the intention of interpretive work is not to generate truth or social laws, this does not mean that interpretive work is not generalizable, although the nature of the generalization in the two traditions differs (Walsham 1995).

When looking at the evaluation of the results of qualitative or flexible research, (Robson 2011) discusses non-positivist research in terms of it being “flexible design” research, it cannot be judged by the yardsticks of the positivist tradition of science. The following examines some of the criteria that are used when discussing judgements of the quality of research results, as described by Robson (2011) for flexible research in general, Yin (2003) regarding case studies, and by Gasson (2003) and Lincoln and Guba (1985) regarding GT. In the case of Gasson and Lincoln and Guba, some of the factors that they include are common, and in these cases they are discussed together.

There are alternative ways of viewing what is good qualitative or flexible research. There are a number of characteristics of good flexible design that should be adhered to when
designing and performing the study. The characteristics of a good flexible design are (Robson 2011, p. 132):

- Rigorous data collection, where multiple data collection techniques are used
- The study is designed with the assumptions and characteristics of a qualitative approach to research
- The study is informed by an existing strategy, where the researcher identifies, studies and employs one or more traditions of enquiry. The tradition need not be pure, and procedures can be brought together from several traditions.
- The study begins with a problem or idea that the researcher seeks to understand.
- The researcher uses detailed methods for the collection, analysis and writing
- Analysis is performed on several layers
- The report is written clearly and in an engaging fashion, allowing the reader to experience “being there”

When looking at this work on a general level, these characteristics have informed the design and performance of the research. As previously discussed, the approach was qualitative, and multiple data collection methods have been used, such as observations, participant observations, document and literature studies, workshops and interviews. The study was informed by the strategies from CMD, action research, case studies, and grounded theory, and although they have not been used in their pure form at all stages in the research, the combination of methods has been designed to take into account and deal with the specifics of the research environment and the questions at hand. The research began with a broad intention to study, understand and improve usability work in organizations, and over time has crystallised into a number of research challenges, with connected research questions. The material has been collected and analysed according to the research approaches mentioned above, in particular GT. Once again, the approach has not always been pure, but has been adapted to the situation at hand, and has contributed to the process of analysis on several levels. The main factor that differs, as mentioned previously in this chapter, is the fact that the presentation of the research has not focused on allowing the reader to experience “being there”, as is typical for a e.g. traditional ethnography, but has been adapted to the expectations of the reader.

Looking more closely at the evaluation of case studies, Yin states there are several criteria that have commonly been used to establish the quality of empirical social research, including case study research. These criteria are important during the design phase of the study, but should also be applied during the conduct of the case study. They deal with construct validity, internal validity, external validity and reliability (Yin 2003, pp. 35-39).

The concept of construct validity deals with establishing the correctness of the concepts being studied. Construct validity can be strengthened through the use of multiple sources of information, by ensuring a chain of evidence, and by having the key participants review the results of the case study, in a process of member checking. As already discussed above, many sources of information were used, and the chain of evidence has been maintained in a
“study database” consisting mainly of a logbook containing an account of the study, including records of meetings, transcriptions, observation notes, research codes, and reflections on the research and emerging theories. The results of the study have been subjected to member checking, partly through presenting the results to those who have been involved in the study, and partly through a cooperative writing process where some of the participants have been co-authors in publications.

In exploratory case studies, where the investigator tries to make sense of how one event leads to another, *internal validity* is especially important, as the intention is to show that events are causal, and that no other events have caused the change. It is only by doing this that threats to internal validity are dealt with. It is hard to identify specific analytic tactics to ensure internal validity, but it can be done through pattern matching, explanation building, addressing rival explanations, and using logic models. In this research, which has been a mix of explanatory and exploratory studies, emphasis has been placed on analysing the data according to the principles of grounded theory, thus supplying a reliable toolkit for the analysis process.

*External validity* concerns the ability to know whether the results of a case study are generalizable to other contexts or cases. Single case studies have been seen as a poor basis for generalization, and this has been seen as a serious hinder to doing case studies. This is, however, based on a faulty reasoning. Critics contrast the situation to survey research, where samples readily generalise to a larger population, whereas in a case study, researcher tries to generalise a set of results to a wider theory. In the research presented in this thesis, the individual chapters present the theories that have been suggested by the data. Some of the theories are built are substantive theories, which means that they address delimited problems in specific substantive areas, but in particular the theory presented in the final chapter appears to have a greater relevance, and can be regarded as closer to a formal theory.

*Reliability* deals with the possibility of replicating a study. The general way to ensure reliability is to conduct the study so that someone else could repeat the procedures and arrive at the same result (Yin 2003). It should be possible for a later investigator to follow the same procedures as the first investigator, and arrive at the same findings and conclusions. Ensuring reliability means minimizing errors and bias in a study. Documenting the procedures followed in research work is important, and this can be done by maintaining a case study protocol, or by developing a case study database. The case study protocol is intended to guide the investigator in carrying out the data collection. It contains both the instrument and the procedures and general rules for data collection. It should contain an overview of the project, the field procedures, case study questions, and a guide for the case study report (Yin 2003, p. 69). As already mentioned, this has been a central part of this research approach. However, as already mentioned above in connection with GT studies, and discussed further below, studies are performed in a context, which can never be reconstructed. Furthermore, the researcher has to take into account their own role in the process. We must be aware that there is a degree of subjectivity involved in the research process, even though this is dealt with through reflexive self-awareness, to
minimize the impact of the researcher’s own role. Thus, we must be aware that the results we reach are not the single verifiable truth, and that it is impossible to replicate a study involving individuals, situations and contexts.

Dealing with the quality of a case study was particularly relevant during the earlier phases of the research, where the research was primarily informed by the case study approach. As previously mentioned, as the research progressed, the methodological focus changed from a case study approach to become more influenced by GT, particularly CGT, so the discussion now moves to rigour in GT. Even though there are similarities, the discussion takes a somewhat different form.

Gasson (2003), when discussing rigour in GT research, maintains that we must have alternative notions of rigour that reflect the same sorts of quality notions as those used in positivist research. Where positivist research discusses results on the basis of objectivity, reliability, internal and external validity, interpretive research must be judged on the basis of Confirmability, Transferability, Dependability/Auditability, and Internal consistency. Lincoln and Guba (1985) also present principles for the evaluation of naturalistic research, for establishing trustworthiness. Their concepts are: confirmability, transferability, dependability, and credibility. As can be seen, there is some overlap in the terminology. Where there is an overlap, the different authors’ criteria are discussed together.

Concerning confirmability, which is common to both discussions, Gasson (2003) states that the main issue is that findings should represent the situation being researched, rather than the beliefs of the researcher. Interpretive GT research focuses on reflexive self-awareness, in an attempt to guard against implicit influences and prejudices. The only way to deal with subjectivity is through a constant process of reflexivity, where we as researchers are aware of ourselves as part of a social context, affecting the phenomena under observation, and awareness that we apply biases and prejudices to the collection, analysis and interpretation of data. Only through making our assumptions and frameworks explicit can we minimize the effect of our own role. When discussing confirmability, Lincoln and Guba (1985) also discuss reflexivity, maintaining that it relates to a systematic treatment of theory building, in particular the influence of the researcher, at all stages of the research. They state that a technique for dealing with this is for the researcher to write a reflexive journal, to hold regular entries made during the research process, describing methodological choices, the logistics of the study, and reflections upon what is happening in the study, and why. This is in line with the technique that has already been discussed, concerning the case study database and logbook, where the research process is recorded in detail, including reflections about the issues that arose in the study, and thoughts about emerging and developing theories, and the way in which they change over time.

Gasson (2003) states that the core issue of transferability is to which extent a researcher can claim general application of a theory, and this depends on similarities or differences in the context where the theory is to be applied. According to Lincoln and Guba (1985), transferability concerns the ability to show that the findings are applicable in other contexts, which entails providing sufficient data to make such judgements possible. It is not the researcher’s task to provide an index of transferability, but they must provide enough
data to allow potential applicants to make the comparison themselves. One technique for establishing transferability is to provide thick descriptions. As previously mentioned, the discourse where the research has been published made it difficult to provide thick descriptions. These descriptions are available, in the form of the study database, with all of the associated material and records of the process. However, when publishing this research, it has been found necessary to adapt the presentation to the audience, who do not expect results to be presented as thick descriptions. Thus, it has been necessary to present the results in a way that shows how they relate to the case in question, and then discuss whether they can be generalised to other similar situations.

Regarding dependability/auditability, Gasson (2003) writes that the main issue is that the way in which a study is conducted should be consistent, and we need to show that we have used clear and repeatable procedures. In the interpretive view, it is seen that reliability cannot be achieved, since it is recognized that we apply and interpret social constructions of reality rather than an objective reality. Therefore, it is important to ensure dependability by making explicit the processes whereby findings are derived. Achieving dependability and auditability can be done through defining the procedures whereby data are collected and analysed, and ensuring that these processes are recorded so that others can understand them. It is important to maintain a research journal where records are kept of all data collection and analysis. Lincoln and Guba (1985) state that dependability, or consistency, entails showing that the findings are consistent and could be repeated, and includes taking into account factors of instability in the individuals, the context and the design. Dependability is closely connected to credibility. Techniques for establishing dependability are closely connected to those for credibility, discussed below, and include overlapping methods, and providing an inquiry audit, allowing an examination of the process and the product, and how well they fit. Once again, the study database contains the records of the study and its procedures. These procedures have been subjected to peer-review by research colleagues, in workshops and meetings where the process and progress of the research has been reported and discussed. Having said this, it is still important to remember that, even though the findings of a study can be judged to be consistent, it will still be difficult to achieve repeatability, since the mechanisms of reflexivity, and the fact that, as already discussed, our theories are constructed through our past and present interactions with people, practices and perspectives.

Gasson (2003) includes internal consistency, and to achieve this, it is necessary to explain from which data and how the theoretical results are derived, and whose perspectives these constructs reflect. The source data and the process whereby it is collected should be described in enough detail to demonstrate a fit between the theory and the data. This is closely related to what Lincoln and Guba (1985) call credibility, also known as the truth value, which they express in more detail. It expresses the degree of confidence in the “truth” of the findings. Here, the findings of the study should be credible to the constructors of the original multiple realities that have been the subject of the study. The techniques for establishing credibility are: Prolonged Engagement, Persistent Observation, Triangulation, Peer debriefing, Negative case analysis, Referential adequacy, and member-checking. Prolonged engagement is to provide scope, and persistent observation is to provide depth.
Triangulation is done through using different sources, methods, investigators or theories. In peer debriefing the emerging theories are tested with others. In negative case analysis, the data and the emerging theories are revisited and revised, until the fit between data and theory is clear. Referential adequacy means archiving the data for future cross analysis. In member checking, the details of the data and the analysis are cross-checked with the groups where the data are collected. In the research at UIQ, all of these techniques have been taken in use, but in the final phase, at ABB, there was no opportunity for prolonged engagement or persistent observation. However, given the fact that the study at ABB involved a wide cross section of stakeholders, it still provides a vivid illustration of the case at hand.

Finally, Charmaz (2010) presents a number of criteria for GT, regarding credibility, originality, resonance and usefulness, and maintains that a strong combination of originality and credibility increases resonance, and the value of the contribution. In GT, any claim to making a scholarly contribution requires a careful study of literature, including literature that lies beyond disciplinary boundaries, and that the GT should be clearly positioned. As far as the study at UIQ is concerned, the length and depth of the study, and the intensity of the research cooperation, combine to give the results credibility. The fact that there was such a cross section of stakeholders in a young company working in the expanding area of interactive technology for smartphones, and that the data has been analysed from so many different angles lends originality to the study. Taken together, this creates resonance. The study at ABB, whilst not achieving the same depth of penetration, is still a vivid illustration of the problems experienced in a global company adapting to new situations regarding technology and use. Usefulness is lent by the fact that this research, by examining usability and UX in this way, and bringing to software engineering knowledge and theories from diverse fields, contributes to solving some of the problems still associated with usability and UX work in software engineering environments.

A more detailed description of the methods and methodology used to address each research challenge, and the way in which the research has been performed according to the principles discussed above, is found in the methods sections of the separate chapters that are included in this thesis. However, in general it can be said that the combination of field work and analysis methods that have been used in the research, complemented by the use of the case study database, the logbook, and by the mechanisms of member checking and peer-reviewing, have all contributed to the quality of this research.

Now, having looked at the lenses and focuses that have been central to the research, the methods that have been used to gather and analyse the data, and the way in which it has been ensured that the methods have been sound, the following section presents a summary of the actual results of the research.

### 1.6 Results

The research presented in this thesis revolves around five research challenges that have arisen during two industrial research cooperation projects. This section summarises the results of the research, as found in the articles that are included as chapter two to chapter
six in this thesis. The results are presented in relation to the research challenges, which are outlined in section 1.3.

1.6.1 Challenge One
The first research challenge (C1, see figure 1.3) dealt with finding appropriate ways of analysing and using the results of usability testing. This is dealt with in chapter two: *Examining Correlations in Usability Data to Effectivize Usability Testing.*

This study examined whether there is a simple measurement to express usability, and whether it is possible to streamline the discovery of problematic use cases. To do this, an analysis was performed of the correlation between test metrics for efficiency, effectiveness and satisfaction, different factors of usability, that were collected during the testing process.

This was done in order to see whether there are correlations that would allow the discovery of usability problems on the basis of a simple metric. To satisfy industrial needs, this metric should preferably be one that can easily be measured without the presence of the test leader. Based on this situation, two research questions were formulated.

The research in this chapter deals firstly with examining the correlation between measurements of effectiveness, efficiency and satisfaction, which are all different aspects of usability. Furthermore, it examines whether a statistical analysis of task-completion time allows the discovery of problematic use cases. This was done in an attempt to discover whether there are correlations that allow the discovery of usability problems on the basis of a simple metric, which could easily be measured without requiring the presence of a usability tester.

The first part of the research is based on the idea that there may be a sufficiently strong correlation between the three factors of usability that measuring one of them would give a reliable indication of the usability of a mobile phone. The second part of the research is based on the hypothesis that there is an expected distribution of completion times for a given use case and that deviations from goodness of fit indicate user problems.

The study showed that the correlations between efficiency, effectiveness and satisfaction were not sufficiently strong to permit basing usability evaluations on the basis of one single metric, which means that it is important to measure all three factors.

However, it was found that it may be possible to discover potentially problematic use cases by analysing the distribution of time taken to complete use cases. This would allow the collection of data that could indicate which use cases are problematic for the users, allowing companies to concentrate scarce testing resources where they are most needed.

Data collection could even be performed remotely, allowing the data to be communicated to the development organisation. This would allow the collection and use of data from a real-world use context, rather than performing testing in an planned test situation, together with a test leader who steers and observes the testing, or in an artificial test situation, such as in a usability laboratory.

However, it was still found that the role of the test leader is important, since it is not possible to identify the character of a problem simply by looking at completion times.
Through observing and analysing the performance of use cases, the test leader is able to develop knowledge of the causes of problems that user experience, and can mediate this knowledge to designers and developers when developing design suggestions and solutions, and even to other stakeholders who have an interest in the usability and UX of the product.

1.6.2 Challenge Two

The second research challenge (C2, see figure 1.3) involves dealing with the problems associated with understanding and using the definitions and concepts of UX. It is dealt with in chapter three: *The Concept of UX Depends on the Intellectual Landscape of HCI Paradigms.*

This challenge involves examining and discussing the importance of definitions and the way in which UX is defined and treated in the research community, and the problems associated with UX measurement, particularly for the field of software engineering.

The discussion in this chapter was grounded in the fact that there are a large number of definitions of UX, but that little discussion has taken place regarding the theoretical groundings of the definitions that exist. The chapter shows that there must be a richer debate regarding the underlying theoretical foundation applied when defining and discussing UX, particularly to allow the use of UX in a software engineering context.

By presenting a representative selection of UX definitions, the chapter illustrates the diversity of UX definitions, and the growing richness of UX concepts. The definitions are discussed in terms of three paradigms that are identified as existing in the field of HCI, each of which has a unique understanding of what science is, and how to go about being scientific (Harrison, Tatar et al. 2007). In these paradigms, certain phenomena are seen as central, whilst others are seen as marginal. The main point in this chapter is that these centres and margins are central for an understanding of UX as a concept. Concepts and definitions are seen as bearers of different meanings, depending on the paradigm they are grounded in. A definition is dependent on the paradigm that influences its formulation. Furthermore, the interpretation given to a definition also reflects the paradigm that influences the interpreter. Each paradigm opens up for different interpretations of a definition. Without traceability and discussion about the paradigms that have influenced definitions of UX, we can end up in confusion, or with a lack of powerful contributions.

The chapter maintains that definitions have positive qualities. They can be used as boundary objects between researchers with different ontologies, belonging to different research paradigms. They can ease discussions, and the placement of work and the results of work. They can ease mutual learning and knowledge sharing between actors from different paradigms and fields, and they could contribute as learning vehicles for beginners in the area. However, it also maintains that for a definition to be truly useful, it is important that it be positioned in the researcher’s background and in the paradigm where they belong. We must look to the origins of the underlying theoretical foundation of a definition of UX and define the origins of it in order to understand and use it.

Furthermore, the chapter points out reasons why there are difficulties introducing UX into software engineering. HCI is the primary academic community when it comes to research
within UX, whereas Software Engineering (SE), which has slowly incorporated usability as a quality and product measurement, has barely touched upon the topic of UX. Still, UX will become increasingly important even for SE in order to measure quality on a broader framework, including customer satisfaction and business needs. Until we reach some kind of agreement regarding the meaning and use of definitions, then we cannot reach a stage where UX becomes a meaningful tool in the software engineering toolbox, allowing us to measure and make use of this important and highly relevant concept.

1.6.3 Challenge Three

The third research challenge (C3, see figure 1.3) dealt with finding ways to use the results of usability and UX testing to satisfy the different information needs of different stakeholders in organisations. This is dealt with in chapter four: Satisfying Stakeholders’ needs – Balancing Agile and Formal Usability Test Results.

The results are based on two separate but related studies, and on a number of research questions, all centred on the best way to capture usability data through usability testing, and the best ways to represent that data.

The research involved finding factors that would allow us to identify the different types of stakeholders present in the organization who had an interest in usability data. Furthermore, it dealt with discovering the types of testing methods and result presentations that were most appropriate for these different groups, which are involved at different levels and in different stages of the product design and development process.

The research was also aimed at studying the importance of the test leader when presenting the results of the testing, to see if test results could be presented in a meaningful way without the test leader being present, or whether presentations could be sufficient in themselves, without the test leader’s presence to explain the results.

It was found that the participants in the study could be divided into two separate groups of stakeholders, called Designers and Product Owners, who were found to have interests in different elements of the usability test data. We found that they were on two sides of a spectrum, regarding the level of detail included in the results, and the time scale that the results concerned, where designers needed rapid feedback, and detailed descriptions, whilst product owners had a longer perspective, and needed more overarching information on the product. To satisfy the needs of the different groups, it was found that there must be a balance between formalism and agility, and that both of these elements are important. It was shown that usability testing must be flexible to accommodate the needs of many different stakeholders.

The research also showed that the presence of the test leader was important. Some kinds of presentations were sufficient in themselves, but other types of presentation required that the test leader was present to explain the context, and the results and background factors that led to them. The findings illustrated the importance of having a specialised tester, who brings in the knowledge and experience of end-users. It was also shown how the test leader works as a proxy for the users in a mass market, where PD methods are lacking today.
Beyond these factors, the research confirmed how learning and drawing from experience may be as important as taking a rational approach to testing, and confirms theories that there are good organisational reasons for practices that might otherwise be considered to be bad practice. It shows that the reasons for not adopting what is otherwise regarded as “best practice” are that the testing is oriented to meeting the actual organisational needs.

1.6.4 Challenge Four
The fourth research challenge (C4, see figure 1.3) dealt with finding ways in which research in usability and UX can contribute to other process improvement methods, and can in turn be improved by studying other process improvement methods. This is dealt with in chapter five: SPI success factors within product usability evaluation.

This challenge involved examining how the results of usability testing could be used to improved design and development processes, and involved seeing if there are similarities and differences between the principles and practice of usability work and the principles for Software Process Improvement (SPI). The chapter reports eight years’ experience of product-related usability testing and evaluation is compared to the principles of software process improvement (SPI).

The results show that there are many similarities when comparing SPI and usability testing. The new knowledge that is presented could affect ways of thinking about both SPI and usability testing.

Concrete lessons for usability work, which may also have implications and be relevant for SPI, were:

- It is important to ensure the participation of specialists who are well versed in the organisation;
- it is important to adopt a flexible approach to the design and performance of the programme, to adapt to environmental changes;
- it is important to adapt presentation needs to the needs of different types of stakeholders;
- it is important to find suitable people to collect metrics, and to analyse and present the results of the testing – to be an advocate of the user perspective.

As far as SPI is concerned, it was found that:

- Flexibility is important in a changing world, which means that we must be open to re-formulating goals during an on-going improvement process, rather that assuming that the goals that were formulated at the beginning of the process are stable, and that fulfilling them is enough;
- It is important to include knowledge that is based on intuition and experience, rather than formally defined knowledge, and in this case we found that the test leader was a central figure who was most suited to communicate this knowledge to other stakeholders. This confirms the findings of our previous work, where the test leader is found to be a central figure and mediator of knowledge;
• It is important to have both agile and formal results, to satisfy the needs of different stakeholders. Formal results support agile results, and ensure that the test results are accountable.

In conclusion, it was found that usability testing was a useful vehicle that was in accordance with the principles of SPI, and that is also had a clear business orientation. There were parallels between usability testing and SPI, but they are not equivalent. There is a need for SPI, but there is also a need for usability process improvement, taking into account international standards.

1.6.5 Challenge Five

The fifth and final research challenge (C5, see figure 1.3) concerns understanding the way in which organizational factors affect how usability work is discussed and performed in a large organisation, and the way in which the history and structure of the organization affect what gets done within the organization. This is dealt with in chapter six: Identifying organizational barriers– a case study of usability work when developing software in the automation industry.

Here, the results show that working with usability Key Performance Indicators (KPIs) in the kind of organization that was studied is difficult, and that achieving a situation where this is natural would require what would probably be far reaching cultural and organizational changes that would be difficult to manage and achieve. However, the results suggest that, to be successful in the marketplace with the kinds of products that the company are hoping to produce and sell, it is necessary to make these changes.

The organization we studied is well-established and successful. The company has succeeded by using an engineering strategy in product design and development, but now experiences difficulties when operating in a changing situation and marketplace, where new demands are placed on the software that is becoming a more and more important integral part of their products. The context is influenced by the history and traditions of the company and by the organizational structures that are rooted in strategies that have functioned well in the past. However, changes in technology and user demands for usable products are forcing changes in the products that are produced, and the way the products are used. Changes in the marketplace mean that the customers no longer have a strong engineering background, and the customers are now parts of smaller networks of actors without a traditional engineering background. Thus, the existing structures are being challenged. In order to produce competitive products, it becomes necessary to change the view of what constitutes quality, and how this quality can be achieved. It is suggested that the situation that was observed is also representative of other large organizations, where changes in technology are forcing changes in assumptions about what constitutes quality, and how products should be produced.

Some of the problems that are illustrated in this study are behavioural; based on organizational factors that have led to conflicts, breakdowns and barriers. The problems originated on several levels of the organization, and they affected several different parts of
the organisation. Due to this, they found that they had problems when deciding how to
define and use usability KPIs, which the organisation appreciates is an important part of
maintaining a leading position in the marketplace.

The results suggest that addressing the problems that lead to the difficulties when working
with usability KPIs needs a holistic view of the structures and processes involved, even in
the everyday work of the end-users, and this means placing the context of the end-users in
the foreground. However, this view is difficult to achieve, given the situation that exists in
the described organization. The organizational problems, in conjunction with the
complexity of the marketplace, and insufficient knowledge of the user domain and users’
needs, mean that a wicked problem exists. When addressing this problem, the traditional
engineering approaches that have been used when developing products are not sufficient.
This is because the problems are not based in the technology, but in behavioural problems.

However, the behavioural problems that exist make it difficult to address the problems.
Solving the problems would require that all stakeholders bring their contributions to the
design process, taking context into account. However, because of the many breakdowns
that occur and the barriers that exist, it is hard to get input from key stakeholders. This
makes it difficult to include context, reach consensus, and produce design solutions. To
remedy this would require far reaching changes in the organizational culture, and although
this is a long and difficult process, we believe that it is only if these changes are made that
usability will become a natural part of the design and development processes, where
incentives are created that make it possible to work with usability as an integral part of the
organization, and where it is possible to work with generally agreed upon usability KPIs.

This change must have backing from management, and should involve staff and their
representatives from all levels of the organization. Management must devote resources to
enable the change processes, and the change processes must include staff from all levels of
the organization, working not solely with the change processes, but also with the creation
of KPIs. The focus given to KPIs must reinforce the importance given to usability, and
there should be incentives for meeting the goals set in the KPIs. The requisite knowledge to
address some of the problems is distributed between several stakeholders, including end-
users, and all of these stakeholders must participate, cooperate and collaborate when
finding solutions. Placing context in the foreground requires the active participation of end-
users. To create relevant KPIs, it is important to have knowledge of the users and how they
use the products in their work life, and this requires mechanisms for user participation and
user research. These methods are today not well established in the engineering context that
was studied in this case.

The analysis shows that the most important factor to work with is to find ways of including
actual end-users, with the goal of making UCD an integral part of the design and
development processes. This could be the start of the process of change that is needed to
work effectively with usability KPIs. Furthermore, it is found that UCD is vitally important
when designing and developing competitive products in a changing marketplace.
1.7 Contribution to Quality, Organizations and Institutions

In this section, the overall contribution of this research will be discussed, in relation to the three lenses: quality, organizations, and institutions, with their particular areas of focus: usability and UX, participatory design, user-centred design, and wicked problems.

1.7.1 Lens: Quality

Two research challenges were outlined when dealing with the area of quality, with its focus upon two aspects of product quality: usability and UX. The first research challenge dealt with finding appropriate ways of analysing and using the results of usability testing. The second research challenge involved dealing with the problems associated with understanding and using the definitions and concepts of UX.

As discussed in the overview of usability (section 1.4.2), much of the work within usability and UX has been performed in areas such as Human-Computer Interaction and design science, whereas Software Engineering (SE) has found it difficult to incorporate usability and UX in the SE discourse. With its focus on usability and UX, and its use of theories taken from diverse research fields, this research contributes to bridging the gap, and bringing new theories to the field of SE.

Firstly, the research demonstrates that although measuring usability is important, it is not sufficient. The results of the measurement must be interpreted and explained by a usability expert, who must also observe how a product is used, in order to understand the problems that users have. Secondly, it is important to measure quality on the basis of other factors than faults and failures. This research contributes to one step on the path to achieving this, by developing ways in which usability and (to some degree) UX can be measured. On the basis of these measurements, information can be presented to different stakeholders, thereby affecting product design and development processes. Furthermore, the research supports the existing view that quality cannot be shown by one single measure. It demonstrates that it is necessary to work with several different types of measurement in order to gain a true picture of product usability. This is discussed further in the next section, 1.7.2, in connection with usability and UX.

Stakeholders have different views of quality and different acceptable limits for it, and this research contributes to the discussion of this subject. The research results illustrate the importance of studying the different information needs of stakeholders, and point out the importance of adapting the presentation of test results to these needs. Through this adaption, it is possible to contribute to fulfilling these different needs, providing stakeholders with illustrations of quality at a given point in time, and allowing discussions between different groups, since the results are based on the same test data, but adapted to particular needs.

By studying metrics for efficiency, effectiveness and satisfaction, this research demonstrates that the correlations between these factors are not sufficiently strong to permit measuring usability on the basis of a single metric. Instead, it shows the importance of measuring all three factors. This is in relation to previous studies, which have reached contradictory conclusions regarding the correlations between these factors. This research
did however reveal that discovering potentially problematic use cases may be done by collecting one metric. This approach involves analysing the distribution of the time taken to complete specific use cases. This would simplify the collection of data that can indicate which use cases are problematic for the users. Companies could thereby concentrate scarce testing resources in an efficient manner. Data collection could be performed remotely, and the data then communicated to the development organisation. This would allow the collection and use of data from a real-world use context, rather than from an artificial test situation.

The research contributes by demonstrating the importance of the test leader, as a person who, in their role as observer, gains deep knowledge about the way in which technology is used, and can mediate this knowledge to other stakeholders. For example, the research results demonstrate that, when presenting the results of usability testing for different stakeholders, some kinds of presentations were sufficient in themselves, whilst other types of presentation required that the test leader was present to explain the context, and the results and background factors that led to them.

The study at ABB illustrates the difficulties of driving usability work, even though it has long been known that this work is important, and it illustrates how institutions hinder the establishment of “usability champions”. The study found stakeholders who wanted a mandate to push the usability issues that they saw as being important, but that organizational factors made this impossible. This confirms other studies that point out the need for “usability champions”, who should have the mandate to decide on usability matters, alleviating the risk that the skilful and experienced work of the usability designers is ignored in later phases of the development project.

When developing the UTUM test, much effort was applied to finding ways of visualizing results in appropriate forms, and this research contributes to developing ways of visualizing usability and UX, which is becoming more important for organizations. For example, section 1.4.1 discussed the way in which quality and function are intertwined, and this research shows how the Kano model, with its analysis of functions and the way they affect satisfaction, can be used to discuss usability and UX. The use of the Kano model influenced the way in which results UTUM results are presented, contributing to creating new ways of visualizing UX.

The research suggests that it is important to look at the theoretical foundations of UX definitions. This is in order to reveal the different perspectives that are reflected in the multitude of definitions, thereby leading to an understanding of the definitions themselves. Furthermore, it shows how achieving this is important for the field of SE. Until UX can be understood as a concept, which includes understanding the meaning of the definitions, it will be impossible to measure UX, which is what the field of SE needs to be able to incorporate UX in the SE toolbox.

1.7.2 Lens: Organizations

Two research challenges were outlined when dealing with the area of organizations, focusing upon two areas: participatory design (PD) and user-centred design (UCD). The
first of these research challenges deals with finding ways to use the results of usability and UX testing to satisfy the different information needs of different stakeholders in organisations. The second of these research challenges deals with finding ways in which research in usability and UX can contribute to other process improvement methods, and can in turn be improved by studying other process improvement methods.

This research provides a clear illustration of the organizational barriers that stop usability and UX work from being done. It demonstrates how the barriers stop organizations from finding out who the real users are, and how they hinder them from engaging the users in design and development processes. The research is based on long-term cooperation at UIQ, where organizational participation at many levels affected the research, and at ABB, where the diversity of the participants contributes to the depth of the study. The research has included shop-floor practitioners, together with other stakeholders, from management, marketing, and product planning. Thus it provides a vivid picture of the organizational factors that affect the way in which usability and UX work gets done, or is hindered. According to the review in 1.4.1, it is important to associate innovation with the context where it is embedded, and to consider the rational decisions as well as the social forces of the processes. Furthermore, it is stated that relevant theories are probably grounded in social and organizational themes. The research in this thesis contributes in these areas.

Section 1.4.2 discusses how resistance to change is caused by organizational constraints such as lack of management support, or lack of time to negotiate or communicate. The research in this thesis strengthens these findings, showing the presence of these mechanisms, particularly at ABB, but contributes by extending them to show how these factors also affect what gets done, or gets neglected, and by illustrating the kinds of issues that are prioritized or de-prioritized in projects.

At UIQ, this research discovered that stakeholders with an interest in usability data could be divided into two different groups, with different types of information needs on different time scales. These groups were designated as Designers and Product Owners, and their information needs were shown to be on a scale between agility and plan-driven, with clear parallels to the agile movement. Furthermore, the research results illustrate how learning and drawing from experience may be as important as taking a rational approach to testing, and confirms earlier theories that there are good organisational reasons for practices that might otherwise be considered to be bad practice. It shows that the reasons for not adopting what is otherwise regarded as “best practice” are that the testing is oriented to meeting the actual organisational needs of different stakeholders.

This research contributes by showing the role that usability testing can play in improving both products and processes, and by showing the way in which usability can improve usability processes, in the same way as Software Process Improvement (SPI) has been found to improve other SE processes. Here once again, the research shows that the test leader is a central figure, collecting metrics, analysing and communicating the results of the testing, and advocating the user perspective. Furthermore, the analysis of the use of the usability test results also suggests ways in which SPI could be adapted, to take into account the rapid changes that are occurring in the marketplace.
Regarding PD, the bulk of this research has taken place in an environment where changes in technology are changing the concepts of work, the workplace, and work practices, from the traditional workplace, to all of the places where people perform tasks, and all of the tasks that they perform, including communication and entertainment. This research confirms the problems that are already known, concerning PD in a mass market. However, it contributes by finding ways in which the test leader can contribute to bringing in the point of view of the users to the design and development process, even when direct participation is not feasible. Thus, the test leader works as a proxy for the users in the mass market, where PD methods are lacking today. The testing has thereby been done with real users using realistic use cases, and the test leader has taken the knowledge of the users to the design and development context, so this is one step on the road. In addition to this, this research has shown that the character of the test leader is important, in the same way as the character of a flexible researcher is important and, experience from this study shows that the best testers can come from a social work or nursing background, rather than from a technical background.

The user is still seen as the “great unknown”, but in spite of this, user involvement has not been a particular research topic in SE. Once again, this research makes a contribution by illustrating the importance of having a specialised tester, who brings in the knowledge and experience of end-users to the design and development processes. The test leader is necessary as an advocate of the user perspective. This helps alleviate the situation that exists today, where users are often black-boxed, and the concept of the user is underdeveloped.

This research is a strong illustration of the role that power and organizational politics play in design and development processes, showing how problems arise that can put a stop to UCD efforts, or at least stop them having their potential influence. With its focus on these factors, this research contributes to alleviating the situation where power and politics have been found to be absent in the UCD literature. As an illustrative example, previous research has shown that there is a growing realisation that usability is important, but that developers do not know how to achieve it. This was also clear in our research, but in addition to this, the research demonstrates that even in cases where this knowledge did exist, organizational factors stopped the usability work, as other issues were judged as being more important to deal with.

1.7.3 Institutions

One final research challenge was outlined when dealing with the area of organizations, focusing upon wicked problems. This research challenge concerns understanding the way in which organizational factors affect how usability work is discussed and performed in a large organisation, and the way in which the history and structure of the organization affect what gets done within the organization.

The study at ABB is an example of work aimed at the previously identified need for research into the ways that regulative processes, normative systems and cultural frameworks shape the design and use of technical systems.
The research at ABB contributes by providing a rich illustration of an institutionalized organization that is moving from a systems engineering context to a situation where new and unaccustomed problems are arising. As described in 1.4.3, systems engineering approaches have successfully addressed well-bounded problems, and engineering disciplines have been directed at solving clearly defined problems, with clear goals, that can be attacked from an established body of theory and knowledge. In this kind of environment, SE has been an accepted and useful approach. However, as this research demonstrates, these conditions are no longer the rule. The environment is characterised by ill-defined and potentially tacit goals that are value-laden and that may be impossible to define explicitly. In this situation, complex systems problems are emerging that must be addressed.

This research showed that there was little knowledge of end-users in the institution. In effect the users were “black-boxed”. The research demonstrates how this phenomenon of black-boxing users was caused by organizational factors, which stopped communication between the users and different stakeholders within the organization. Furthermore, the research illustrates a situation where there were problems in involving multiple external stakeholders, and problems in gaining a clear view of the larger picture. There were clear differences between the views of users and project boards, regarding the importance of different issues, and important issues were left unattended because of fixed views of what was important to include in the project budget. All of these issues are illustrative of the situation described in section 1.4.3. All of these issues are instrumental in the way wicked problems occur and must be approached and dealt with.

The rich picture given in this thesis show that, in the case of ABB, all of these issues that have been demonstrated within the institutionalized organization make working with usability and UX into a wicked problem.

1.7.4 Summing up

In summary, for the three areas that have been central to this research, the main contributions are:

**Quality**

Even though it is well known that usability work is important, there are still many obstacles to getting it done. Within usability work, measurement is important, and it is important to measure several factors, to get a clear view of the whole concept. But the measurements must be interpreted by a usability expert, who has a central role when explaining the measurements to different stakeholders, who all have their information needs that must be satisfied. To enable us to measure UX, allowing it to become accepted as a tool for SE, it is important that we understand the theoretical foundations of UX definitions.

**Organizations**

Usability testing can be used to improve usability processes, in the same way as Software Process Improvement improves other SE processes. However, usability and UX work is hindered by organizational barriers, such as lack of management support, or lack of time to negotiate or communicate, which stop organizations from finding out who users are, and...
hinder user participation. Within organizations, it was possible to distinguish two broad stakeholder groups, Designers and Product Owners, with different information needs on different time scales. It is important to adapt the testing and the way results are shown to these groups, and even here, there is a need for a specialised tester, who brings in the knowledge and experience of end-users.

**Institutions**

The research is a rich illustration of an institutionalized organization that is moving from a systems engineering context to a situation where new and unaccustomed problems are arising. It shows how regulative processes, normative systems and cultural frameworks shape the design and use of technical systems, in combination with organizational factors that stop communication between the users and different stakeholders within the organization. Important issues are left unattended because of fixed views of what was important to include in the project budget. All of these issues make working with usability and UX into a wicked problem.

**1.8 Conclusion**

The research in this thesis has stemmed from concrete needs identified within industry, from its beginnings together with UIQ, to the final stage together with ABB. The research process has been aimed at working towards solving these identified needs, within the areas of usability and UX. The methods used, and the questions addressed have varied over time, but they have all centred upon the questions of how industry can incorporate usability and UX in development processes for technological products.

An important part of the contribution this research makes is that it gives a graphic comparison between two very different types of companies, with different backgrounds, structures, and areas of operations. However, both companies were operating in a situation where the usability and UX of their products were seen as vital factors for improving, and even maintaining, their position in the marketplace. Together, they faced many common problems, and a number of individual problems.

This thesis is a vivid illustration of the current situation within the technology industry in a real-world setting, based as it is on many years of action research cooperation with industrial actors, giving an insight into the everyday situation in the companies, the changes that are taking place, the new demands that companies face, and the problems that are experienced in many parts of the organisations.

The foundation of the research has led to the scope of the questions asked, the breadth of the methods used, and the diversity of the answers. Together, this has led to a focused attempt to benefit usability and UX work in a real-life industrial context.

**1.9 Future work**

The research presented in this thesis provides a number of answers. These answers are the results of addressing a number of research challenges over a long period of time. As with all research, new questions arise as the research progresses. Some questions are left
unanswered. Even for the questions that are answered, there are new questions branching off from the answers. What follows here are but a few suggestions for further research. This research gives indications that it may possible to identify usability problems through a single metric, allowing a new and simpler way of performing usability testing in the real world. Work remains to be done testing these findings, performing tests on a greater number of devices, to find ways of doing this in practice. Beyond this, work can be done to test the heuristics that were used in this research, to find accurate ways to distinguish between high and low problem use cases.

This research has also shown many instances of how important the role of the test leader is. This concerns the process of performing usability and UX tests, the interpretation of the results, and the presentation of results. Much work remains to be done regarding the role of the test leader. Interesting questions remain as to who is a good usability and UX tester, and what the characteristics are that make a good usability tester.

The visualization of test results is an important field to develop, in order to enable stakeholders to gain answers to their questions about product usability and UX. This research has provided some presentation methods, but there is work remaining, discovering which types of information different stakeholders need, whether it is possible to find measurements to fulfill these needs, and how to visualize the results in the best possible way. This is particularly true for UX. There is still no agreement on how to measure UX, yet this is important to solve if UX is to become an established tool in the way that usability, despite many obstacles, has slowly become established and accepted in the software engineering discourse.

Despite all of the work that has been done on UX, the field is still young, and there is much left to do. This research indicates the importance of studying the theoretical foundations of definitions, in order to gain an understanding of the definitions. Work remains to achieve this in practice, and it may involve a study of the existing definitions and an analysis of which theoretical paradigm the creator of the definition is associated with.

More and more products and services are marketed in a mass market, or for use outside a traditional workplace. Thus, there is important work to do, finding ways that participatory design can be extended to find ways of including users in these situations. An alternative solution is to find ways to refine the role of the test leader, so that the test leader can truly work as a spokesperson for the users.

This research shows how usability and UX work is hindered by organizational barriers. This became particularly obvious when looking at usability work and UX in an institutionalized organization. When usability and UX work is approached in this kind of milieu, many new and challenging difficulties arise. Here, there are many possibilities for future research, looking at the field from a variety of standpoints, using theories from a variety of scientific fields, combining these to reach an understanding of the mechanisms at work, and thereby solving some of the problems that have been shown in this research.

Thus, there are endless questions left to answer, but from the point of view of research, that is exactly as it should be.
Chapter Two

Article Three
Chapter Two

Examining correlations in Usability Data to Effectivise Usability Testing


Jeff Winter, Mark Hinely

Abstract

Based on a case study performed in industry, this work deals with a statistical analysis of data collected during usability testing. The data is from tests performed by usability testers from two companies in two different countries. One problem in the industrial situation is the scarcity of testing resources, and a need to use these resources in the most efficient way. Therefore, the data from the testing is analysed to see whether it is possible to measure usability on the basis of one single metric, and whether it is possible to judge usability problems on the basis of the distribution of use case completion times. This would allow test leaders to concentrate on situations where there are obvious problems. We find that it is not possible to measure usability through the use of one metric, but that it may be possible to gain indications of usability problems on the basis of an analysis of time taken to perform use cases. This knowledge would allow the collection of usability data from distributed user groups, and a more efficient use of scarce testing resources.

2.1 Introduction

The background to this study is the situation faced by companies developing and testing consumer products for a mass market. The study is based on a long research cooperation between Blekinge Institute of Technology (BTH) and UIQ Technology AB (UIQ), an international company established in 1999. UIQ, who developed and licensed a user interface platform for mobile phones, identified a need to develop a flexible test method for measuring the usability of mobile phones, to give input to design and development processes, and to present usability findings for a number of stakeholders at different levels in the organization. This need resulted in the development of UIQ Technology Usability Metrics (UTUM). UTUM was successfully used in operations at UIQ until the closure of the company in 2009.

Together with UIQ we found that there is a need for methods that can simplify the discovery of usability problems in mobile phones. There is also a desire to find ways of identifying usability problems in phones without having to engage the test leader in every step of the process, with the ability to do it for geographically dispersed user groups. However, we also realise that even if it is found to be possible to identify problem areas, for example through a simple measurement of one metric, or through an analysis of
completion times, this would not identify the particular aspects of the use cases that are problematic for the users. It would simply indicate use cases where the users experienced problems. This means that further studies would still have to be performed by test leaders together with users, to examine and understand what the actual problems consist of, and how they affect the way that users experience the use of the phone. This must be done in order to create design solutions to alleviate the problems.

As we discuss in greater detail in section 3 of this article, the role of the usability tester is central in many ways, and it is a role that is not easily filled. It demands particular personal qualities, knowledge and experience. It involves the ability to communicate with people on many organisational levels, the ability to observe, record and analyse the testing process, and the ability to present the results of testing to many different stakeholders. Since there is a scarcity of people who can fill this role, it would ease the situation for companies wanting to perform usability testing if these resources could be used in the most efficient way possible. This is the principle behind the need to identify problematic use cases without having to involve the test leader in every step of the process.

If it is possible to identify use cases that are problematic, without requiring the presence of the test leader, this will allow companies to pinpoint which areas require further testing, so that test leaders can work more efficiently. Since we are working with a mass-market product, being able to do this remotely, for widely dispersed groups, would also be an advantage for the company, in order to test solutions in different geographical areas without requiring the usability tester to travel to these areas before there is seen to be a need, and to reduce the amount of testing that needs to be done on-site.

These needs are the basis of this article. In this work, we examine the metrics collected in the UTUM testing, to study the correlations between the metrics for efficiency, effectiveness, and satisfaction, to see whether we can measure usability on the basis of one metric, and we examine whether it is possible to develop a simple method of automatically identifying problem areas simply by measuring and analysing the time taken to perform different use cases.

2.2 Research Questions

The aim of this study is to examine whether there is a simple measurement to express usability, and to find if it is possible to streamline the discovery of problematic use cases. To do this, we examine the correlation between metrics for efficiency, effectiveness and satisfaction that have been collected during the testing process. These are the different elements of usability as specified in ISO 9241-11:1998 (ISO 9241-11 1998). This is done in order to see whether there are correlations that allow us to discover usability problems on the basis of a simple metric. To satisfy the needs within industry, this metric should preferably be one that can easily be measured without the presence of the test leader. Based on this situation, we have formulated two research questions:

– RQ1: What is the correlation between the different aspects of usability (Effectiveness, Efficiency and Satisfaction)?
RQ2: Can a statistical analysis of task-completion time allow us to discover problematic use cases?

The first research question is based on the idea that there may be a sufficiently strong correlation between the 3 factors of usability that measuring one of them would give a reliable indication of the usability of a mobile phone. The second research question is based on the theory that there is an expected distribution of completion times for a given use case and that deviations from goodness of fit indicate user problems.

This study is a continuation of previous efforts to examine the correlations between metrics for efficiency, effectiveness and satisfaction. A previous study by Frøkjær, Hertzum et al (2000) found only weak correlations between the different factors of usability, whereas a study by Sauro and Kindlund (2005) showed stronger correlations between the different elements. The results of this study will be placed in relation to these studies, to extend knowledge in the field. This is also a continuation of our previous work, where we have examined how the UTUM test contributes to quality assurance, and how it balances the agile and plan-driven paradigms (see e.g. Winter, Rönkkö et al. 2007; Winter, Rönkkö et al. 2008; Winter 2009; Winter, Rönkkö et al. 2009).

2.3 Usability and the UTUM Test

UTUM is an industrial application developed and evolved through a long term cooperation between BTH and UIQ. UTUM is a simple and flexible usability test framework grounded in usability theory and guidelines, and in industrial software engineering practice and experience.

According to (ISO 9241-11 1998) Usability is 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. Effectiveness is the accuracy and completeness with which users achieve specified goals. Efficiency concerns the resources expended in relation to the accuracy and completeness with which users achieve goals. Satisfaction concerns freedom from discomfort, and positive attitudes towards the use of the product.

UTUM measures the usability of products on a general level, as well as on a functional level. According to Hornbæk (2006), amongst the challenges when measuring usability are to distinguish and compare subjective and objective measures of usability, to study correlations between usability measures as a means for validation, and to use both micro and macro tasks and corresponding measures of usability. Emphasis is also placed on the need to represent the entire construct of usability as a single metric, in order to increase the meaningfulness and strategic importance of usability data (Sauro and Kindlund 2005). UTUM is an attempt to address some of these challenges.

An important characteristic of the UTUM test is the approach to understanding users and getting user input. Instead of simply observing use, a test expert interacts and works together with the users to gain insight into how they experience being a mobile phone user, in order to gain an understanding of the users’ perspective. Therefore, users who help with UTUM testing are referred to as testers, because they are doing the testing, rather than
being tested. The representative of the development company is referred to as the test leader, or test expert, emphasising the qualified role that this person assumes.

The test experts are specialists who bring in and communicate the knowledge that users have, in accordance with Pettichord (2000), who claims that good testers think empirically in terms of observed behaviour, and must be encouraged to understand customers’ needs. Evidence in (Martin, Rooksby et al. 2007) suggests that drawing and learning from experience may be as important as taking a rational approach to testing. The fact that the test leaders involved in the testing are usability experts working in the field in their everyday work activities means that they have considerable experience of their products and their field. They have specialist knowledge, gained over a period of time through interaction with end-users, customers, developers, and other parties that have an interest in the testing process and results. However, these demands placed on the background and skills of test leaders mean that these types of resources are scarce, and must be used in the most efficient way possible.

A second characteristic of UTUM is making use of the inventiveness of phone users, by allowing users to participate actively in the design process. The participatory design tradition (Schuler and Namioka 1993) respects the expertise and skills of the users, and this, combined with the inventiveness observed when users use their phones, means that users provide important input for system development. The test expert has an important role to play as an advocate and representative of the user perspective. Thus, the participation of the user provides designers, with the test expert as an intermediary, with good user input throughout the development process.

The user input gained through the testing is used directly in design and decision processes. Since the tempo of software development in the area of mobile phones is high, it is difficult to channel meaningful testing results to recipients at the right time in the design process. To address, this problem, the role of the test expert has been integrated into the daily design process. UTUM testing is usually performed in-house, and results of testing can be channelled to the most critical issues. The continual process of testing and informal relaying of testing results to designers leads to a short time span between discovering a problem and implementing a solution.

The results of testing are summarised in a clear and concise fashion that still retains a focus on understanding the user perspective, rather than simply observing and measuring user behaviour. The results of what is actually qualitative research are summarised by using quantitative methods. This gives decision makers results in the type of presentations they are used to dealing with. Statistical results are not based on methods that supplant the qualitative methods that are based on PD and ethnography, but are ways of capturing in numbers the users’ attitudes towards the product they are testing.

A UTUM test does not take place in a laboratory environment, but should preferably take place in an environment that is familiar to the person who is participating in the test, in order that he or she should feel comfortable. When this is not possible, it should take place
in an environment that is as neutral as possible. Although the test itself usually takes about 20 minutes, the test leader books one hour with the tester, in order to avoid creating an atmosphere of stress. The roles in testing are the test leader, who is usually a usability expert, and the tester.

In the test, the test leader welcomes the tester, and tries to put the tester at their ease. This includes explaining the purpose of the test, and saying that it is the telephone that is being tested, not the performance of the tester. The tester is instructed to tell the test leader when she or he is ready to begin the use case, so that the test leader can start the stopwatch to time the use case, and the tester should also tell the test leader when the use case is complete.

The tester begins by filling in some of their personal details and some general information about their phone usage. This includes name, age, gender, previous telephone use, and other data that can have an effect on the result of the test, such as which applications they find most important or useful. In some circumstances, this data can also be used to choose use cases for testing, based on the tester’s use patterns.

For each phone to be tested, the tester is given time to get acquainted with the device. If several devices are to be tested, all of the use cases are performed on one device before moving on to the next phone. The tester is given a few minutes to get acquainted with the device, so that he or she can get a feeling for the look and feel of the phone. When this has been done, the tester fills in a Hardware Evaluation, a questionnaire based on the System Usability Scale (SUS) (Brooke 1986) about attitudes to the look and feel of the device. The SUS was developed in 1986 by John Brooke, then working at the Digital Equipment Company. The SUS consists of 10 statements, where even-numbered statements are worded negatively, and odd-numbered statements are worded positively.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The answers in the SUS are based on Likert style responses, ranging from “Strongly disagree” to “Strongly agree”. The Likert scale is a widely used summated rating that is easy to develop and use. People often enjoy completing this type of scale, and they are
likely to give considered answers and be more prepared to participate in this than in a test that they perceive as boring (Robson 2002, p. 293).

Brooke characterised the SUS as being a “Quick and Dirty” method of measuring usability. However, Lewis and Sauro state that although SUS may be quick, it is probably not dirty, and they cite studies that show that SUS has been found to be a reliable method of rating usability (Lewis and Sauro 2009). SUS has been widely used in the industrial setting, and Lewis and Sauro state that the SUS has stood the test of time, and they encourage practitioners using the SUS to continue to do Usability and Learnability components, beyond showing the overall SUS score (Lewis and Sauro 2009). In a study of questionnaires for assessing the usability of a website, Tullis and Stetson found that the SUS, which was one of the simplest questionnaires studied, was found to yield amongst the most reliable results across sample sizes, and that SUS was the only questionnaire of those studied that addressed all of the aspects of the users’ reactions to the website as a whole (Tullis and Stetson 2004).

In a UTUM test, the users perform the use cases, the test leader observes what happens, and records the time taken to execute the tasks, observes hesitation or divergences from a natural flow use, notes errors, and counts the number of clicks to complete the task. Data is recorded in a form where the test leader can make notes of their observations. The test leader ranks the results of the use case on a scale between 0 - 4, where 4 is the best result. This judgement is based on the experience and knowledge of the test leader. This means that the result is not simply based on the time taken to perform the use case, but also on the flow of events, and events that may have affected the completion of the use case.

After performing each use case, the tester completes a Task Effectiveness Evaluation, a shortened SUS questionnaire (Brooke 1986) concerning the phone in relation to the specific use case performed. This is repeated for each use case. Between use cases, there is time to discuss what happened, and to explain why things happened the way they did. The test leader can discuss things that were noticed during the test, and see whether his or her impressions were correct, and make notes of comments and observations. Even though the test leader in our case does not usually actively probe the tester’s understanding of what is being tested, this gives the opportunity to ask follow up questions if anything untoward occurs, and the chance to converse with the tester to glean information about what has occurred during the test.

The final step is an attitudinal metric representing the user’s subjective impressions of how easy the phone is to use. This is found through the SUS (Brooke 1986), and it expresses the tester’s opinion of the phone as a whole. The statements in the original SUS questionnaire are modified slightly, where the main difference is the replacement of the word “system” with the word “phone”, to reflect the fact that a handheld device is being tested, rather than a system. This SUS questionnaire results in a number that expresses a measure of the overall usability of the phone as a whole. In general, SUS is used after the user has had a chance to use the system being evaluated, but before any debriefing or discussion of the
test. In UTUM testing, the tester fills in the SUS form together with the test leader, giving an opportunity to discuss issues that arose during the test situation.

The data collected during the test situation is used to calculate a number of metrics, which are then used to make different presentations of the results to different stakeholders. These include the Task Effectiveness Metric, which is determined by looking at each use case and determining how well the telephone supports the user in carrying out each task. It is in the form of a response to the statement “This telephone provides an effective way to complete the given task”. It is based on the test leader’s judgement of how well the use case was performed, recorded in the test leader’s record and the answers to the Task Effectiveness Evaluation. The Task Efficiency Metric is a response to the statement “This telephone is efficient for accomplishing the given task”. This is calculated by looking at the distribution of times taken for each user to complete each use case. The distribution of completion times is used to calculate an average value for each device per use case. The User Satisfaction Metric, is calculated as an average score for the answers in the SUS, and is a composite response to the statement “This telephone is easy to use”. For more information regarding different ways of presenting these metrics and data, see (Winter 2009, Appendix A).

A previous study by Winter and Rönkkö (2009) showed that two different groups of stakeholders existed within UIQ. The first group was designated as Designers represented by e.g. interaction designers and system and interaction architects, representing the shop floor perspective. The second group was designated as Product Owners, including management, product planning, and marketing, representing the management perspective. These two groups were found to have different needs regarding the presentation of test results. These differences concerned the level of detail included in the presentation, the ease with which the information can be interpreted, and the presence of contextual information included in the presentation. Designers prioritised presentations that gave specific information about the device and its features, whilst Product Owners prioritised presentations that gave more overarching information about the product as a whole, and that were not dependent on including contextual information.

These results, and more information on UTUM in general, are presented in greater detail in (Winter 2009, Chapter 4 and Appendix A). A video demonstration of the test process (ca. 6 minutes) can be found on YouTube (BTH 2008).

### 2.4 Research Method

The cooperative research and development work that led to the development of UTUM has been based on an action research approach according to the research and method development methodology called Cooperative Method Development (CMD) (see e.g. Dittrich, Rönkkö et al. 2007). CMD is an approach to research that combines qualitative social science fieldwork, with problem-oriented method, technique and process improvement. CMD has as its starting point existing practice in industrial settings, and
although it is motivated by an interest in use-oriented design and development of software, it is not specific for these methods, tools and processes.

This particular work is based on a case study (Yin 2003) and grounded theory (Glaser and Strauss 1967) approach. A case study is “an empirical enquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” (Yin 2003, p. 13). The focus is on a particular case, taking the context into account, involving multiple methods of data collection; data can be both qualitative and quantitative, but qualitative data are almost always collected (Robson 2002, p. 178). Case studies have their basis in a desire to understand complex social phenomena, and are useful when “how” or “why” questions are being asked, and where the researcher has little control over events (Yin 2003, p. 7). A case study approach allows the retention of characteristics of real life events (Yin 2003).

The data in this case study has been analysed in a grounded theory approach. Grounded theory (GT) is both a strategy for performing research and a style of analysing the data that arises from the research (Robson 2002, p. 191). It is a systematic but flexible research style that gives detailed descriptions for data analysis and generation of theory. It is applicable to a large variety of phenomena and is often interview-based and (Robson 2002, p. 90) but other methods such as observation and document analysis can also be used (Robson 2002, p. 191). We have not attempted to work according to pure GT practice, and have applied a case study perspective, using ethnography (Rönkkö 2010) and participatory design (Schuler and Namioka 1993).

2.5 Subjects and Context

The data in this study were collected in tests performed by UIQ in Sweden and by a tester from a mobile phone manufacturer in England. The testing was performed in a situation where there are complex relationships between customers, clients, and end-users, and complexities of how and where results were to be used. The phones were a UIQ phone, a “Smart phone” of a competing brand, and a popular consumer phone. The use cases were decided by the English company, and were chosen from their 20 most important use cases for a certain mobile phone. The use cases were:

- UC1. Receive and answer an incoming call
- UC2. Save the incoming call as a new contact - “Joanne”
- UC3. Set an alarm for 8 o’clock tomorrow morning
- UC4. Read an incoming SMS and reply with “I’m fine”
- UC5. Make a phone call to Mårten (0708570XXX)
- UC6. Create a new SMS - “Hi meet at 5” and send to Joanne (0708570XXX)

The test group consisted of 48 testers. The group consisted of 24 testers from Sweden, and 24 testers from England, split into 3 age groups: 17 - 24; 25 - 34; 35+. Each age group consisted of 8 females and 8 males. The size of the group was in order to get results from a wide range of testers to obtain general views, and to enable comparisons between age
groups, cultures and genders. Normally, it was not deemed necessary to include so many testers, as small samples have been found to be sufficient to evaluate products. Dumas and Redish (1999) for example, refer to previous studies that indicate in one case that almost half of all major usability problems were found with as few as three participants, and in a second case that a test with four to five participants detected 80% of usability problems, whilst ten participants detected 90% of all problems. This indicates that the inclusion of additional participants is less and less likely to contribute new information. The number of people to include in a test thus depends on how many user groups are needed to satisfy the test goals, the time and money allocated for the test, and the importance of being able to calculate statistical significance.

However, even though this can be seen from the point of view of the participating organisations as a large test, compared to their normal testing needs, where the data collected consisted of more than 10 000 data points, the testing was still found to be a process where results were produced quickly and efficiently. In this case, the intention of using a larger number of testers was to obtain a greater number of tests, to create a baseline for future validation of products, to identify and measure differences or similarities between two countries, and to identify issues with the most common use-cases. Testers were drawn from a database of mobile phone users who have expressed an interest in being testers, and who may or may not have been testers in previous projects.

2.6 Validity

Regarding internal reliability, the data used in this study have been collected according to a specified testing plan that has been developed over a long period of time, and that has been used and found to be a useful tool in many design and development projects. The risk of participant error in data collection is small, as the test is monitored, and the data is verified by the test leader. The risk of participant bias is also small, as the testers are normal phone users, using a variety of different phones, and they gain no particular benefits from participating in the tests or from rating one device as being better than another. The fact that much of the data has been in the form of self-evaluations completed by the testers themselves, and that the testing has been performed by specialized usability experts minimizes the risk of observer error. The risk of observer bias is dealt with by the presence of the two independent test leaders, allowing us to compute inter-observer agreements. The use of multiple methods of data collection, including self-assessment, test leader observation and measurement, and the collection of qualitative data, allow us to base our findings on many types and ways of collecting data.

In regard to external validity, the fact that the testing has been performed in two different countries may be seen as a risk, but the two countries, Sweden and England, are culturally relatively close, which should mean that the results are comparable across the national boundaries. The tasks performed in the testing are standard tasks that are common to most types of mobile phones, and should therefore not affect the performance or results of the tests. The users are a cross section of phone users, and the results should thus be generalizable to the general population of phone users.
To ensure the statistical conclusion validity, we use statistical methods that are standard in the field, and use the SPSS software package PASW Statistics 18 for statistical analysis.

2.7 Data Sets, Possible Correlations and Analysis Steps

The test data has been split into three sets of data. This division is based on the metrics collected in the attitudinal questionnaires and the times recorded by the test leader during testing. These data sets concern satisfaction, effectiveness and efficiency, as called for by ISO 9241-11:1998 (ISO 9241-11 1998). The sets of test data are:

Set 1: SUSuapp - Based on the System Usability Scale (SUS) (Brooke 1986), which consists of 10, 5-scale Likert questions. The evaluation is a user appraisal of satisfaction, based on one evaluation per phone and tester. It is a summary of the use cases performed on the individual phones. It provides us with a total of 144 data points - 48 per phone (48 testers, 3 phones, 1 SUS per phone).

Set 2: TEEuapp - Based on a Task Effectiveness Evaluation (TEE), which consists of 6, 5-scale Likert questions. It is a user appraisal based on one evaluation per phone, use case and tester. The tester fills in this evaluation directly after completing each of the 6 use cases on each of the three phones. It provides us with a total of 864 data points - 144 per use case task (48 testers, 6 use cases, 3 phones).

Set 3: TIMEreal - This is used to represent efficiency, and is the time taken in seconds to complete a use case task. It is a test leader measurement based on one number per phone, use case and tester. The test leader measures the time for the tester to complete each of the use cases on each of the phones. This provides us with 864 data points- 144 per use case task (48 testers, 6 use cases, 3 phones).

As a complement to these data, we also make use of a spreadsheet, the Structured Data Summary (SDS) (Denman 2008) that is used to record qualitative data based on the progress of the testing. This contains some of the qualitative findings of the testing and the SDS shows issues that have been found, for each tester, and each device, for every use case. Comments made by the testers and observations made by the test leader are stored as comments in the SDS.

The first step in the data analysis is to investigate the strength of the correlations between the metrics for satisfaction, effectiveness and efficiency. The second step is to investigate if the distribution of time taken to perform use cases can provide a reliable indication of problematic use cases, and in which way this should be analysed and shown. If this is successful, it should be possible to discover use cases that exhibit poor usability by looking at the shape of the distribution curve. The third step is to verify the fact that the distribution of time can be used to illustrate the fact that certain use cases exhibit poor usability. This can be done by comparing with the data recorded in the SDS for these use cases, to see if the test leader has noted problems that users experienced. If this is found to be the case, this indication could be used when testing devices, to identify the areas where test leader resources should be directed, thus allowing a more efficient use of testing resources.
Chapter Two
Examining Correlations in usability data

STEP 1: Investigate the correlation between Satisfaction, Effectiveness and Efficiency. For each phone each tester completed a SUS-evaluation (SUSuapp). SUSuapp gives an appraisal score from 0-40. The correlation between the SUSuapp, and TEEuapp might be calculated using Pearson’s correlation coefficient, Spearman rank correlation coefficient, or Kendall’s rank correlation coefficient (Kendall’s Tau). The most reasonable method could be Spearman or Kendall’s tau, as these deal with data in the form of ranks or ordering of data, and do not assume normal distribution of the data, on which the Pearson coefficient is based. Spearman is preferred by some, as it is in effect a Pearson coefficient performed on ranks, but Kendall’s Tau is usually preferred, as it deals with ties more consistently (Robson 2002)

The SUSuapp data is the result the 144 Likert appraisals, which could normally be assumed to exhibit a normal distribution. However, in some of the other data distributions, we have observed a positive skew that also suggests that Spearman may be a better choice. Also, the central concentration of the data causes many ties in ranks, which could make Kendall’s Tau more appropriate.

The tests that include TIMEreal may be more difficult to deal with. Since the TIMEreal data is continuous, while the other data is of Likert-type, it may be difficult to see any linear relationships. However, the same tests should still be performed.

The results of the analysis are found in table 2.1. The analysis shows only weak to moderate correlations between the different factors. This is particularly obvious regarding Kendall’s tau, which as previously mentioned is probably the best indicator given the type of data involved here. This supports the findings of Frøkjær, Hertzum et al (2000), who state that all three factors must be measured to gain a complete picture of usability. It contradicts the results of Sauro and Kindlund (2005), who showed stronger correlations, although even Sauro et al state that it is important to measure all three factors, since each measure contains information not contained in the other measures.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Sig. (2-tailed)</th>
<th>Correlation</th>
<th>Sig. (2-tailed)</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kendall's tau b</td>
<td>Spearman's rho</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUSuapp/</td>
<td>0.599**</td>
<td>0.000</td>
<td>0.758**</td>
<td>0.000</td>
<td>0.710**</td>
</tr>
<tr>
<td>TEEuapp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUSuapp/</td>
<td>-0.408**</td>
<td>0.000</td>
<td>-0.573**</td>
<td>0.000</td>
<td>-0.485**</td>
</tr>
<tr>
<td>TIMEreal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.490**</td>
<td>0.000</td>
<td>-0.663**</td>
<td>0.000</td>
<td>-0.595**</td>
<td>0.000</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
These results do not support our conjecture that there is a sufficiently strong correlation between the 3 factors of usability that simply measuring one of them would give a reliable indication of the usability of a mobile phone.

STEP 2: Investigating if the distribution of time can provide a reliable indication of problematic use cases. We find that TIME real data, for time taken to complete a given use case, corresponds well with a Rayleigh distribution (Ray(2*mean)) with a shape parameter that is twice the mean of the data. Data points that end up in the tail fall under a specific degree of probability of belonging to the Ray(2*mean) distribution. This means that the use cases with a “long tail” are those that the testers found to be troublesome (see Figure 2.1).

Figure 2.1 illustrates one use case. The right hand diagram is the seconds to complete the use case divided into ten evenly spaced frequency intervals. The diagram to the left is the Ray(2*mean) probability distribution. For example, we see that 2 on the x-axis has a 28% chance belonging to the Rayleigh distribution, and that is where we have a frequency of 70+ data points. 6 on the x-axis has a less than 1% chance of belonging to the distribution and we see that 6 in our data is empty. This would mean that the points in our data set in ranges 7-10 are beyond all probability influenced by something more than the excepted random difference between different testers. Our interpretation is that these are the use cases where “something went wrong”. This result suggests that it may be possible to discover use cases where users have problems, by examining the distribution of the time taken to perform the use case.

STEP 3: Verifying the “long tail” method of identifying troublesome use cases. Here we analyse which use cases the testers have experienced as exhibiting poor usability by analysing the distribution of time taken to complete the use case. This is cross tabulated with data from the SDS (Denman 2008), the spreadsheet containing some of the qualitative findings of the testing. The SDS shows issues that have been found, for each tester, and each device, for every use case. Comments made by the testers and observations made by the test leader are stored as comments in the spreadsheet.
Given the fact that the intention of this work is to find ways that simplify the discovery of problematic use cases, and the fact that the test is designed to be flexible and simple to perform and analyse, we attempted to find some simple heuristic that could help us differentiate between the use cases with high and low levels of problems. We ordered the use cases according to their coefficients of variation, which is the standard deviation, divided by the mean time taken to perform the use case. This allows us to standardize the use cases, in order to give a basis for comparison.

This calculation gave us a spread between 0.481 and 1.074. To give a simple cut-off point between Lower and Higher problem use cases, we set a boundary where a coefficient of variation of 0.6 is regarded as High problem, thus dividing the set of use cases into two groups. We also use a simple heuristic to judge an acceptable level of problems when performing a use case. The test leader registers problems observed whilst performing the use case by a letter “y” in the SDS, with an explanatory comment. One tester may have experienced more than one problem, and all of these are noted separately, but we chose to count the number of individuals who had experienced problems, rather than the number of problems. The seriousness of the problems could range from minor to major. However, since the ambition was to find a simple heuristic, we have not performed any qualitative analysis of the severity of the problems, but have simply noted number of users who had problems. We refer to this as No_USERS.

In this case, the cut-off point was set as being less that 33% of the total number of testers. We assume that use cases where more than 33% of users had some kind of problem are High problem, and worthy of further examination.

Table 2.2 Use cases and their categorization as High or Low problem

<table>
<thead>
<tr>
<th>Case</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient of variation</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Severity</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient of variation</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Severity</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

Table 2.2 illustrates the cases and their categorization as High or Low problem for Coefficient of variation and No_USERS.

We performed Fisher’s exact test on the set of data shown in table 2.2. This test can be used in the analysis of contingency tables with a small sample. It is a statistical test that is used to determine if there are non-random associations between two categorical variables. The results of performing Fisher’s exact test are shown in table 2.3.
Since the values given by Fisher’s exact test are below 0.05 they can be regarded as significant, meaning that there is a statistically significant association between Coefficient of variation and No_USERS as we have defined them.

As can be seen in table 2.3, all of the cases (5) where No_USERS indicated a high rate of problems are discovered by the coefficient of variation being high. On the other hand, a high coefficient of variation also points to just as many cases that do not have a high rate of problems. However, the results still show that a number of use cases (8) can, with high probability, be excluded from the testing process, allowing for more efficient use of testing resources. Simply by calculating the coefficient of variation, 8 of 18 cases could be excluded from more expensive testing.

To conclude, the SDS records the fact that the test leader observed that users experienced problems when performing use cases, and there is found to be an association between the use cases where a larger proportion of users experienced problems, and those use cases with a high coefficient of variation. This suggests that it is possible to identify potentially problematic use cases simply by measuring the time taken to perform use cases and analysing the distribution of those times.

This article is based on research that was performed previous to the cessation of activities in UIQ. The limited number of tests that were available to be included for analysis in this study, the fact that the testing as it was performed was not designed as an experiment with this purpose in mind, and that this is a post factum analysis mean that the results must be read with some caution. However, the results we have obtained from this analysis do indicate that this is an interesting area to study more closely.

This means that it may be possible to formulate a “time it and know” formula that can be tested in new trials. This could be used to give a “problem rating” to individual use cases that could categorize the degree of problems that the user experienced. It would allow a simple categorization of use cases without needing the presence of the test leader, simply by measuring the time taken to perform the use cases, in order to identify the areas where
test leader resources should be directed, thus allowing a more efficient use of testing resources.

2.8 Discussion

The aims of this study have been twofold: to examine the correlation between the different aspects of usability (Effectiveness, Efficiency and Satisfaction) to find whether there is one simple measurement that would express usability, and; to discover if it is possible to streamline the discovery of problematic use cases through a statistical analysis of task-completion time, which would allow scarce testing resources to be concentrated on problematic areas.

The analysis detailed above shows that, for the material collected in our study, the correlations between the factors of usability are not sufficiently strong to allow us to base usability evaluations on the basis of one single metric. This means that it is important that all three factors are measured and analysed, and as discussed previously, the test leader is an important figure in this process. This supports previous work that stresses the importance of measuring all of these aspects. This was stated to be the case even by those researchers who found stronger correlations between the different aspects measured.

However, we do find that it may be possible to discover potentially problematic use cases by analysing the distribution of use case completion times. This would mean that it is possible to collect data which indicate which use cases are most important to concentrate testing resources on. This could be done without the presence of a test leader. Many companies involved in developing and producing mass-market products already have a large base of testers and customers who participate in different ways in evaluating features and product solutions. By distributing trial versions of software to different user groups, and by using an application in a mobile phone that measures use case completion time, and submits this data to the development company, it should be possible to collect data in a convenient manner. The development company could distribute instructions to users and testers, who could perform use cases based on these instructions, and the telephone itself could transmit data to the company, which could form the basis of the continued analysis and testing process. This data would be especially valuable since it could be based more on the use of the telephone in an actual use context, rather than in a test situation.

From an analysis of the distribution of completion times it is thus possible to gain indications of problem areas that need further attention. However, it is impossible to say, simply by looking at the completion times, what the problem may be. To discover this, and to develop design suggestions and solutions, it is still necessary for the test leader to observe and analyse the performance of the use cases that are indicated as problematic.

Future work would be to test the findings made here, by performing further tests on a greater number of devices, and comparing the results with UTUM testing as it is normally performed. It is also possible to study cases where the statistics indicate that there are problems, and other devices where this was not apparent in the statistics, and compare the
results. Further work would also be to test the heuristics used in our analysis, to find if there are more accurate ways of distinguishing between low and high problem use cases.

Acknowledgements
This work was partly funded by The Knowledge Foundation in Sweden under a research grant for the software development project “Blekinge – Engineering Software Qualities”, www.bth.se/besq. We would like to thank Associate Professor Claes Jogréus for reading our work and giving advice regarding statistical methods, and all of our partners from UIQ Technology AB, for their help and cooperation.
Chapter Three

Article One
Chapter Three

The concept of UX depends on the intellectual landscape of HCI paradigms

Submitted to International Journal of Human-Computer Studies, Special Issue on the Interplay between User Experience Evaluation and Software Development

Kari Rönkkö, Jeff Winter

Abstract

Since the turn of the millennium, the topic of User Experience (UX) has become more and more frequent in academic discourses and today it is one of the central topics in Human Computer Interaction (HCI). We have pressure from businesses and industry saying that usability is not enough anymore. Today the market expects usability to be good as default, and buyers will be irritated if it is not, but what really sells products is a good user experience. HCI is the primary academic community when it comes to research within UX, whereas Software Engineering (SE), that slowly incorporated usability as a quality and product measurement, has barely touched upon the topic of UX. Still, UX will become increasingly important even for SE, in order to measure quality on a broader framework, including customer satisfaction and business needs. Today we have a number of UX definitions, but despite so many definitions of UX, it is still difficult to understand UX as concept. Understanding UX as a concept can be viewed as a bottleneck influencing the success rate of customer satisfaction within both HCI and SE. Until we reach some kind of agreement regarding the meaning and use of definitions, then we cannot reach a stage where UX becomes a meaningful tool in the software engineering toolbox, allowing us to measure and make use of this important and highly relevant concept. This paper concerns itself with and revolves around the subject of definitions of User Experience (UX). Starting in this situation that we have identified, that shared definitions has not been the answer hoped for, we reflect in this paper over what role a definition can have in this effort to strive towards a better understanding of the concept of UX. It suggests that we must look to the origins of the underlying theoretical foundation of a definition of UX and define the origins of it in order to understand it.

3.1 Introduction

This paper concerns itself with and revolves around the subject of definitions of User Experience (UX). Part of this concern is why, despite so many definitions of UX, it is still difficult to understand UX as concept. It suggests that we must look to the origins of the underlying theoretical foundation of a definition of UX and define the origins of it in order to understand it. Since the turn of the millennium, the topic has become more and more
frequent in academic discourses and today it is one of the central topics in Human Computer Interaction (HCI). HCI is also the primary academic community when it comes to research within UX, whereas Software Engineering, that slowly incorporated usability as a quality and product measurement aspect (Winter 2013), has barely touched upon the topic of UX. As a roadmap for future work in software engineering, Pfleeger and Atlee (2006, p. 613) stated that it will become increasingly important to measure quality on a broader framework, including customer satisfaction and business needs. From the present authors point of view this would need to include the incorporation of UX.

The research within UX has increased significantly within HCI since the turn of the millennium. Definitions and measurements are two important focus areas related to UX. These and many other types of prominent initiatives that have improved our understanding of UX can be found at conferences such as NordiCHI ’06, HCI ’07, CHI ’08, CHI ’09, INTERACT ’09. Examples of influential UX concepts that have profoundly shaped our understanding of UX include hedonic (Hassenzahl 2001), joy (Hassenzahl, Beu et al. 2001) and pleasure (Jordan 2002). of the state of the art concerning empirical research on UX. Based on 51 publications Bargas-Avila and Hornbaek (2011) have recently provided a study the authors investigate how empirical research on UX has been conducted. Among their findings they revealed that emotions, enjoyment, and aesthetics are the most frequently assessed dimensions of empirical research, but also that context of use is rarely researched. The latter is surprising, since understanding context is repeatedly emphasized as being a crucial part of understanding UX. Other examples of UX discourses that have been published over the years include:

- The nature of experience (Jacko, Mahlke et al. 2007; Hassenzahl 2008; Karapanos, Zimmerman et al. 2009),
- Characteristics of UX (Law, Roto et al. 2008; Roto, Obrist et al. 2009; Kuutti 2010),
- Giving UX its shape (McNamara and Kirakowski 2006; Hassenzahl, Diefenbach et al. 2010; Obrist, Tscheligi et al. 2010),
- Evaluating UX (Agarwal and Meyer 2009; Barnum and Palmer 2010; Wimmer, Wöckl et al. 2010)

This paper is concerned with whether background theories already presented by the UX community are useful or not, and if so, in what sense. It is obvious that researchers have found UX important, and have also interpreted the term UX into a wide range of meanings from different background theories, assumptions and viewpoints. Today the concept of UX is widely used and also understood in a variety of different ways, dependent on e.g. technologies applied, identified users and relevant stakeholders, activities involved, different contexts, researchers’ background and the researchers’ own interest in UX.

As already mentioned, measurement is an area in HCI in which UX is discussed. It is not surprising that the initiatives of measurement of UX, and how to achieve standardization,
are progressing but still in their infancy. As early as 2008, Bevan (2008) claimed that the interpretations of UX at that point of time were even more diverse than those of usability. We cannot see any change in this situation. As is visible in the overview of UX provided above, the multitude of angles on UX has actually increased. Bevan stated that appropriate UX measures would be simplified if the different perspectives on UX were identified and distinguished. However, he did not elaborate on what this means and what it would take to accomplish it. Instead, his approach here was to contrast UX with usability. Bevan suggested a common framework for classifying usability and UX measures. The framework built upon the idea of comparing existing and evolving UX issues with already established usability measures such as effectiveness, efficiency, and satisfaction.

In a later paper, Bevan (2009) once again addresses the differences between usability and UX methods. In this paper Bevan points out that the concept of UX broadens both the range of human responses to be measured and the circumstances in which measurement occurs. One especially tricky aspect identified in his paper is that of subjectivity and objectivity. Researchers working in the field have argued that there are substantial differences between objective task-related measures of usability and subjective measurements of users’ feelings related to UX (Roto, Obrist et al. 2009). In contrast, Bevan suggests that we should focus upon optimization instead, using the argument that both usability and UX in the end are all about optimizing human performance and user satisfaction. This is a very pragmatic approach. Bevan suggests that the specific product and design objectives will decide which methods to apply. In such a view it is difficult to know whether the background theories already presented by the UX community are not useful enough to be included in the discussion of measurements, or is it rather so that a piece of the puzzle is missing? A piece that would explain how to relate the difference in background theories to the practical achievements of measurement. A piece that would also explain why sometimes it is not possible to mix different UX theories. The authors of the present paper believe so. We provide a first outline towards identifying what that missing piece might be.

It is interesting to note that definitions played an important role in the UX community when discussing UX. As is the case in Bevan’s argumentation in both of the papers discussed above. Definitions were also the means Bevan himself used for differentiating and distinguishing between different perspectives on UX. Many other researchers during the last decade have also put their hopes in the power of definitions as the means to help the community to create a shared understanding of UX. Unfortunately, the debate surrounding the idea of approaching UX through definitions is weak, or frankly speaking non-existent. This paper contributes here by introducing a discourse in which we exemplify how researchers’ HCI paradigms are tightly intertwined with understanding definitions of UX. In fact, we claim that the concept of UX cannot be understood without reference to the researchers’ applied research paradigms. This statement is a substantial part of the identified missing piece mentioned above.

Many explanations exist for why the definition of UX is so hard. Law et al (2008) suggested several reasons why it is hard to formulate a universal definition of UX:

:: 95 ::
Chapter three
The concept of UX and HCI paradigms

- First, UX is associated with a broad range of fuzzy and dynamic concepts, including emotional, affective, experiential, hedonic, and aesthetic variables. Typical examples of so-called elemental attributes of UX like fun, pleasure, pride, joy, surprise, and intimacy are but a subset of a growing list of human values. Inclusion and exclusion of particular values or attributes seem arbitrary, depending on the author’s background and interest.
- Second, the unit of analysis for UX is too malleable, ranging from a single aspect of an individual end-user’s interaction with a standalone application to all aspects of multiple end-users’ interactions with the company and the merging of the services of multiple disciplines.
- Third, the landscape of UX research is fragmented and complicated by diverse theoretical models with different foci such as emotion, affect, experience, value, pleasure, beauty, etc.

We agree with Law et al. that the above are legitimate reasons for why the concept of UX is so hard to understand. But, we also suggest that rather than considering it as an inconvenience, to see it as a strength. This statement is closely related to the main contribution made in the present paper. We will explain how, after we have provided more background information that helps us to place our discourse in ongoing academic discourses.

In this paper we present and reflect over ten definitions provided in a timespan between 1990 and 2012. The definitions were identified through research on related literature and through browsing the web pages of prominent usability organizations. It is interesting to conclude that despite, or perhaps because of, the variety of diverse viewpoints and suggested definitions of UX, it has been and still is difficult to understand UX as a concept. In this paper our use of the term “definition” includes published descriptions that we have identified to have the same aim as a definition, i.e. to describe the core concepts of UX and to distinguish it from related concepts. For the present authors, there is no doubt that definitions have been important for shaping and isolating the concept of UX, and also helped developing a shared vocabulary of it. Still, it does not seem to be the case that shared definitions is the answer to understanding the concept of UX. Let’s take a look at a recent UX initiative to elaborate our statement.

A strong initiative for bringing clarity to the concept of UX took place in September 2010 at a Dagstuhl seminar held by the Schloss Dagstuhl Centre for Informatics in western Germany. In this seminar, about 30 UX experts were gathered to define the basic and important concepts of UX, in order to differentiate UX from the allied concepts. Judging by the number of participating UX experts in this seminar, this is perhaps the strongest joint face-to-face event aimed at clarifying UX, so far. Many of the participants had also provided prominent individual contributions preceding this event. Hence it is worthwhile to reflect over its results. Before this seminar, the organizers came up with a “white paper” on
UX to provide a starting point for discussing insight about UX (Kuutti 2010). This “User Experience White Paper” was distributed beforehand to the participants of the Dagstuhl UX workshop, and was amended after the seminar. It aimed to describe the core concepts of UX and to distinguish it from related concepts. It was agreed among the UX professionals gathered at the Dagstuhl seminar that a clear description of UX is of help to teach, communicate, clarify, ground and advance UX as a concept in the HCI community. It is worth noticing that at the seminar, consensus on a definition was achieved, not in relation to researchers own personal achievements, but rather through the recently published ISO 9241-210 standard, where UX is defined as a person’s perception and responses that result from the use or anticipated use of a product, system or service (ISO DIS 9241 2010). Hence one important result from the Dagstuhl seminar was obviously the agreement of understanding UX through yet another definition.

In response to the Dagstuhl seminar’s UX white paper aiming at bringing clarity to the concept of user experience, Kari Kuutti (2010), who also was one of the participants of the seminar, raised some interesting points. First and foremost he noticed that the selection of UX features discussed during the seminar was based on practical considerations rather than theoretical analysis. Prediction and usefulness were more valued than explanations and understanding. He also reflected over the means for reaching the results. Conceptual analysis was applied at the seminar, i.e. the development of core concepts was produced inductively, based on the opinions of the participants. Besides admitting that such conceptual analysis is a fully legitimate method for reaching consensus among researchers, Kuutti asked if this consolidation of personal, practice-grounded opinions is really the only way to define what user experience means (p. 716).

Besides the reflections above, Kuutti also identified that, despite the fact that the User Experience White Paper was aimed at defining the basic and important concepts of UX in order to differentiate UX from the allied concepts, it failed to do so. He argued that the result of the white paper, i.e. that it emphasized the ISO 9241-210 standard definition, lacked traceability and enough discussion related to the currently used different definitions, frameworks and models of UX that have been developed by the UX research community. This is a surprising result, as many of the senior UX researchers were part of the seminar. Neither was any mention made to any background theories, nor to any possible need for them. Hence, he rightly raised the question whether these background theories, frameworks and definitions already presented by UX community were not useful enough to be included in that discussion. Kuutti assured that his critique was by no means meant to scorn or ridicule the respectable efforts of organizers and participants; rather the opposite – he was proud of being invited and participating in this seminar full of energy and determination.

Kuutti did what any good scientist does, i.e. reflect over and respectfully question achieved academic results, identify issues and gaps and provide suggestions for alternatives and improvements.

What Kuutti had identified was that besides the practically oriented needs leading up to a shared definition, there also is a need to define the origins of UX as a concept, in order to characterize UX as concept. This was from a theoretical point of view. The need for
conceptual-theoretical knowledge and a commonly agreed concept of user experience is still an unresolved challenge (Kuutti 2010). In his published response to the UX white paper Bringing clarity to the concept of user experience Kuutti makes an analogy between HCI treatment of UX to the situation than astronomy faced in 400-600 B.C., where the results of both Babylonians (practitioners) and Ionians (theorists) were needed for building the foundation of today’s astronomy. Analogically it can be claimed today that the HCI version of the Ionians’ results is not incorporated in the foundation of the concept of UX.

In this paper we continue on Kuutti’s track by asking if there is an element missing in the discourse related to the approach towards a better understanding through UX definitions. As mentioned earlier we identified definitions from a period of twenty years of UX discourses. We chose to include and look more closely at eight UX definitions published in peer-reviewed publications, and at two definitions given by prominent UX actors and organizations. Obviously, many researchers in the HCI research community and practitioners have put their hope in definitions for better explaining UX. Unfortunately the result of this effort has not been as strong as hoped for. Starting in this situation that we have identified, that shared definitions has not been the answer hoped for, we reflect in this paper over what role a definition can have in this effort to strive towards a better understanding of the concept of UX. We do so by emphasizing the need of a clear relation between a UX definition and the paradigm it belongs to, i.e. where the theoretical foundation can be found. We also conclude this paper by suggesting ways to improve the situation.

As mentioned earlier, this paper concerns itself with and revolves around the subject of definitions, and suggests that we must look to the origins of the underlying theoretical foundation. In this paper we discuss why, despite many definitions of UX, it is still difficult to understand UX as concept. Here, we agree with Kuutti’s critique concerning the lack of theoretical discourses discussing the concept of UX in the HCI community. We need to define the origins of UX as a concept, in order to characterize UX as concept. We suggest that theoretical attempts at understanding UX as a concept, including the choice to use definitions, profoundly benefits from clarifying which HCI paradigm is applied. We also reveal how challenging it can be to make an impact with suggested UX concepts outside your own scientific paradigm. Before deepening in the subject of definitions and paradigms we will place measurements and UX in a software engineering context.

3.2 Measurement in Software Engineering

In software engineering, measurement is an important activity that helps us understand what is happening, to control what is happening, and to improve processes and products (Fenton and Pfleeger 1998). One of the ambitions regarding UX is to ensure that it is measurable, in the same way as, for example, usability is measurable, allowing it to become part of the software engineering toolbox.

Software measurement is a way of applying software measures to software engineering objects to achieve predefined goals, which vary according to what is being measured, why
it is being measured, who is interested in the measurement, and the environment where the measurement is performed (Basili, Caldiera et al. 1994). Improving processes, resources and methods, as a help to producing better products, is seen as a driving force in software engineering research, and software measurement is seen as a key aspect of good software management, helping to reach improvement goals (Pfleeger and Atlee 2006). However, software measurement has been characterized as a collection of fringe topics which have not been regarded as mainstream software engineering (Fenton and Pfleeger 1998), and as a result of the complexity and uncertainty involved when implementing metrics activities, companies seem to find it difficult to implement metrics programmes as part of their process improvement activities, even though metrics programmes are also seen as vital when attempting to improve software processes (Iversen and Kautz 2001).

Fenton and Pfleeger (1998) discuss three classes of software entities that we wish to measure. These are: processes; products; and resources. Each of these have internal and external attributes. Internal attributes can be measured by studying the entity in isolation from its behavior, and are measured in terms of the process, resource or product itself. External measures, which are usually harder to measure than internal attributes, can only be measured in terms of the behaviour of the product in its environment. Managers and users are often concerned with external attributes, such as reliability, usability and portability (Fenton and Pfleeger 1998). However, when looking at the range of measurements that are detailed by Pfleeger and Atlee (2006), we see that the majority of measurement models that are discussed are internal measures, such as general sets of measurement for object-oriented systems, measures of coherence, cohesion and coupling for object-oriented systems, internal attributes that affect maintenance characteristics. Although some external measurements, such as those regarding maintainability, are also discussed, these are also mainly discussed in terms of the impact on internal processes, rather than as a direct attempt to increase product quality from a user point of view. As already mentioned in the introduction Pfleeger and Atlee (2006), in a roadmap for future work, discuss Wasserman’s Steps to Maturity, and they state that in the future it will become increasingly important to measure quality on a broader framework, including customer satisfaction and business needs, and other factors beyond faults and failures.

Usability, which has had a strong focus on measurement and metrics, has typically been concerned with measuring efficiency, effectiveness and satisfaction, in accordance with the ISO standards, e.g. (ISO 9241-11 1998), and there is a general agreement from standards boards on these dimensions, although even here, there is less agreement on which metrics should be used to quantify these dimensions (Sauro and Kindlund 2005). Usability has concentrated on an objective, task related view of quality. In many ways, it has concentrated on barriers, frustrations and failures. Usability testing, in common with many other types of testing, has concentrated on identifying and removing the negative. This has been, and still is, an important part of product design and production. In this way, measurement has been used to identify “hygiene factors” (Tietjen and Myers 1998), which are dissatisfiers, whose absence will lead to dissatisfaction, but more important, whose
presence will not necessarily lead to increased satisfaction! The former can be equated with the “must be” factors in the Kano model (Sauerwein, Bailom et al. 1996).

Usability has traditionally been concerned with measurements and metrics, and this is one of the reasons why it has been possible to include usability in the software engineering paradigms. However, even here, where there are clear correlations between the interests and areas, with a focus on measurement and metrics, there have traditionally been gaps that have been and are difficult to bridge.

In many ways, it is still uncertain in what way usability differs from the UX perspective (UXEM 2008), what the connection is between UX and usability, and why the kind of thinking that usability has traditionally been concerned with lacks relevance when looking at UX. Whereas usability has concentrated on the negative and the objective, UX has focused on the holistic, subjective and positive (Hassenzahl, Lai-Chong Law et al. 2006). Both usability and UX can concentrate on efficiency, effectiveness and satisfaction, but UX attempts to understand satisfaction in terms of fun, pride, pleasure and joy, and attempts to understand, define and quantify these elements (Lai-Chong Law, Hvannberg et al. 2006). The focus that is given to the negative in usability work can still be found in UX, but greater emphasis is placed on the fact that the positive is not simply the absence of the negative. The presence of the positive is an indicator of product quality, and high levels of satisfaction can be a result of positive aspects, rather than being a result of the absence of dissatisfying features. This can also be related to the Kano model, where the presence of attractive features can be found to outweigh the absence of “must have” features (Sauerwein, Bailom et al. 1996).

However, when concentrating on UX, particularly when it is at the stage that exists today, we are moving away from a situation where usability has objectively measured product quality and quality in use, to a situation where we attempt to capture subjective factors such as beauty, fun, pleasure and pride. It has been found difficult enough to make use of usability work in software engineering, (see e.g. Winter, Rönkkö et al. 2013) for a discussion of the problems found related to introducing usability KPIs in an industrial situation) even though much usability work has been based on measurement and metrics. As long as there is a situation where there is no consensus about what can be measured, and in which way, then we believe that it will be difficult to incorporate UX in a software engineering paradigm. A finding supported by Bargas-Avila and Hornbaek (2011) recent study of how empirical research on UX has been conducted, identifying that context of use is rarely researched.

3.3 Definitions of UX in HCI

During the last decade more and more researchers have interested themselves in better understanding UX. One approach in that effort has concerned the challenging task of defining UX. This is visible in the ten examples of definitions discussed in this paper. A literature overview including the use of the snowball method, looking at definitions from the years between 1990 and 2010 was done. The definitions discussed were published
between the years 1996 to 2010. Further to the definitions presented in this paper, there also exist many efforts among reputable commercial and voluntary organizations to define UX. These are organizations working with usability, user experience and design that have provided definitions of UX on their web pages. We have included two of these in the discussion. As mentioned earlier the remaining eight of the ten discussed definitions are from peer reviewed papers.

The earliest identified definition was by Alben (1996) *All the aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they’re using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it.*

Unsurprisingly, the UX concepts found in this early definition are influenced by practically focused usability and performance, i.e. *feel in the hand, feel about it, understand how it works, serving purpose, experience and expectations, motivated action, and fit into context.*

The focus on these concepts matches the development in the landscape of HCI research that was initially dominated by Human Factors and Engineering (Harrison, Tatar et al. 2007). It is also interesting to notice how the terms context and experience have been with us from the very start, whilst having a different value and methodological treatment. In this early stage of HCI, those concepts were still primarily perceived as providing yet another set of factors influencing the usage of artifacts under design. Focus was placed on concrete problems that cause disruption, and the phenomena that underlie interaction were not of concern if they did not cause noticeable problems (ibid.).

Five years later, Mäkelä and Fulton Suri (2001) provided another early definition: *A result of motivated action in a certain context. Users’ previous experiences and expectations influence the present experience; this present experience leads to more experiences and modified expectations.*

Shortly after, Henke (2004) identified that user experience can be defined simply as the *sum total of everything that the user experiences while interacting with any aspect of the product.*

In the first definition, a new concept was included, *motivated action,* and more precision was hinted by adding certain to the term context. Also the cycle of users’ previous experiences and expectations affecting new experiences and expectations is highlighted, pointing to the phenomenon that users always come with a baggage of experiences which will unavoidably influence the user experience. The second definition takes a shortcut by referencing the *sum total of everything.*

In the period between 2006 and 2010, the following definitions where found:

Hassenzahl and Tractinsky (2006): *a consequence of a user’s internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.).*
And Jetter and Gerken (2007): UX incorporates not only the traditional qualities like reliability, functionality or usability but also novel and hard-to-grasp concepts from visual or industrial design, psychology or marketing research e.g., attractiveness, stimulation, fun, coolness, sexiness or the successful delivery of the brand proposition.

New concepts, difficult to define and measure, such as attractiveness, stimulation, fun, coolness, sexiness, simplicity, elegance, joy to own and joy to use appeared together with predispositions, expectations, needs, motivation, mood, meaningfulness of the activity, voluntariness of use. These concepts relate to subjective knowledge and personal preferences of users. During these years many HCI researchers took inspiration from psychology to better understand internal personal concepts. At this point of time, psychology included a cognitive view where rationality and rational analysis was emphasized. i.e. how information goes in, what transformations it undergoes, and what is the output. It was a fully legitimate approach to sum up meaning by mapping to the perceived information flow (Harrison, Tatar et al. 2007). Also, context was more specified, through pointing out both organizational and social setting. Related to the previous definitions, it is visible that the development of UX definitions has shifted in emphasis towards the motivation behind, multiple views, and the importance of more precise understanding of situations and contexts. The movement from human factors and engineering to include cognitive influences from psychology and marketing research is also clearly spelled out in one of the definitions. Other identified definitions are:

Desmet and Hekkert (2007): The entire set of affects that is elicited by the interaction between a user and a product including the degree to which all our senses are gratified (aesthetic experience) the meanings we attach to the product (experience of meaning) and the feelings and emotions that are elicited (emotional experience).

Sward and Macarthur (2007): the value derived from interaction(s) [or anticipated interaction(s)] with a product or service and the supporting cast in the context of use (e.g. time, location, and user disposition).

Hassenzahl (2008): UX as a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service. Good UX is the consequence of fulfilling the human needs for autonomy, competency, stimulation (self-oriented), relatedness, and popularity (others-oriented) through interacting with the product or service (i.e., hedonic quality). Pragmatic quality facilitates the potential fulfilment of be-goals.

Once again, new and hard to grasp concepts are introduced, such as elicited by the interaction between a user and a product, including aesthetic experience, experience of meaning, together with emotional experience. UX is related to the value derived from interaction(s) or anticipated interaction(s), related to aspects of context such as time, location, and user disposition. Furthermore there is the addition of good-bad feeling when interacting, self-oriented, concepts such as autonomy, competency, stimulation, and others-oriented, concepts such as relatedness, and popularity, as well as hedonic quality of experience. Now the fast growing pile of theoretical concepts needed to understand the
concept of UX has become a huge challenge. Yes, the concept of UX is maturing as it becomes richer but also more and more complicated.

Two recent web definitions are:

Usability Body of Knowledge, UPA (2012): User Experience (UE). Every aspect of the user's interaction with a product, service, or company that make up the user's perceptions of the whole. User experience design as a discipline is concerned with all the elements that together make up that interface, including layout, visual design, text, brand, sound, and interaction. UE works to coordinate these elements to allow for the best possible interaction by users.

Nielsen-Norman Group (2012) defines UX as all aspects of the end-user's interaction with the company, its services, and its products. The first requirement for an exemplary user experience is to meet the exact needs of the customer without fuss or bother. Next come simplicity and elegance that produce products that are a joy to own, a joy to use. True user experience goes far beyond giving customers what they say they want or providing checklist features.

It is noticeable that both actors above emphasize the entirety in their definitions, by using the terms every and all. In one of the definitions the best possible is emphasized, and in the other to meet the exact needs of. It is clear that it is the subject of UX investigation that also defines what is good UX.

From an overview point of view of all definitions above: The concept of interaction mentioned in the last five definitions was first mentioned in the definition from 2004, and is later mentioned in the majority of the identified definitions. As will be discussed later in section 4, interaction can be approached through the use of many different lenses, leading to very different understandings of the concept. As mentioned, both latter definitions frame the concept of UX by applying general high level terminology, i.e. every aspect, the whole, all elements, that together, all aspects of, beyond checklist. In fact, when reflecting over all of the identified definitions, many more use similar high-level terminology such as: all the aspects of, sum total of everything, the entire set of, all aspect of, feeling, the sum of, result from. Other definitions emphasized a consequential relationship: a result of, a consequence of, the quality of, the value derived from, feels about, a person has, result from, perceived by, perceives. Taken together, this indicates that UX is either still under elaboration, or that UX is not possible to specify. Or at least not possible to specify it without more specific concrete references to its context and framing, which will be different also dependent on researchers belong in the HCI paradigm map, which is one point we make in this paper. Or, as Kuutti puts it (2011), that besides the practical oriented needs leading up to a shared definition, there is also a need to define the origins of UX as a concept in order to characterize UX as a concept. This is something that has to be done from a theoretical point of view, and it is still an unresolved challenge (Kuutti 2010). In this paper we approach this subject in greater depth, and elaborate what such an effort might mean. One way to remedy the situation could be to be clear about which paradigm’s view is applied to understand a
concept, e.g. Human factors, Engineering, Psychology, or a situated paradigm. This will be discussed in next section.

It can be seen that the growing richness of UX concepts slowly creates a more open mindset towards, and a need for the inclusion of, approaches and lenses from other sources of definitions. Despite this, between the years 2006-2007 research approaches and lenses such as ethnomethodology, value-sensitive design, and ethnography, addressing concepts such as understanding people’s own methods for sense making, situated meaning, embodiment, values, and social issues - were still mainly discussed in corners and cafes, rather than being part of the HCI conference agendas. That is the situated paradigm when using Harrisson, Tatar and Sengers terminology (2007). These research approaches have slowly become more and more accepted even in mainstream HCI conferences since 2007, but can still to be considered as a minority. This peripheral placement of the situated paradigm in HCI, a paradigm that has heavy focus on context and situation, also indicates that context and framing is not fully accepted or appreciated yet.

3.4 Paradigms Understood as Generative Metaphors

In order to elaborate on definitions and the challenge of understanding “UX as concept” we will introduce the idea of the generative metaphor, together with an overview of three suggested intellectual directions or paradigms in HCI as presented by Harrison, Tattar and Sengers (2007). We will not present the paradigms in any depth; instead we present just as much information as we believe is necessary to make our point in this paper. Extensive details of these paradigms can be found in Harrison, Tattar and Senger’s paper. Similar development of different generative metaphors can also be found in software engineering. Harrison, Tattar and Sengers presented in 2007 how the intellectual field of HCI has evolved in at least three separate intellectual directions. HCI started as an atheoretical and entirely pragmatic human factors and engineering focused field, which was later expanded by the cognitive view, with a focus on abstract reasoning and theories, emphasizing models and deductive proofs. More recently, it has also included a phenomenological view that insists on putting the interaction itself in the center, where meaning and meaning-making is subjective and related to multiple interpretations based on multiple mutually defining factors. These are three different developments of interest within HCI, which all have their own distinct theoretical base. Each of these constitutes examples of different theoretical foundations which will influence the understanding of concepts such as UX differently.

The first intellectual direction or paradigm, called Human Factors and Engineering, focused on man-machine coupling and targeted how to optimize the fit between man and machine, often through a pragmatic atheoretical approach. In this community, research is often based on observations, and the question of interest surrounds how to fix specific problems that arise in interaction. The second paradigm called the Cognitive Revolution, focuses on the communication between man and machine and strives to optimize the accuracy and efficiency of information transfer. Research here is often based on experiments and typically includes the identification of mismatches, how to accurately model information.
flow and the human mind, as well as how to improve the efficiency of computers. The third paradigm, Situated Perspectives, investigates interaction as situated phenomenon, with the aim to support it in the world. Here research is often based on an ethnographic approach; it typically includes how users appropriate technology, what are the politics and values at the site of interaction, and how these can these be supported.

Each of the three parallel paradigms in the HCI community has unique understandings of what science is, how to go about being scientific, as well as what is worthwhile to study. By paying attention to the details and the difference in nature in relation to each of these paradigms, Harrison et al realized that the traditional mechanisms in Kuhn’s understanding of the paradigm (Kuhn 1970) is not at work in the HCI community. The main reason being that the idea of having shared access to a force which can act as an undisputable determiner to verify and validate scientific results against does not exist when studying humans and technology in HCI, as opposed to the laws of nature in physics. Instead they found that researchers themselves in each of the three paradigms had developed their own expectations on scientific rules, research methods and outcomes.

To explain the development of science in the HCI community, Harrison, Tattar and Sengers applied Agre’s theory of generative metaphors in technical work (Agre 1997, pp. 33-48). In this view, knowledge development and research in HCI are structured around particular generative metaphors. In comparison to Kuhn’s paradigm, Agre’s generative metaphor offers a different explanation of why and how researchers’ change of research interest takes place. Central to a generative metaphor is the introduction of …‘centers’ and ‘margins’ that drive choices about what phenomena constitute important descriptive qualities of interactions, what questions are interesting to ask about interaction, what methods are appropriate for studying and designing interaction and what validation procedures are requested to establish knowledge claims about interaction. (Harrison, Tatar et al. 2007, p.3). In this view, paradigm shifts (keeping to the term paradigm) do not occur based on undisputable repeatable scientific proof. Instead, the shift of focus occurs when some interest is spreading in popularity in the research community. An interesting attribute of these metaphors is that they do not strictly dictate what is of relevance; instead it is a joint movement of interest that brings some phenomena into the centre whilst marginalizing other phenomenon. Shifts driving new choices of discourse of what to do research on are often based on needs. This is not always consciously acted out; sometimes it spreads more like a trend from many actors, who are not always strongly connected. The shift of interest in this kind of generative paradigm can be identified when challenges that used to be marginalized appear to be the central discourse. In the case of UX, we have pressure from businesses industry saying that usability is not enough anymore. Today the market expects usability to be good as default and buyers will be irritated if it is not, but what really sells products is a good user experience. Relating this to the Kano model, usability has become a must-be requirement, whilst UX has become an attractive requirement, the sort of thing that delights the customer and that has a great impact on customer satisfaction (Sauerwein, Bailom et al. 1996).
In contrast to Kuhn’s paradigms, Agre’s generative metaphors can exist side-by-side. If we want to keep the term paradigm (as Harrison, Tattar and Sengers did in their paper (2007) and which we decided to do in the present paper) we need to accept that the generative metaphors in HCI offer a different understanding or model of the paradigm and the reasons for shifts.

3.5 Centers and Margins in the three paradigms of HCI

In this section we will go right to the point of comparing the identified centers and margins as they were presented by Harrisson et al in 2007, i.e. compare the situated paradigm with the human factors and engineering paradigm and the cognitive paradigm. Note that, even though we agree with the authors’ message, here we simply make use of their paper to make our point about different co-existing scientific foundations. We realize that much has happened in each community since 2007, and that different researchers might have painted the landscape of paradigms or generative metaphors in different ways. Our ambition is not to elaborate or update Harrisson et al’s contribution; we merely want to borrow the published and available idea of paradigms and generative metaphors to make our point about definitions of UX in this paper. So let’s start with right on the point comparison, contrasting the three different paradigm views.

In the situated paradigm, meaning and meaning-making is at the heart, as compared to the Human factors and Engineering paradigm, that tended to ignore this, as long as it did not cause problems. In the latter, the goal was to optimize the fit between humans and machines, and the questions to be answered focused on identifying problems of coupling between man and machine, and developing pragmatic solutions to them. At the center of the Human Factors and Engineering paradigm has been concrete problems that cause disruption; the phenomena that underlie interaction have not been of concern as long as they did not cause noticeable troubles. In the situated Paradigm, on the other hand, the latter is of paramount concern. The situated paradigm adopts a view where meaning is constructed on the fly, often collaboratively, by people acting in contexts and situations. Hence, focus on the interaction itself is of central importance for understanding the process of meaning construction. An interesting consequence of this stance is that meaning cannot meaningfully be summed up by mapping information flow, which is a focus in the cognitive paradigm. This is because, in the situated view, meaning is irreducibly connected to the studied people’s personality, which has arisen out of a unique history, the situation at hand, future endeavours, context, community and human institutions, culture and subcultures, together with related human and technical systems with which people interact. Hence, the situated view also differs substantially from the Human Factors and Engineering view, where context was primarily perceived as providing yet another set of factors influencing the usage of artifacts under design. In the Situated paradigm, context is a key component for understanding, designing and evaluating results.

Designing for human computer interaction and UX in the situated view means to overcome the challenge of establishing an understanding of multiple, and sometimes conflicting, views held both by diverse stakeholders and users. The relevant knowledge and meanings
will be found in the local and situated circumstances perceived from the different actors’ own points of view. This is a view which is in resonance with the challenges inherent in the concept of wicked problems (Rittel and Webber 1973). It is also acknowledged that, whilst it is impossible to design UX for all possibilities, both the meaning and nature of the interaction itself will be defined by the specifics of particular contexts. This can again be contrasted with the cognitive paradigm, where rationality and rational analysis is emphasized, i.e. how information goes in, what transformations it undergoes, and what is the output. Here, phenomena that are difficult to assimilate, such as peoples’ feelings, situational and contextual information, will be placed in the margin.

Adapting a situated view of meaning and meaning making also obliges the researcher to ask questions about what it means for a design to be good from the point of view of the people who are studied, which is about understanding user values. This is an act that opens up for a variety of viewpoints which are all potentially equally valid. Hence, a consequence is that evaluations also by necessity need to be linked to the many values of many people. Measures claimed to be universally valid are no longer enough. This lead to problems of evaluations that are comparable to other evaluations. If we again compare with the cognitive paradigm, that emphasizes the importance of using theory and data for further theory building, the situated paradigm sees that theory has its limitations. In fact, in the situated paradigm, theory is perceived as a resource for analyzing and making sense of what is going on at the site of interaction. This together with context, which is perceived as being an equally important resource, and sometimes an even more important resource than theory.

3.6 Discussion and conclusion

The point we want to make clear in this paper is that these centers and margins are central for the understanding of UX as a concept. Take the concept interaction, which is included many of the definitions, what does it actually mean for the understanding of UX? The answer is that it depends. From a Situated view interaction is closely related to the process of meaning construction. Meaning is constructed on the fly, often collaboratively, by people who always exist in contexts and specific situations. Here, it would be possible to explain UX with the help of pure and rich field descriptions, capturing the interplay with situation and context, without the necessity of applying any type of theory. In this view it could not be meaningfully summed up by mapping information flow. From a Cognitive paradigm point of view, it could be meaningfully summed up by mapping information flow, and it would be explained by applying external theory and be part of further theory building rather than providing detailed relations to context and situation. From a Human Factors and Engineering point of view, interaction or UX and their intertwinement with context could be safely ignored for all practical purposes as long as they did not cause any troubles. Context would constitute yet another set of factors influencing the usage of artifacts under design. It is obvious that each paradigm opens up for different possible understandings of the concept of interaction as well as UX. Without traceability and sufficient discussion related to the underlying frameworks and models of UX that have
been developed in and by the different UX researchers, we end up either in a situation of confusion, or with a lack of clear and powerful contributions.

As presented above, there will be different consequences when approaching UX from the point of view of different and contrasting paradigms, such as in cases:

A) where one paradigm places concrete problems that cause disruption in the center, whilst viewing meaning and meaning making that underlie interaction to be of less importance as long as they do not cause noticeable trouble, in contrast to another paradigm that places meaning and meaning making at the heart, and focuses on the interaction itself for understanding the process of meaning making.

B) if one paradigm places theory in the forefront, whilst at the same time perceiving context as yet another set of factors influencing the usage of artifacts under design, in contrast to another paradigm that sees context as being an equally important or even more important resource than theory for analyzing and making sense of what is going on at the site of interaction.

C) if the understanding of theory in one paradigm relates to generalizable theory, in contrast to another paradigm where theory is seen as being similar to what Merton (1957) identified as theories of middle range.¹

To use a metaphor borrowing an old saying from Garfinkel, the act of not considering the underlying paradigm seriously when understanding the concept of UX would be very much like complaining that if the walls of a building were gotten out of the way, one could see better what was keeping the roof up (Garfinkel 1967, p. 22). Of course, UX definitions have many independent and general positive qualities. Definitions could function as boundary objects (Star and Griesemer 1989) between HCI researchers with different ontologies, belonging to different research paradigms. They could provide pointers to specific aspects that ease discussions and the placement of work and work results. As boundary objects they could ease mutual learning and knowledge integration between different actors across interdisciplinary borders. They could also contribute as a learning vehicle for beginners, which would become unnecessary when they became more expert. In any case, they would require a clear positioning of researchers’ backgrounds and to which paradigm they belong.

Bargas-Avila and Hornbaek’s (2011) paper showed that context of use is rarely dealt with in empirical research of UX. This could mean either that it is too complicated to be handled in empirical research, or that the situated paradigm for some reason is still not fully

¹ Middle range theories were originally explained as: theories that lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization, and social change ... It is intermediate to general theories of social systems which are too remote from particular classes of social behavior, organization, and change to account for what is observed and to those detailed orderly descriptions of particulars that are not generalized at all (1957).
accepted in HCI. Ethnographic research on revealing details from context in relation to UX would constitute a fully legitimate HCI subject in publications if the situated paradigm was fully accepted. In the situated paradigm, context is a key component for understanding, designing and evaluating results UX.

If we reflect upon the results of the Dagstuhl Seminar, there were many senior UX researchers who were part of the seminar who neither advocated their background theories, nor saw any possible need for them in the process of understand UX. One reason for this could be that it is hard and takes a lot of time to share the fundamentals of one’s own paradigm with others who have another foundation from another paradigm. It would require a lot of time and reflection to understand and discuss the consequences of similarities and differences, which this seminar most probably was not prepared for and had not scheduled time for.

We finish by asking why this discussion is of importance for a field such as software engineering. We have seen the difficulties of creating and using definitions in the HCI community, and even on reaching agreement on and understanding of the meaning of definitions. Until we reach some kind of agreement regarding the meaning and use of definitions, then we cannot reach a stage where UX becomes a meaningful tool in the software engineering toolbox, allowing us to measure and make use of this important and highly relevant concept.

To summarize, our argumentation points to the need for a much richer debate on which underlying theoretical foundation is applied when discussing UX definitions and concepts, where, whichever ontological orientation researchers have, they all have an important contribution to make.

Acknowledgements

The challenge of understanding the concept of UX and the interest in its definitions started with two separate master projects. Two excellent and energetic students, Ali Mansoor and Tabassum Riaz, wanted to write a paper based on a literature review of UX definitions after having successfully defended their respective master theses discussing UX within software engineering. For some reason, despite many attempts and intensive hours of thinking, that paper was never produced. Intuitively something felt very wrong with the identified focus of the paper, i.e. categorizing, classifying and building framework for UX definitions. A “why?” started gnaw in our minds, and later ideas for this paper slowly grew out of the need to sort out why the first paper was never produced. Thanks, Ali and Tabassum, for implicitly forcing us to take the subject of UX definitions one step further. We are also sorry that we at that time were unable to explain to you why the first idea for the paper did not work out.
Chapter Four

Article Five
Chapter Four

Satisfying Stakeholders’ Needs
– Balancing Agile and Formal Usability Test Results


Jeff Winter, Kari Rönkkö

Abstract
This paper deals with a case study of testing with a usability testing package (UTUM), which is also a tool for quality assurance, developed in cooperation between industry and research. It shows that within the studied company, there is a need to balance agility and formalism when producing and presenting results of usability testing to groups who we have called Designers and Product Owners. We have found that these groups have different needs, which can be placed on opposite sides of a scale, based on the agile manifesto. This becomes a Designer and a Product Owner Manifesto. The test package is seen as a successful hybrid method combining agility with formalism, satisfying organisational needs, and fulfilling the desire to create a closer relation between industry and research.

4.1 Introduction
Product quality is becoming the dominant success criterion in the software industry, and Osterweil (1996) states that the challenge for research is to provide the industry with the means to deploy quality software, allowing companies to compete effectively. Quality is multi-dimensional, and impossible to show through one simple measure, and research should focus on identifying various dimensions of quality and measures appropriate for it (Osterweil 1996). A more effective collaboration between practitioners and researchers would be of great value (Osterweil 1996). Quality is also important owing to the criticality of software systems (a view supported by Harrold (2000) in her roadmap for testing) and even to changes in legislation that make executives responsible for damages caused by faulty software. One approach to achieving quality has been to rely on complete, testable and consistent requirements, traceability to design, code and test cases, and heavyweight documentation. However, a demand for continuous and rapid results in a world of continuously changing business decisions often makes this approach impractical or impossible, pointing to a need for agility. At a keynote speech at the 5th Workshop on Software Quality, held at ICSE 2007 (WoSQ 2007), Boehm stated that both agility and quality are becoming more and more important. Many areas of technology exhibit a tremendous pace of change, due to changes in technology and related infrastructures, the dynamics of the marketplace and competition, and organisational change. This is particularly obvious in mobile phone development, where their pace of development and

:: 113 ::
penetration into the market has exploded over the last 5 years. This kind of situation demands an agile approach (Boehm 2007).

This article is based on two case studies of a usability evaluation framework called UIQ Technology Usability Metrics (UTUM) (UIQ Technology 2006), the result of a long research cooperation between the research group “Use-Oriented Design and Development” (U-ODD) (U-ODD 2011) at Blekinge Institute of Technology (BTH), and UIQ Technology (UIQ) (UIQ Technology 2008). With the help of Martin, Rooksby et al.’s (2007) study and our own case studies, it presents an approach to achieving quality, related to an organizational need for agile and formal usability test results. We use concepts such as “agility understood as good organizational reasons” and “plan driven processes as the formal side in testing”, to identify and exemplify a practical solution to assuring quality through an agile approach. The research question for the first case study was:

– How can we balance demands for agile results with demands for formal results when performing usability testing for quality assurance?

We use the term “formal” as a contrast to the term “agile” not because we see agile processes as being informal or unstructured, but since “formal” is more representative than “plan driven” to characterise the results of testing and how they are presented to certain stakeholders. We examine how the results of the UTUM test are suitable for use in an agile process. eXtreme Programming (XP) is used as an illustrative example in this article, but note that there is no strong connection to any particular agile methodology; rather, there is a philosophical connection between the test and the ideas behind the agile movement. We examine how the test satisfies requirements for formal and informal statements of usability and quality.

In the first study, we identify two groups of stakeholders that we designated as Designers (D) and Product Owners (PO), with an interest in the different elements of the test data. A further case study was performed to discover if these findings could be confirmed. It attempted to answer the following research questions:

– Are there any presentation methods that are generally preferred?
– Is it possible to find factors in the data that allow us to identify differences between the separate groups (D & PO) that were tentatively identified in the case study presented in the previous study?
– Are there methods that the respondents think are lacking in the presentation methods currently in use within UTUM?
– Do the information needs, and preferred methods change during different phases of a design and development project?
– Can results be presented in a meaningful way without the test leader being present?

The structure of the article is as follows. An overview of two testing paradigms is provided. A description of the test method comes next, followed by a presentation of the methodology, and the material from the case studies, examining the balance between agility
and formalism, the information needs of different stakeholders, the relationship between agility, formality and quality, and the need for research/industry cooperation.

The article ends with a discussion of the work, and conclusions.

4.2 Testing – Prevailing Models vs. Agile Testing

Testing is performed to support quality assurance, and an emphasis on software quality requires improved testing methodologies that can be used by practitioners to test their software (Harrold 2000). Since we regard the test framework as an agile testing methodology, this section presents a discussion of testing from the viewpoints of both the software engineering community and the agile community.

Within software engineering, there are many types of testing, in many process models, (e.g. the Waterfall model (Royce 1987), Boehm’s Spiral model (Boehm 1988)). Testing is often phase based, and the typical stages of testing (see e.g. Pfleeger and Atlee 2006; Sommerville 2007) are Unit testing, Integration testing, Function testing, Performance testing, Acceptance testing, and Installation testing. The stages from Function testing and onwards are characterised as System Testing, where the system is tested as a whole rather than as individual pieces (Pfleeger and Atlee 2006). Usability testing (otherwise named Human Factors Testing) has been characterised as investigating requirements dealing with the user interface, and has been regarded as a part of Performance testing (Pfleeger and Atlee 2006). The prevailing approach to testing is reliant on formal aspects and best practice.

Agile software development changes how software development organisations work, especially regarding testing (Talby, Hazzan et al. 2006). In agile development, exemplified here by XP (Beck 2000), a key tenet is that testing is performed continuously by developers. Tests should be isolated, i.e. should not interact with the other tests that are written, and should preferably be automatic, although not all companies applying XP automate all tests (Martin, Rooksby et al. 2007). Tests come from both programmers and customers, who create tests that serve to increase their confidence in the operation of the program. Customers specify functional tests to show that the system works how they expect it to, and developers write unit tests to ensure that the programs work how they think it does. These are the main testing methods in XP, but can be complemented by other types of tests when necessary. Some XP teams may have dedicated testers, who help customers translate their test needs into tests, who can help customers create tools to write, run and maintain their own tests, and who translate the customer’s testing ideas into automatic, isolated tests (Beck 2000).

The role of the tester is a matter of debate. It is primarily developers who design and perform testing. However, within industry, there are seen to be fundamental differences between the people who are “good” testers and those who are good developers. In theory, it is often assumed that the tester is also a developer, even when teams use dedicated testers. Within industry, however, it is common that the roles are clearly separated, and that testers are generalists with the kind of knowledge that users have, who complement the perspectives and skills of the testers. A good tester can have traits that are in direct contrast
with the traits that good developers need (see e.g. Pettichord 2000) for a discussion regarding this). Pettichord claims that good testers think empirically in terms of observed behaviour, and must be encouraged to understand customers’ needs. Thus, although there are similarities, there are substantial differences in testing paradigms, how they treat testing, and the role of the tester and test designer. In our testing, the test leaders are specialists in the area of usability and testing, and generalists in the area of the product and process as a whole.

4.3 The UTUM Usability Evaluation Framework

UTUM is a usability evaluation framework for mass market mobile devices, and is a tool for quality assurance, measuring usability empirically on the basis of metrics for satisfaction, efficiency and effectiveness, complemented by a test leader’s observations. Its primary aim is to measure usability, based on the definition in ISO 9241-11, where usability is defined 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 9241-11 1998). This is similar to the definition of quality in use defined in ISO 9126-1, where usability is instead defined as understandability, learnability and operability (ISO/IEC 9126-4 2004). The intention of the test is also to measure “The User eXperience” (UX), which is seen as more encompassing than the view of usability that is contained in e.g. the ISO standards (UIQ Technology 2006), although it is still uncertain how UX differs from the traditional usability perspective (UXEM 2008) and exactly how UX should be defined (for some definitions, see e.g. Hassenzahl, Lai-Chong Law et al. 2006; Hassenzahl and Tractinsky 2006; UXNet: 2008).

In UTUM testing, one or more test leaders carry out the test according to predefined requirements and procedure. The test itself takes place in a neutral environment rather than a lab, in order to put the test participant at ease. The test is led by a test leader, and it is performed together with one tester at a time. The test leader welcomes the tester, and the process begins with the collection of some data regarding the tester and his or her current phone and typical phone use. Whilst the test leader is preparing the test, the tester has the opportunity to get acquainted with the device to be tested, and after a few minutes is asked to fill in a hardware evaluation, a questionnaire regarding attitudes to the look and feel of the device.

The tester performs a number of use cases on the device, based on the tester’s normal phone use or organisational testing needs. The test leader observes what happens during the use case performance, and records any observations, the time taken to complete the use cases, and answers to follow-up questions that arise. After the use case is complete, the tester answers questions about how well the telephone lets the user accomplish the use case.

When all of the use cases are completed, the tester completes a questionnaire based on the System Usability Scale (SUS) (Brooke 1986) about his or her subjective impressions of how easy the interface is to use. It expresses the tester’s opinion of the phone as a whole.
The tester is finally thanked for their participation in the test, and is usually given a small gift, such as a cinema ticket, to thank them for their help.

The data obtained are transferred to spreadsheets. These contain both quantitative data, such as use case completion times and attitude assessments, and qualitative data, such as comments made by testers and information about problems that arose. The data is used to calculate metrics for performance, efficiency, effectiveness and satisfaction, and the relationships between them, leading to a view of usability for the device as a whole. The test leader is an important source of data and information in this process, as he or she has detailed knowledge of what happened during testing.

Figure 4.1 illustrates the flow of data and knowledge contained in the test and the test results, and how the test is related to different groups of stakeholders. Stakeholders, who can be within the organisation, or licensees, or customers in other organisations, can be seen at the top of the flow, as interested parties. Their requirements influence the design and contents of the test. The data collected is found both as knowledge stored in the mind of the test leader, and as metrics and qualitative data in spreadsheets.

The results of the testing are thereby a combination of metrics and knowledge, where the different types of data confirm one another. Metrics based material is presented in the form of diagrams, graphs and charts, showing comparisons, relations and tendencies. This can be corroborated by the knowledge possessed by the test leader, who has interacted with the testers and who knows most about the process and context of the testing. Knowledge
material is often presented verbally, but can if necessary be supported and confirmed by metrics and visual presentations of the data.

UTUM has been found to be a customer driven tool that is quick and efficient, is easily transferable to new environments, and that handles complexity (Winter, Rönkkö et al. 2007). For more detailed information on the contents and performance of the UTUM test and the principles behind it, see (UIQ Technology 2006; Winter, Rönkkö et al. 2007). A brief video presentation of the whole test process (6 minutes) can be found on YouTube (BTH 2008).

4.4 The Study Methodology and the Case Studies

This work has been part of a long-term research cooperation between U-ODD and UIQ, which has centred on the development and evaluation of a usability evaluation framework (for more information, see e.g. Winter, Rönkkö et al. 2007; UIQ Technology 2008). The case studies in this phase of the research cooperation were based on tests performed by together by UIQ in Ronneby, and by Sony Ericsson Mobile Development in Manchester.

The process of research cooperation is action research (AR) according to the research and method development methodology called Cooperative Method Development (CMD), see (Dittrich 2002; Dittrich, Rönkkö et al. 2005; Rönkkö 2005, Chapter 8; Dittrich, Rönkkö et al. 2007) for further details. AR “involves practical problem solving which has theoretical relevance” (Mumford 2001, p. 12). It involves gaining an understanding of a problem, generating and spreading practical improvement ideas, applying the ideas in a real world situation and spreading the theoretical conclusions within academia (Mumford 2001). Improvement and involvement are central to AR, and its purpose is to influence or change some aspect of whatever the research has as its focus (Robson 2002, p. 215). A central aspect of AR is collaboration between researchers and those who are the focus of the research. It is often called participatory research or participatory action research (Robson 2002, p. 216). CMD is built upon guidelines that include the use of ethnomethodological and ethnographically inspired empirical research, combined with other methods if suitable. Ethnography is a research strategy taken from sociology, with foundations in anthropology (Rönkkö 2010). It relies upon the first-hand experience of a field worker who is directly involved in the setting that is under investigation (Rönkkö 2010). CMD focuses on shop floor development practices, taking the practitioners’ perspective when evaluating the empirical research and deliberating improvements, and involving the practitioners in the improvements. This approach is inspired by a participatory design (PD) perspective. PD is an approach towards system design in which those who are expected to use the system are actively involved and play a critical role in its design. It includes stakeholders in design processes, and demands shared responsibility, participation, and a partnership between users and implementers (Schuler and Namioka 1993).

These studies have been performed as case studies, defined by Yin as “an empirical enquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” (Yin 2003, p. 13). Yin presents a number of criteria that are used to establish the quality of empirical social
Chapter Four
Satisfying Stakeholders’ needs

research and states that they should be applied both in the design and conduct of a case study. They deal with construct validity, internal validity, external validity and reliability (Yin 2003, pp. 35–39).

Three tactics are available to increase construct validity, which deals with establishing correct measures for the concepts being studied, and is especially problematic in case study research. These are: using multiple sources of information; ensuring a chain of evidence and; using member checking, i.e. having the key participants review the case study report.

In this study, we have used many different sources of information. The data was obtained through observation, through a series of unstructured and semi-structured interviews (Robson 2002), both face-to-face and via telephone, through participation in meetings between different stakeholders in the process, and from project documents and working material. The interviews have been performed with test leaders, and with staff on management level within the two companies. Interviews have been audio taped, and transcribed, and all material has been stored. The second case study involves the use of a survey. The mix of data and collection methods has given a triangulation of data that serves to validate the results that have been reached. To ensure a chain of evidence a “study database” or research diary has been maintained. It collects all of the information in the study, allowing for traceability and transparency of the material, and reliability (Yin 2003). It is mainly computer based, and is an account of the study recording activities performed in the study, transcriptions of interviews and observation notes, and records of relevant documents and articles. The audio recordings are also stored digitally. The written document contains notations of thoughts concerning themes and concepts that arise when reading or writing material in the account of the study. The chain of evidence is also a part of the writing process.

The most important research collaborators in the industrial organisation have been an integral part of the study, and have been closely involved in many stages of the work. They have been available for testing thoughts and hypotheses during the study, giving opportunities for member checking. They have also been involved as co-authors when writing articles, which also means that member-checking has been an integral part of the research.

Internal validity is especially important in exploratory case studies, where an investigator tries to determine whether one event leads to another. It must be possible to show that these events are causal, and that no other events have caused the change. If this is not done, then the study does not deal with threats to internal validity. Some ways of dealing with this are via pattern matching, explanation building, addressing rival explanations, and using logic models. This study has been a mix of exploratory and explanatory studies. To address the issues of internal validity in the case studies, we have used the general repertoire of data analysis as mentioned in the previous paragraph. The material in the research diary has been analysed to find emerging themes, in an editing approach that is consistent with Grounded Theory (see (Robson 2002) p. 458). The analysis process has affected the further rounds of questioning, narrowing down the focus, and shifting the main area of interest, opening up for the inclusion of new respondents who shed light on new aspects of the
study. A further method for ensuring validity has been through discussions together with research colleagues, giving them the chance to react to the analysis and suggest and discuss alternative explanations or approaches.

External validity, knowing whether the results of a case study are generalisable outside the immediate case study, has been seen as a major hinder to doing case studies, as single case studies have been seen as a poor basis for generalisation. However, this is based on a fallacious analogy, where critics contrast the situation to survey research, where samples readily generalise to a larger population. In a case study, the investigator tries to generalise a set of results to a wider theory, but, generalisation is not automatic, and a theory must be tested by replicating the findings, in much the same way as experiments are replicated. Although Yin advises performing multiple-case studies, since the chances of doing a good case study are better than using a single-case design (Yin 2003, p. 53), this study has been performed as a single-case study and has been performed to generate theory. The case here represents a unique case (Yin 2003, p. 40), since the testing has mainly been performed within UIQ, and it is thereby the only place where it has been possible to evaluate the testing methodology in its actual context. One particular threat is in our study is therefore that most of the data comes from UIQ. Due to close proximity to UIQ, the interaction there has been frequent and informal, and everyday contacts and discussions on many topics have influenced the interviews and their analysis. Interaction with Sony Ericsson has been limited to interviews and discussions, but data from Sony Ericsson confirms what was found at UIQ. A further threat is that most of the data in the case study comes from informants who work within the usability/testing area, but once again, they come from two different organisations and corroborate one another, have been complemented by information from other stakeholders, and thus present a valid picture of industrial reality.

A threat in the second case study is the fact that only ten people have participated. This makes it difficult to draw generalisable conclusions from the results. Also, since the company is now disbanded, it is not possible to return to the field to perform cross checking with the participants in the study. The analysis is therefore based on the knowledge we have of the conditions at the company and the context where they worked, and is supported by discussions with a people who were previously employed within the company, whom we are still in contact with. These people can however mainly be characterised as Designers, and therefore may not accurately reflect the views of Product Owners.

Thus, since this research is mainly based on a study of one company in a limited context, it is not possible to make confident claims about the external validity of the study. However, we can say that we have created theory from the study, and that readings appear to suggest that much of what we have found in this study can also be found in other similar contexts. Further work remains to see how applicable the theory is for other organisations in other or wider contexts. Extending the case study and performing a similar study in another organisation is a way of testing this theory, and further analysis may show that the case at UIQ is actually representative of the situation in other organisations.
Reliability deals with the replicability of a study, whereby a later investigator should be able to follow the same procedures as a previous investigator, and arrive at the same findings and conclusions. By ensuring reliability you minimize errors and bias in a study. One prerequisite for this is to document procedures followed in your work, and this can be done by maintaining a case study protocol to deal with the documentation problem, or the development of a case study database. The general way to ensure reliability is to conduct the study so that someone else could repeat the procedures and arrive at the same result (Yin 2003, pp. 35–39). The case study protocol is intended to guide the investigator in carrying out the data collection. It contains both the instrument and the procedures and general rules for data collection. It should contain an overview of the project, the field procedures, case study questions, and a guide for the case study report (Yin 2003, p. 69). As mentioned previously, a case study database has been maintained, containing the most important details of the data collection and analysis process. This ensures that the study is theoretically replicable. One problem regarding the replicability of this study, however, is that the rapidly changing conditions for the branch that we have studied mean that the context is constantly changing, whereby it is difficult to replicate the exact context of the study.

In the following, we begin by presenting the results of the first case study, and discuss in which way the results are agile or plan-driven/formal, who is interested in the different types of results, and which of the organisational stakeholders needs agile or formal results.

4.5 Agile or Formal

The first focus of the study was the fact that testing was distributed, and the effect this had on the testing and the analysis of the results. During the case study, as often happens in case studies (Yin 2003), the research question changed. Gradually, another area of interest became the elements of agility in the test, and the balance between the formal and informal parts of the testing. The framework has always been regarded as a tool for quality, and verifying this was one purpose of the testing that this case study was based on. Given the need for agility mentioned above, the intention became to see how the test is related to agile processes and whether the items in the agile manifesto can be identified in the results from the test framework. The following is the result of having studied the material from the case study from the perspective of the spectrum of different items that are taken up in the agile manifesto.

The agile movement is based on core values, described in the agile manifesto (The Agile Alliance 2001), and explicated in the agile principles (The Agile Alliance 2001). The agile manifesto states that: “We are uncovering better ways of developing software by doing it and by helping others do it. Through this work we have come to value: Individuals and interactions over processes and tools, Working software over comprehensive documentation, Customer collaboration over contract negotiation, and Responding to change over following a plan. That is, while there is value in the items on the right, we value the items on the left more”. Cockburn stresses that the intention is not to demolish the house of software development, represented here by the items on the right (e.g. working software over comprehensive documentation), but claims that those who embrace the items
on the left rather than those on the right are more likely to succeed in the long run (Cockburn and Highsmith 2002). Even within the agile community there is some disagreement about the choices, but it is accepted that discussions can lead to constructive criticism. Our analysis showed that all these elements could be identified in the test and its results.

In our research we have always been conscious of a division of roles within the company, often expressed as “shop floor” and “management”, and working with a participatory design perspective we have worked very much from the shop floor point of view. During the study, this viewpoint of separate groups emerged and crystallised, and two disparate groups became apparent. We called these groups Designers, represented by e.g. interaction designers and system and interaction architects, representing the shop floor perspective, and Product Owners, including management, product planning, and marketing, representing the management perspective.

When regarding this in light of the Agile manifesto, we began to see how different groups may have an interest in different factors of the framework and the results that it can produce, and it became a point of interest to see how these factors related to the manifesto and which of the groups, Designers (D) or Product Owners (PO), is mainly interested in each particular item in the manifesto. The case study data was analysed on the basis of these emerging thoughts. Where the groups were found to fit on the scale is marked in bold text in the paragraphs that follow. One of the items is changed from “Working software” to “Working information” as we see the information resulting from the testing process as a metaphor for the software that is produced in software development.

- **Individuals and interactions** – The testing process is dependent on the individuals who lead the test, and who actually perform the testing on the devices. The central figure here is the test leader, who functions as a pivot point in the whole process, interacting with the testers, observing and registering the data, and presenting the results. This interaction is clearly important in the long run from a PO perspective, but it is D who has the greatest and immediate benefit of the interaction, showing how users reacted to design decisions, that is a central part of the testing.

- **Processes and Tools** – The test is based upon a well-defined process that can be repeated to collect similar data that can be compared over a period of time. This is important for the designers, but in the short term they are more concerned with the everyday activities of design and development that they are involved in. Therefore we see this as being of greatest interest to PO, who can get a long-term view of the product, its development, and e.g. comparisons with competitors, based on a stable and standardised method.

- **Working information** – The test produces working information quickly. Directly after the short period of testing that is the subject of this case study, before the data was collated in the spreadsheets, the test leaders met and discussed and agreed upon their findings. They could present the most important qualitative findings to system and interaction architects within the two organisations 14 days after the testing began, and changes in the implementation were requested soon after that. An
advantage of doing the testing in-house is having access to the test leaders, who can explain and clarify what has happened and the implications of it. This is obviously of primary interest to D.

- **Comprehensive documentation** – The documentation consists mainly of spreadsheets containing metrics and qualitative data. Metrics back up qualitative data and open up ways to present test results that can be understood without having to include contextual information. They make test results accessible for new groups. The quantitative data gives statistical confirmation of the early qualitative findings, but are regarded as most useful for PO, who want figures of the findings that have been reached. There is less pressure of time to get these results compiled, as the critical findings are already being implemented. The metrics can be subject to stringent analysis to show comparisons and correlations between different factors. In both organisations there is beginning to be a demand for Key Performance Indicators for usability, and although it is still unsure what these may consist of, it is still an indication of a trend that comes from PO level.

- **Customer collaboration** – in the testing procedure it is important for the testers to have easy access to individuals, to gain information about customer needs, end user patterns, etc. The whole idea of the test is to collect the information that is needed at the current time regarding the product and its development. How this is done in practice is obviously of concern to PO in the long run, but in the immediate day to day operation it is primarily of interest to D.

- **Contract negotiation** – On a high level it is up to PO to decide what sort of cooperation should take place between different organisations and customers, and this is not something that involves D, so this is seen as most important for PO.

- **Respond to change** – The test is easily adapted to changes, and is not particularly resource-intensive. If there is a need to change the format of a test, or a new test requirement turns up suddenly, it is easy to change the test without having expended extensive resources on the testing. It is also easy to do a “Light” version of a test to check a particular feature that arises in the everyday work of design, and this has happened several times at UIQ. This is the sort of thing that is a characteristic of the day to day work with interaction design, and is nothing that is of immediate concern for PO, so this is seen as D.

- **Following a plan** – From a short-term perspective, this is important for D, but since they work in a rapidly changing situation, it is more important for them to be able to respond to change. This is however important for PO who are responsible for well-functioning strategies and long-term operations in the company.

4.5.1 On Opposite Sides of the Spectrum

In this analysis, we found that “Designers”, as in the agile manifesto, are interested in the items on the left, rather than the items on the right (see Figure 4.2). We see this as being “A Designer’s Manifesto”. “Product Owners” are more interested in the items on the right. Boehm characterised the items on the right side as being “An Auditor Manifesto” (Boehm 2007). We see it as being “A Product Owner’s Manifesto”. This is of course a sliding scale;
some of the groups may be closer to the middle of the scale. Neither of the two groups is uninterested in what is happening at the opposite end of the spectrum, but as in the agile manifesto, while there is value in the items on one side, they value the items on the other side more. We are conscious of the fact that these two groups are very coarsely drawn, and that some groups and roles will lie between these extremes. We are unsure exactly which roles in the development process belong to which group, but are interested in looking at these extremes to see their information requirements in regard to the results of usability testing. On closer inspection it may be found that none of the groups is on the far side of the spectrum for all of the points in the manifesto. To gain further information regarding this, a case study has been performed, which we present in the next section.

![Figure 4.2 Groups and their diverging interests](image)

### 4.6 Follow-up Study of Preferred Presentation Methods

This study is thus an investigation of attitudes regarding which types of usability findings different stakeholders need to see, and their preferred presentation methods. In the previous study we identified two groups of stakeholders with different information needs, ranging from Designers, who appear to want quick results, often qualitative results rather than quantitative results, to Product Owners, who want more detailed information, are more concerned with quantitative results, but are not as concerned with the speediness of the results. (See figure 4.2) To test this theory, we sent a questionnaire to a number of stakeholders within UIQ and their customers, who are participants in the design and development process.

A document was compiled illustrating ten methods for presenting the results of UTUM tests. It contained a brief description of the presentation method and the information contained in it. The methods were chosen together with a usability expert from UIQ who often presents the results of testing to different groups of stakeholders. The methods were
chosen on the basis of his experience of presenting test results to different stakeholders and are the most used and most representative ways of presenting results. The methods range from a verbal presentation of early findings, to spreadsheets containing all of the quantitative or the qualitative data from the testing, plus a number of graphical representations of the data. The methods were as follows

**Method 1: The Structured Data Summary (the SDS).** A spreadsheet with the qualitative findings of the testing. It shows issues that have been found, on the basis of each tester and each device, for every use case. Comments made by the test participants and observations made by the test leader are stored in the spreadsheet.

**Method 2: A spreadsheet containing all “raw” data.** All of the quantitative data from a series of tests. Worksheets contain the numerical data collected in a specific series of tests, which are also illustrated in a number of graphs. The data includes times taken to complete use cases, and the results of attitude assessments.

**Method 3: A Curve diagram.** A graph illustrating a comparison of time taken to complete one particular use case. One curve illustrates the average time for all tested telephones, and the other curves show the time taken for individual phones.

**Method 4: Comparison of two factors (basic version).** An image showing the results of a series of tests, where three telephones are rated and compared with regard to satisfaction and efficiency. No more information is given in this diagram.

**Method 5: Comparison of two factors (brief details).** The same image as Method 4, with a very brief explanation of the findings.

**Method 6: Comparison of two factors (more in depth details).** The same image as Methods 4 and 5. Here, there is a more extensive explanation of the results, and the findings made by the test leader. The test leader has also written suggestions for short term and long term solutions to issues that have been found.

**Method 7: The “Form Factor” – an immediate response.** A visual comparison of which telephone was preferred by men and women, where the participants were asked to give an immediate response to the phones, and choose a favourite phone on the basis of “Form Factor” – the “pleasingness” of the design.

**Method 8: PowerPoint presentation, no verbal presentation.** A PowerPoint presentation, produced by the test leader. A summary of the main results is presented graphically and briefly in writing. This does not give the opportunity to ask follow-up questions in direct connection with the presentation.

**Method 9: Verbal presentation supported by PowerPoint.** A PowerPoint presentation, given by the test leader. A summary of the main results is presented graphically and briefly in writing, and explained verbally, giving the listener the chance to ask questions about e.g. the findings and suggestions for improvements. This type of presentation takes the longest to prepare and deliver.

**Method 10: Verbal presentation of early results.** The test leader gives a verbal presentation of the results of a series of tests. These are based mainly on his or her
impressions of issues found, rather than an analysis of the metrics, and can be given after having observed a relatively small number of tests. This is the fastest and most informal type of presentation, and can be given early in the testing process.

The participants in the study were chosen together with the usability expert at UIQ. Some of the participants were people who are regularly given presentations of test results, whilst others were people who are not usually recipients of the results, but who in their professional roles could be assumed to have an interest in the results of usability testing. They were asked to read the document and complete the task by filling in their preferences in a spreadsheet. The results of the survey were returned to the researcher via e-mail and have been summarised in a spreadsheet and then analysed on the basis of a number of criteria to see what general conclusions can be drawn from the answers.

The method whereby the participants were asked to prioritize the presentation methods was based on cumulative voting (Investopedia.com 2008), (Wikipedia 2008), a well-known voting system in the political and the corporate sphere (Gordon 1994), (Sawyer and MacRae 1962), also known as the $100 test or $100 method (Leffingwell and Widrig 2003). Cumulative voting is a method that has previously been used in the software engineering context, for e.g. software requirement prioritization (Regnell, Höst et al. 2001) and the prioritization of process improvements (Berander and Wohlin 2003), and in (Berander and Jönsson 2006) where it is compared to and found to be superior to Analytical Hierarchy Process in several respects.

The questionnaire was sent to 29 people, mostly within UIQ but also to some people from UIQ’s licensees. Only six respondents had replied to the questionnaire within the stipulated time, so one day after the first deadline, we sent out a reminder to the respondents who had not answered. This resulted in a further three replies. After one more week, we sent out a final reminder, leading to one more reply. Thus, we received 10 replies to the questionnaire, of which nine were from respondents within UIQ. On further enquiry, the reason given for not replying to the questionnaire was in general the fact that the company was in an intensive working phase for a planned product release, and that the staff at the company could not prioritise allocating the time needed to complete the questionnaire. This makes it impossible to give full answers to the research questions in this study, although it helps us to answer some of the questions, and gives us a better understanding of factors that affect the answers to the other questions. This study helps us formulate hypotheses for further work regarding these questions.

The division of roles amongst the respondents, and the number of respondents in the categories was as follows:

- 2: UI designers
- 2: Product planning
- 4: System design
- 1: Other (Usability)
- 1: Other (CTO Office)
We have divided the respondents according to the tentative schema found in the first case study, between Designers (D) and Product Owners (PO). Some respondents were difficult to place in a particular category. The roles the respondents held in the company were discussed with a member of the management staff at UIQ, with long work experience at the company, who was well versed in the thoughts we had regarding the difference between Designers and Product Owners. Due to turbulence within the company, it was not possible to verify the respondents’ attitudes to their positions, and would have been difficult, since they were not familiar with the terminology that we used, and the meaning of the roles that we had specified.

Five respondents, the two UI designers, the usability specialist and two of the system designers, belonged to the Designer group. The remaining five respondents, the two members of product planning, the respondent from the CTO office and two of the system designers, were representatives of the group of Product Owners.

![Distribution of allocated points](image)

**Figure 4.3 Distribution of points allocated per respondent**

Figure 4.3 is a box and whisker plot that shows the distribution of the points and the mean points allocated per person. As can be seen, the spread of points differs greatly from person to person. Although this reflects the actual needs of the respondent, the way of allocating
points could also reflect tactical choices, or even the respondent’s character. To get more information about how the choices were made would require a further study, where the respondents were interviewed concerning their strategies and choices.

In what follows, we use various ways of summarising the data. To obtain a composite picture of the respondents’ attitudes, the methods are ranked according to a number of criteria. Given the small numbers of respondents in the study, this compilation of results is used to give a more complex picture of the results, rather than simply relying on one aspect of the questionnaire. The methods are ranked according to: the total number of points that were allocated by all respondents; the number of times the method has been chosen, and; the average ranking, which is the sum of the rankings given by each respondent, divided by the number of respondents that chose the method (e.g., if one respondent chose a method in first place, whilst another respondent chose it in third place, the average position is \((1+3)/2 = 2\)). A lower average ranking means a better result in the evaluation, although it gives no information about the number of times it has been chosen.

**Figure 4.4 Comparison: All, Designers & Product Owners (lowest value is best)**

Figure 4.4 shows a summary of the results for all respondents and a comparison with the results for the group of Designers and Product Owners. For all respondents, total rank is very similar to ranking according to points allocated, and only two methods (ranked 5 and
6) have swapped places. Three methods head the list. Two are verbal presentations, one being supported by PowerPoint and the other is purely verbal.

Even within the two groups of Designers and Product Owners, there is little discrepancy between the results for total rank and position according to points awarded. Table 4.1 illustrates the ranks. The Methods are ordered according to the points allocated by all respondents. The next columns show the composite results, for all respondents and according to the two groups. Cases where the opinions differ significantly between Designers and Product Owners (a difference of 3 places or more) will be the subject of a brief discussion, to see whether we can draw any tentative conclusions about the presentation requirements of the different stakeholder groups. These methods, which are shown in bold and underlined in table 4.1, are Methods 1, 3, 4 and 8. Since the company has now ceased operations, it is no longer possible to do a follow-up study of the attitudes of the participants, so the analysis is based on the knowledge we have of the operations at the company and the context where they worked. To verify these results, further studies are needed.

**Table 4.1 Comparison of ranks: All, Designers and Product Owners**

<table>
<thead>
<tr>
<th>Method</th>
<th>All respondents</th>
<th>Designers</th>
<th>Product Owners</th>
<th>Difference between groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. Verbal &amp; PowerPoint</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>6. Comparison (more details)</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>10. Verbal</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td><strong>8. PowerPoint</strong></td>
<td><strong>4</strong></td>
<td><strong>7</strong></td>
<td><strong>1</strong></td>
<td><strong>6</strong></td>
</tr>
<tr>
<td>5. Comparison (brief details)</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td><strong>3. Curve diagram</strong></td>
<td><strong>5</strong></td>
<td><strong>2</strong></td>
<td><strong>9</strong></td>
<td><strong>7</strong></td>
</tr>
<tr>
<td><strong>1. The SDS</strong></td>
<td><strong>7</strong></td>
<td>5</td>
<td><strong>10</strong></td>
<td><strong>5</strong></td>
</tr>
<tr>
<td><strong>4. Comparison (no details)</strong></td>
<td><strong>8</strong></td>
<td><strong>8</strong></td>
<td><strong>5</strong></td>
<td><strong>3</strong></td>
</tr>
<tr>
<td>7. “Form factor”</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>2. Spreadsheet</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

**Method 3: The Curve Diagram.** Designers ranked this presentation highly because if it is interpreted properly, it can give a great deal of information about the use case as it is performed on the device. If the device performs poorly in comparison to the other devices, which can easily be seen by the placement and shape of the curve, this indicates that there are problems that need to be investigated further. Use case performance time indicates the performance of the device, which can be correlated with user satisfaction. The shape of the curve illustrates when problems arose. If problems arise when performing the use case, these will be visible in the diagram and the Designers will know that there are issues that must be attended to.

Product Owners ranked this method poorly because the information is on the level of an individual use case, whilst they need information about the product or device at a coarser level of detail that is easy to interpret, giving an overall view of the product. They trust that
problems at this level of detail are dealt with by the Designers, whilst they have responsibility for the product and process as a whole.

**Method 8: PowerPoint presentation, no verbal presentation.** This can contain several ways of presenting the results of testing. Designers find this type of presentation of limited use because of the lack of contextual information and the lack of opportunity to pose follow-up questions. It gives a lot of information, but does not contain sufficient details to allow Designers to identify the problems or make decisions about solutions. Without details of the context and what happened in the testing situation, it is hard to interpret differences between devices, to know which problems there are in the device, and thereby difficult to know what to do about the problems. The length of time taken to produce the presentation also means that it is not suitable for Designers, who are concerned with fixing product issues as early in the development process as possible. We also believe that there is also a difference in “culture” where Designers are still unused to being presented with results in this fashion, and cannot translate this easily to fit in with their work practices.

This type of presentation is of primary interest to Product Owners because it provides an overall view of the product in comparison to other devices, without including too much information about the context and test situation. It contains sufficient text, and gives an indication of the status of the product. It is also adapted to viewing without the presence of the test leader, so the recipient can view the presentation and return to it at will. Product Owners are often schooled in an engineering tradition and are used to this way of presenting information.

**Method 1: The Structured Data Summary (the SDS).** Designers value this method of presentation because of the extent and character of the contextual information it includes, and because of the way the data is visualised. For every device and use case, there is information on issues that were observed, and records of comments made by the testers. It is easy to see which use cases were problematic, due to the number of comments written by the test leader, and the presence of many user comments also suggests that there are issues that need investigation. The contextual information gives clues to problems and issues that must be dealt with and gives hints on possible solutions. The effort required to read and summarise the information contained in the spreadsheet, leading to a degree of cognitive friction, means however that it is rated in the middle of the field rather than higher. Product Owners rate this method poorly because they are uninterested in products on the level of use cases, which this presentation gives provides, and it is difficult to interpret for the device as a whole. The information is not adapted to the broad view of the product that the Product Owners need. The contextual information is difficult to summarise and does not give a readily understandable of the device as a whole. Product Owners find it difficult to make use of the information contained in this spreadsheet and thereby rank it as least useful for their needs.

**Method 4: Comparison of two factors (basic version).** The lack of detail and of contextual information makes it difficult for Designers to read any information that allows them to identify problems with the device. It simply provides them with a snapshot of how their product compares to other devices at a given moment. Product Owners ranked this in
the middle of the field. This is a simple way of visualising the state of the product at a
given time, which is easy to compare over a period of time, to see whether a device is
competitive with the other devices included in the comparison. This is typically one of the
elements that are included in the PowerPoint presentation that Product Owners have ranked
highest (Method 8). However, this particular method, when taken in isolation, lacks the
richness of the overall picture given in Method 8 and is therefore ranked as lower.

To summarise these results, we find that the greatest difference between the two groups
concerns the level of detail included in the presentation, the ease with which the
information can be interpreted, and the presence of contextual information in the
presentation. Designers prioritise methods that give specific information about the device
and its features. Product Owners prioritise methods that give more overarching information
about the product as a whole, and that is not dependent on including contextual
information.

4.6.1 Changing Information Needs

Participants were informed that the survey was mainly focused on the presentation of
results that are relevant during ongoing design and development. We pointed out that we
believed that different presentation methods may be important in the starting and finishing
phases of these processes. We stated that comments regarding this would be appreciated.
Three respondents wrote comments about this factor.

One respondent (D) stated that the information needed in their everyday work as a UI
designer, in the early stages of projects when the interaction designers are most active, was
best satisfied through the verbal presentations of early results and verbal presentation
supported by PowerPoint, whilst a non-verbal presentation, in conjunction with the metrics
data in the spreadsheet and the SDS would be more appropriate later in the project, where
the project activities were no longer as dependent on the work tasks and activities of the
interaction designers.

A second respondent (D) stated that the verbal presentations are most appropriate in the
requirements/design processes. Once the problem domain is understood, and the task is to
iterate towards the best solution, the metrics data and the SDS would become more
appropriate, because the problem is understood and the qualitative answers are more easily
interpreted than the qualitative answers.

Another respondent (PO) wrote that it was important to move the focus from methods that
were primarily concerned with verification towards methods that could be of assistance in
requirements handling, in prioritisation and decision making in the early phases of
development. In other words, the methods presented are most appropriate for later stages of
a project, and there is a lack of appropriate methods for early stages.

Given the limited number of answers to these questions, it is of course difficult to draw any
general conclusions, although it does appear to be the case that the verbal results are most
important in the early stages of a project, to those who are involved in the actual work of
designing and developing the product, whilst the more quantitative data is more useful as
reference material in the later stages of a project, or further projects.
4.6.2 Attitudes Towards the Role of the Test Leader

The respondents were asked to judge whether or not they would need the help of the test leader in order to understand the presentation method in question. Two of the respondents supplied no answers to this question, and one of the respondents only supplied answers regarding methods 9 and 10, which presuppose the presence of the test leader and are therefore excluded from the analysis. If we exclude these three respondents from the summary, there were seven respondents, of whom four gave answers for all eight methods, one gave five answers, and two gave three answers. The three respondents who did not answer these questions were all Product Owners, meaning that there were five designers and two Product Owners who answered these questions.

Analysis of the answers showed that, with the exception of Method 7 the methods that are primarily graphical representations of the data do not appear to require the presence of the test leader to explain the presentation. Method 7 was found to require the presence of the test leader, presumably because it was not directly concerned with the operations of the company. The spreadsheets however, one containing qualitative and one containing quantitative data, both require the presence of a test leader to explain the contents.

Given the fact that the Designers were in the majority, there were few obvious differences between Designers and Product Owners, although the most consistent findings here regard methods 4, 5, and 6, variations of the same presentation method with different amounts of written information. Here, Product Owners needed the test leader to be present whilst Designers did not.

4.6.3 In Summary

We now summarise the results of the research questions posed in this case study. The answer to the first question, whether any presentation methods are generally preferred, is that the respondents as a whole generally preferred verbal presentations. The primarily verbal methods are found in both first and third place. The most popular form was a PowerPoint presentation that was supported by verbal explanations of the findings. In second place is a non-verbal illustration showing a comparison of two factors, where detailed information is given explaining the diagram and the results it contains. This type of presentation is found in several variants in the study, and those with more explanatory detail are more popular than those with fewer details. Following these is a block of graphical presentation methods that are not designed to be dependent on verbal explanations. Amongst these is a spreadsheet containing qualitative data about the test results. At the bottom of the list is a spreadsheet that contains the quantitative data from the study. This presentation differs in character from the SDS, the spreadsheet containing qualitative data, since the SDS offers a view of the data that allows the identification of problem areas for the tested devices. This illustrates the fact that even a spreadsheet, if it offers a graphical illustration of the data that it contains, can also be found useful for stakeholders, even without an explicit explanation of the data that it contains.

Concerning the second question, we could identify differences between the two groups of stakeholders, and the greatest difference between the groups concerns the level of detail included in the presentation, the ease with which the information can be interpreted, and the
presence of contextual information in the presentation. Designers prioritise methods that give specific information about the device and its features. Product Owners prioritise methods that give more overarching information about the product as a whole, and that is not dependent on including contextual information. We also found that both groups chose PowerPoint presentations as their preferred method, but that the Designers chose a presentation that was primarily verbal, whilst Product Owners preferred the purely visual presentation. Another aspect of this second question is the attitude towards the role of the test leader, where there were few obvious differences between Designers and Product Owners. The most consistent findings here concern variations of the same presentation method with different amounts of written information. Here, Product Owners needed the test leader to be present whilst Designers did not.

Regarding the third question, if there are methods that are lacking in the current presentation methods, it was found that taking into account and visualising aspects of UX is becoming more important, and the results indicate that testing must be adapted to capture these aspects more implicitly. There is also a need for a composite presentation method combining the positive features of all of the current methods – however, given the fact that there do appear to be differences between information needs, it may be found to be difficult to devise one method that satisfies all groups.

No clear answers can be found for the fourth question, whether information needs, and preferred methods change during different phases of a design and development project. However, the replies suggest that the required methods do change during a project, that more verbally oriented and qualitative presentations are important in early stages of a project, in the concrete practice of design and development, and that quantitative orientated methods are important in later stages and as reference material.

Regarding the final question, whether results can be presented without the presence of the test leader, we find that the methods that are primarily graphical representations of the data do not appear to require the presence of the test leader to explain the presentation. The spreadsheets however, containing qualitative and quantitative data, both require the presence of a test leader to explain the contents.

To verify these results, further studies are of course needed. Despite the small scale of this study, the results give a basis for performing a further study, and allow us to formulate a hypothesis for following up our results. In line with the rest of the work performed as part of this research, we feel that this work should be a survey based study in combination with an interview based study, in order to verify the results from the survey and gain a depth of information that is difficult to obtain from a purely survey based study.

We continue by discussing the results of the two case studies in relation to the industrial situation where we have been working, and the need for quality assurance in development and design processes.

4.7 Discussion

We begin by discussing our results in relation to academic discourses, to answer our first research question: How can we balance demands for agile results with demands for formal
results when performing usability testing for quality assurance? We also comment upon two related discourses from the introductory chapter, i.e. the relation between quality and a need for cooperation between industry and research, and the relationship between quality and agility.

Since we work in a mass-market situation, and the system that we are looking at is too large and complex for a single customer to specify, the testing process must be flexible enough to accommodate the needs of many different stakeholders. The product must appeal to the broadest possible group, so it is difficult for customers to operate in dedicated mode with a development team, with sufficient knowledge to span the whole range of the application, which is what an agile approach requires to work best (Boehm 2002). In this case, test leaders work as proxies for the user in the mass market. We had a dedicated specialist test leader who brought in the knowledge that users have, in accordance with Pettichord (2000). Evidence suggests that drawing and learning from experience may be as important as taking a rational approach to testing (Martin, Rooksby et al. 2007). The fact that the test leaders involved in the testing are usability experts working in the field in their everyday work activities means that they have considerable experience of their products and their field. They have specialist knowledge, gained over a period of time through interaction with end-users, customers, developers, and other parties that have an interest in the testing process and results. This is in line with the idea that agile methods get much of their agility from a reliance on tacit knowledge embodied in a team, rather than from knowledge written down in plans (Boehm 2002).

It would be difficult to gain acceptance of the test results within the whole organisation without the element of formalism. In sectors with large customer bases, companies require both rapid value and high assurance. This cannot be met by pure agility or plan-driven discipline; only a mix of these is sufficient, and organisations must evolve towards the mix that suits them best (Boehm 2002). In our case this evolution has taken place during the whole period of the research cooperation, and has reached a phase where it has become apparent that this mix is desirable and even necessary.

In relation to the above, Osterweil (1996) states that there is a body of knowledge that could do much to improve quality, but that there is “a yawning chasm separating practice from research that blocks needed improvements in both communities”, thereby hindering quality. Practice is not as effective as it must be, and research suffers from a lack of validation of good ideas and redirection that result from serious use in the real world. This case study is part of a successful cooperation between research and industry, where the results enrich the work of both parts. Osterweil (1996) also requests the identification of dimensions of quality and measures appropriate for it. The particular understanding of agility discussed in our case study can be an answer to this request. The agility of the test process is in accordance with the “good organisational reasons” for “bad testing” that are argued by Martin, Rooksby et al (2007). These authors state that testing research has concentrated mainly on improving the formal aspects of testing, such as measuring test coverage and designing tools to support testing. However, despite advances in formal and automated fault discovery and their adoption in industry, the principal approach for
validation and verification appears to be demonstrating that the software is “good enough”.
Hence, improving formal aspects does not necessarily help to design the testing that most
efficiently satisfies organisational needs and minimises the effort needed to perform testing.
In the results of this work, the main reason for not adopting “best practice” is precisely to
orient testing to meet organisational needs. Our case is a confirmation of (Martin, Rooksby
et al. 2007). Here, it is based on the dynamics of customer relationships, using limited
effort in the most effective way, and the timing of software releases to the needs of
customers as to which features to release. The present paper illustrates how this happens in
industry, since the agile type of testing studied here is not according to “best practice” but
is a complement that meets organisational needs for a mass-market product in a rapidly
changing marketplace, with many different customers and end-users.

To summarise our second case study, the findings presented here are the results of a
preliminary study that indicates the needs of different actors in the telecom industry. They
are a validation of the ways in which UTUM results have been presented. They provide
guidelines to improving the ways in which the results can be presented in the future. They
are also a confirmation of the fact that there are different groups of stakeholders, the
Designers and Product Owners found in our first case study, who have different
information requirements. Further studies are obviously needed, but despite the small scale
of this study, it is a basis for performing a wider and deeper study, and it lets us formulate a
hypothesis regarding the presentation of testing results. We feel that the continuation of this
work should be a survey based study in combination with an interview based study.

4.8 Conclusions and Further Work
In the usability evaluation framework, we have managed to implement a working balance
between agility and plan driven formalism to satisfy practitioners in many roles. The
industrial reality that has driven the development of this test package confirms the fact that
quality and agility are vital for a company that is working in a rapidly changing
environment, attempting to develop a product for a mass market. There is also an obvious
need for formal data that can support the quick and agile results. The UTUM test package
demonstrates one way to balance demands for agile results with demands for formal results
when performing usability testing for quality assurance. The test package conforms to both
the Designer's manifesto, and the Product Owner’s manifesto, and ensures that there is a
mix of agility and formalism in the process.

The case in the present paper confirms the argumentation emphasizing 'good
organizational reasons’, since this type of testing is not according to “best practice” but is a
complement that meets organisational needs for a mass-market product in a rapidly
changing marketplace, with many different customers and end-users. This is partly an
illustration of the chasm between industry and research, and partly an illustration of how
agile approaches are taken to adjust to industrial reality. In relation to the former this case
study is a successful cooperation between research and industry. It has been ongoing since
2001, and the work has an impact in industry, and results enrich the work of both parts. The
inclusion of Sony Ericsson in this case study gave even greater possibilities to spread the
benefits of the cooperative research. More and more hybrid methods are emerging, where agile and plan driven methods are combined, and success stories are beginning to emerge. We see the results of this case study and the UTUM test as being one of these success stories. How do we know that the test is successful? By seeing that it is in successful use in everyday practice in an industrial environment. We have found a successful balance between agility and formalism that works in industry and that exhibits qualities that can be of interest to both the agile and the software engineering community.

Acknowledgements
This work was partly funded by The Knowledge Foundation in Sweden under a research grant for the software development project “Blekinge – Engineering Software Qualities”, www.bth.se/besq. Thanks to the participants in the study and to my colleagues in the U-ODD research group for their help in data analysis and structuring my writing. Thanks also to Gary Denman for permission to use the Structured Data Summary.
Chapter Five

Article Four
Chapter Five

SPI Success factors within product usability evaluation


Jeff Winter, Kari Rönkkö

Abstract

This article presents an experience report where we compare 8 years of experience of product related usability testing and evaluation with principles for software process improvement (SPI). In theory the product and the process views are often seen to be complementary, but studies of industry have demonstrated the opposite. Therefore, more empirical studies are needed to understand and improve the present situation. We find areas of close agreement as well as areas where our work illuminates new characteristics. It has been identified that successful SPI is dependent upon being successfully combined with a business orientation. Usability and business orientation also have strong connections although this has not been extensively addressed in SPI publications. Reasons for this could be that usability focuses on product metrics whilst today’s SPI mainly focuses on process metrics. Also because today’s SPI is dominated by striving towards a standardized, controllable, and predictable software engineering process; whilst successful usability efforts in organisations are more about creating a creative organisational culture advocating a useful product throughout the development and product life cycle. We provide a study and discussion that supports future development when combining usability and product focus with SPI, in particular if these efforts are related to usability process improvement efforts.

5.1 Introduction

This article presents an experience report that is based on 8 years of our own experiences of implementing usability product metrics as a quality assurance approach. The material that it is grounded in has been collected in close cooperation with an industrial actor, in an action research based approach, using ethnographic methods and a case study approach.

Software engineers often believe that software development is negatively affected by measurements with no quantitative description of where we are and where we would like to go (Pfleeger and Atlee 2006). Measurement is seen as being indispensable to introducing engineering discipline to software development, maintenance and use (Basili, Caldiera et al. 1994). Software testing permits measurements that give organisations better insight into the processes being monitored. Measurement can help us to understand what is happening, to control what is happening, and to improve processes, and products (Fenton and Pfleeger 1998). Usability testing is one part of the testing area, although it is not a commonly
addressed area within software engineering. This may be because usability testing is an interdisciplinary subject, often performed by people with a social sciences background that emphasises qualitative aspects. These people are thus professionals from other paradigms and backgrounds, trained for the usability product test situation, rather than for software development process aspects. To be a professional in usability requires extensive knowledge of interaction theory, user-research methodologies, and especially the principles of user testing. It can often take an interaction designer several years to build up this knowledge. Specialization might also be important here, as an individual’s orientation towards design or usability seems to require different personality types. This is visible in usability and interaction design education targeting industrial actors (Nielsen 2002).

Another explanation might be that usability focuses on product metrics whilst software engineering mainly focuses on process metrics. It is recognised in software engineering that usability does not only affect the design of user interfaces, but the development as a whole, especially related to the requirement process. It is also recognised that both areas require own professional knowledge and approaches. Still, the interplay between these two fields and activities they advocate to be undertaken in software development has proved to be difficult (Hellman and Rönkkö 2008a).

Today’s prevailing focus on process metrics has not always been the major trend. It has also been questioned how successful the process focused approach actually is. Based on an extensive review of quality management, organisational learning and software process improvement (SPI) literature regarding key factors for success, Dybå (2000a) identified three trends: The first trend is that current developments in quality management are heavily influenced by recent attention towards the improvement of production processes. Hence, SPI is viewed mainly from one point of view. In continuation Dybå claims that such a one-sided view causes confusion and is a misunderstanding “. . .partly related to quasi-religious and sectarian controversies and partly related to the most important features of modern quality management: it draws attention to the improvement of production processes and not simply to the characteristic of the product.” Whereby, the original quality control principle, assuring that the characteristics of the final product fall within decided tolerance, is opposed. The second trend is that a recent revision of the ISO 9000 family of standards also follows the same tendency, by distancing itself from the term “quality assurance” and instead emphasising the adoption of a process approach to “quality management”. The third trend is that no single model has yet established itself as a basis for total quality management theory.

In relation to the above, it is worth noting that aligning SPI goals and action with explicit and implicit business goals and strategies is one of the factors that has the strongest influence on SPI success (Dybå 2005). Success is dependent upon successfully combining SPI and business orientation. The question is how this can be achieved. Related to our usability focus it is well known that usability and business orientation have strong connections; good usability is part of the public’s perception of a company, affecting the value of a brand and the share of the market. Products are often publicly “talked about” in magazines, journals and national newspapers in terms of: ease of learning, ease of use,
quality of documentation, i.e. common usability aspects (Bias and Mayhew 2005, p. 18). It is also well known that usability efforts lead to increased user productivity, decreased user errors, decreased training costs, early changes in the development lifecycle, and decreased user support; which lead to business benefits in the form of increased sales, decreased customer support, reduced costs of providing training support, increased savings from making changes early in design lifecycle, and finally increased perception of the value of company by stakeholders (Bias and Mayhew 2005). Here usability product metrics provide one potential area for increased and improved business, which can be related to SPI. Unfortunately the one-sided focus on process that currently prevails in SPI, together with a lack of publications discussing the combination of product and process, is a hinder for positive development in this area. Even though our study does not explicitly address SPI in agreement with the product line approach that currently dominates (the rationalistic paradigm), with standardized, controllable, and predictable software engineering processes, we still believe that our study provides knowledge here. On the other hand there are other views that relate closer to our case, i.e. creating an organisational culture.

In software engineering in general it is believed that the correct path follows imitating technical procedures identified by others. Based on a quantitative study of 120 software organisations Dybå suggests an approach that differs from most of existing SPI literature, i.e. to focus the SPI efforts instead on creating an organisational culture within which technical procedures can thrive (Dybå 2005). An earlier study by Iversen and Kautz (2001) supports Dybå’s idea of “thriving” by means of advocating the need to tailor methods to organisational needs. Iversen and Kautz have studied and summarised recommendations and advice regarding SPI metrics implementations, and suggest that whilst guidelines and frameworks are useful, organisations must tailor methods to fit their particular situation. This is particularly true since much of the advice given disregards the fact that many companies do not employ large numbers of software developers, and that organisational cultures differ between different countries. Advice aimed at, e.g., North American companies may not fit in with organisational cultures in other countries (Iversen and Kautz 2001, pp. 288–289). Iversen and Kautz’s study is of particular relevance to us, since our metrics efforts took place in Sweden, in a Scandinavian context (see also Dybå’s critique of applying SPI results from North America in European contexts in general, and particularly in Scandinavian contexts (Dybå 2005)).

Another topic of relevance for our study is exploitation and exploration. Software organisations can engage in exploitation, the adoption and use of existing knowledge and experience; or exploration, the search for new knowledge either through imitation or innovation. According to Dybå (2000b), which of these two becomes dominant depends on the size of the company. Smaller companies often keep the same level of exploitation, independently of whether they operate in a stable or turbulent environment, whilst at the same time they increase their exploration in turbulent environments. Larger companies (more than 200 developers) had similar results for exploitation, but in contrast to smaller companies, they do not increase their levels of exploration when the level of turbulence increases. Dybå explains that management groups within large organisations still rely on their previous experience to prepare for the future rather than exploring new possibilities.
They generate the same response even though stimuli change, and in a sense keep doing what they do well, rather than risk failure (Dybå 2000b). One suggested explanation for this is that the dominant perspective in today’s software development is the rationalistic paradigm emphasising standardized, controllable, and predictable software engineering process. A large part of the SPI community has also promoted this rationalistic view, e.g. through explicit support of the use of rigorous statistical techniques. Without suggesting the abandonment of the discipline altogether, it is still claimed that software development practice does not conform well to such a rationalistic view (Dybå 2000b).

Of course, both creativity and discipline are needed. Dybå observed that small successful organisations tend to put more emphasis on exploring new possibilities, embracing diversity and creativity, whilst larger successful organisations tend to emphasise “best practices” and formal routines (Dybå 2003). This is a balance that is challenging. In our study we refer to these two sides by using the terms agile and formal (Winter and Rönkkö 2009). This does not imply that we see agile processes as being informal or unstructured, but because the word “formal” is more representative than e.g. “plan driven” to characterise the results of testing and how these results are presented to certain stakeholders. When we started the cooperation with the studied company it was a small company with approximately 50 employees. When we concluded our cooperation it was a large company with approximately 400 employees (more than 200 developers; Dybå’s definition in (2000b). In relation to the usability evaluation framework developed by us, we kept an explorative mindset throughout its implementation with help of a Cooperative Method Development (CMD) (Dittrich, Rönkkö et al. 2007) research approach. We also adjusted our test evaluation framework to satisfy management needs for standardized, controllable, and predictable software engineering process.

The present article will use our achieved knowledge and recommendations found within the SPI literature to discuss 8 years of our own experiences of implementing usability product metrics as a quality assurance approach, including growth in an industrial organisation and strategies to support requests from management and business. The experiences presented of introducing usability metrics in an industrial organisation are also well in line with Dybå’s identified need of re-introducing product quality assurance in SPI. We provide an example of what it might mean to create an organisational culture within which technical procedures can thrive. Our discussion is framed within the area of usability and organisational improvement, which as we mentioned earlier seems to be an unusual topic in software engineering, despite its close relation to business and the recognised importance of combining SPI and business orientation. In this article we discuss our experiences of influencing and improving a software development organisation by applying the language of usability metrics. We discuss how we tailored our metrics, based on work heavily influenced by a creative qualitative tradition, to fit an organisation influenced by a quantitative rationalistic paradigm and software engineering tradition.

Iversen and Kautz’s study dealt with SPI, and involved collecting data on the work performed by developers on an individual level, whilst our work has been the development and implementation of usability testing. Dybå’s study provided extended knowledge of
success criteria for SPI based on a quantitative survey where 120 software organisations helped verify previously known key factors for success in SPI. We have found many interesting parallels to their work and results and the work performed and findings made whilst developing and implementing our usability framework called UIQ Technology Usability Metrics (UTUM). UTUM is a usability test framework developed in an action research (Avison, Lau et al. 1999) fashion, based on the principles of CMD (Dittrich, Rönkkö et al. 2007). The research was a cooperation between, in particular, interaction designers from UIQ Technology AB and researchers (the present authors) from Blekinge Institute of Technology. This cooperation is presented in more detail later in this article. The comparison made, besides discussing usability product metrics experiences and SPI, also aims to help the reader assess the applicability of the lessons intended for companies attempting to implement software metrics programs. We believe this paper is of value for organisations in search for a starting point to tailor their usability metrics or introducing SPI. The experiences presented are from a Scandinavian based company in the telecommunications branch.

First, we introduce the telecommunications branch, followed by a short presentation of the research and the research partners. Thereafter we present the UTUM. After this we compare the areas and principles listed by Iversen and Kautz and Dybå with the conclusions we have drawn from our research and implementation of usability metrics. Finally, we discuss some issues we confronted and conclude our main experiences.

5.2 The Telecom area

Telecommunications, where we have cooperated with a company that developed and licensed a user-interface platform for mobile phones, is a market driven software development area. In market driven areas, there are potential groups of people who fit an imagined profile of intended users rather than a distinct set of users. Requirements elicitation is mainly managed through marketing, technical support, trade publication reviewers and user groups. Recent study in this area has revealed that the constant flow of requirements caused by the variety of stakeholders with different demands on the product is an issue. In market driven companies requirements are often invented based on strategic business objectives, domain knowledge and product visions (see Karlsson, Dahlstedt et al. 2007) for an overview).

It is thus a great challenge to develop a usability test framework for mass market products, where economic benefits are gained through approaching the broadest possible category with one single product. Whilst specific end-user groups must be targeted (Grudin and Pruitt 2002), there is an unwillingness to exclude other potential end-user categories (Rönkkö, Kilander et al. 2004). When developing software designed for single organisations, end-user participation, empowerment and the development of routines to be supported by technology are easier to identify, scope, and handle. In a mass market, it is harder to identify and portray the complexity and representativeness of end-users. Social and political aspects that influence the usefulness of products might even be filtered out by the evaluation techniques used in mass markets. A study demonstrating how socio-political circumstances made a usability and participatory design technique a too precise technique
to be applied is Rönkkö, Kilander et al. (2004). The design of products is influenced by competitors launching new products and features, and by technical magazines publishing reviews and comparisons. Timing aspects give competitive advantages and influence design decisions. Telecommunications focuses on providing the market with new and improved technology rather than fulfilling end-user needs. Taken together, these branch characteristics have so far challenged and bounded the design space for requirements elicitation and validation, and user testing. Since Apple’s launching of iPhone, claimed to be the first ‘real’ user experience product in the market, the above branch characteristics may change.

5.3 Research Cooperation

UIQ Technology AB, founded in 1999 and closed in January 2009, was an international company that early in 2008 had more than 320 employees in Sweden, and around 400 employees in total. The company developed and licensed a user-interface platform for mobile phones using Symbian OS. The product, UIQ’s user-interface platform, enabled mobile phone manufacturers to create different kinds of phones for different market segments, all based on one codeline. Through its independent position (not directly tied to a specific phone manufacturer) the company promoted the introduction of new advanced mobile phones onto the market. Its main assets were technical architecture, a unique independent product, and skilled, experienced staff. More than 20 UIQ-based phones were released.

The research group that has participated in the development of UTUM is U-ODD, Use-Oriented Design and Development (U-ODD 2011) which was also part of the research environment BESQ (2008). U-ODD approaches software engineering via use orientation, influenced by the application of a social science qualitative research methodology and the end-user’s perspective. The human role in software development is an area needing further research in software engineering. The task of understanding human behaviour is complex and necessitates the use of qualitative methods, since quantitative and statistical methods have been found to be insufficient (Seaman 1999). The strength of qualitative methodologies is in exploring and illuminating everyday practices of software engineers, through observing, interpreting and implementing the methods and processes of the practitioners.

Although UIQ as a whole worked as a traditional development organisation, influenced by a software engineering tradition, the particular work that we have studied at UIQ has been performed according to ISO 13407:1999 “Human-centred design processes for interactive systems” (ISO 13407 1999), an international standard concerning usability and human-centred design. ISO 13407:1999 “provides guidance on human-centred design activities throughout the life cycle of computer-based interactive systems. It is aimed at those managing design processes and provides guidance on sources of information and standards relevant to the human-centred approach” (ISO 13407 1999, p. 1). It defines human-centred design (HCD) as “an approach to interactive system development that focuses specifically on making systems usable. It is a multi-disciplinary activity which
incorporates human factors and ergonomics knowledge and techniques” (ISO 13407 1999, p. iv).

Bevan (2009) states that international standards for usability should be more widely used, and that if you are going to read only one standard, it should be ISO 9241-210 or ISO 13407. These provide a high level framework for usability work. The activities that are given in ISO 13407 are present in many design methods, and therefore the standard offers limited practical guidance; however, ISO 18529 contains a detailed list of activities and base practices, and including these activities in a design process ensures a continuous focus on the users of the system (Kesseler and Knapen 2006). It can be used in the specification, assessment and improvement of human-centred processes in system development and operation, and is intended to make the contents of ISO 13407 available to those familiar with or involved in process modelling (Bevan and Bogomolni 2000). Bevan (2009) also states that ISO 18529 or ISO 18152 should be used as a tool to improve usability capability within an organisation. This is supported by Earthy, Jones et al. (2001) who claim that the implication of ISO 13407 and ISO 18529 is that there is a professional responsibility for software engineers, system engineers and usability professionals to adopt good practice as defined in ISO 13407 and ISO 18529 as their baseline. This was not implemented in this company, so we cannot provide any details of it in this case, but we return to the reasoning behind it in our conclusions and discussion.

![Diagram of human-centred design process](image-url)

**Figure 5.1 Human-centred design (ISO 13407, 1999).**
HCD has four main principles and four main activities (see Figure 5.1 for an illustration of the activities), which are closely connected. In the following, we look at these principles and activities, and how they are connected to the work performed at UIQ. The inclusion of UTUM in an HCD-centred process was important, since this has allowed the results of usability testing to have an impact on design and development processes within the company.

The first principle in HCD is active involvement of users. This needs a clear understanding of user and task requirements. User involvement gives knowledge about the use context and how users are likely to work with a future system or product. Even when developing generic or consumer products, which by their nature have a dispersed user population, it is still important to involve users or “appropriate representatives” in development, in order to identify relevant requirements and to provide feedback through testing of proposed design solutions. The first activity in HCD is therefore understanding and specifying the use context. The context should be identified in terms of the characteristics of the intended users, the tasks the users are to perform and the environment where the users are to use the system.

In the context of this company, this was not a trivial task. The company worked in a rapidly changing mass market situation, with a product aimed at satisfying several market segments simultaneously. This is a situation that demands knowledge of end users and their current and expected future use of mobile phones, in a situation that can change over a short period of time; for example, if a competitor releases a new model with new and exciting features. Fortunately, there was a great deal of branch knowledge within the company, and within the different companies that the company worked together with. This meant that the people who were involved were familiar with the market, and were aware of what was happening in the marketplace, even though details of coming models are shrouded in secrecy. There is a great deal of information to be gleaned from many sources, and the company gained information about trends through participation in trade fairs, attending conferences and workshops, and kept abreast of publications within the field. They also gained input from mobile phone manufacturers, regarding future models that were designed to use the UIQ user interface. This information together with other branch knowledge meant that it was possible to understand and specify the use context.

The second principle is an appropriate allocation of function between users and technology, which means the specification of which functions should be carried out by the users and which by the technology. The second activity is therefore specifying the user and organisational requirements. A major step in most design processes is specifying functional and other requirements for products or systems, but in HCD this must be extended to create statements of user and organisational requirements in relation to the use context. This includes identifying relevant users, formulating clear design goals, and providing measurable criteria for testing the emergent design, that can be confirmed by users or their representatives. Once again, in a mass market situation, where the product is aimed at many potential customers, this is a potentially difficult task. In our study, the use context, and
therefore the requirements, is dependent on the type of user and their use of the phone. Once again, there was an extensive knowledge base regarding phone use, much of which was gained through having access to panels of users, and people who were recurrent phone testers. The company had access to user representatives who were long-term users, and focus group meetings and workshops were organised together with users. In this field, since the use of mobile phone technology is primarily individual and is not explicitly affected by organisational issues, we did not look in depth at organisational requirements. Therefore what follows deals explicitly with user requirements.

The third principle is the iteration of design solutions. Iterative design allows the results of testing “real world” design solutions to be fed into the design process. Combining user involvement with iterative design minimizes risks that systems do not meet requirements, and feedback from users is an important source of information. The third activity is therefore to produce design solutions. Potential solutions are produced by drawing on the state of the art, the knowledge of the participants, and the context of use. The process involves using existing knowledge to develop design inputs with multi-disciplinary input, concretizing the design through simulations, mock-ups, etc., presenting the design to users and allowing them to perform tasks – real or simulated – and altering the design in response to user feedback. This process should be iterated until design objectives are met. The interaction design department at UIQ has been a good example of the type of actor involved in this process. They had a mix of many different competencies, from graphical designers, interaction designers, and usability experts, working closely not only with customers and end users, but also together with development teams within their own organisation. This method of working allows the design and development of features that are acceptable in the marketplace, and are feasible to develop. The way in which UTUM has been used in this phase will be detailed more clearly in Section 4.

The fourth principle concerns multi-disciplinary design, and states that HCD requires a range of personnel with a variety of skills in order to address human aspects of design. The multidisciplinary design teams need not be large, but should be capable of making design trade-off decisions, and should reflect the relationship between the customer and the development organisation. The fourth activity is to evaluate the design against requirements. Evaluation is an essential step in HCD. It can be used to provide feedback that can improve design, can assess whether design objectives have been met, and can monitor long term use of the system. It should take place at all stages of the system life cycle (ISO 13407 1999). This is one of the activities where UTUM is the central factor in our study, allowing the evaluation and validation of solutions, from a feature level to the level of the complete product. For more information regarding this, see Section 4.

Our cooperation with UIQ technology has focused on usability issues at the company (Rönkkö, Winter et al. 2009). It began with a 12-month ethnographic (Rönkkö 2010) study that led to an attempt to implement the participatory design method (PD) Personas (see e.g. Grudin and Pruitt 2002; Pruitt and Grudin 2003) in 2001. The Personas attempt was abandoned in 2004, as it was found not to be a suitable method within the company for branch related reasons that can be found in (Rönkkö, Dittrich et al. 2005). To our surprise,
the original reason for introducing the PD method had disappeared during our attempt at method implementation, and internal socio-political developments had solved the power struggle that the PD method was aimed to mediate (Rönkkö, Hellman et al. 2008).

In parallel with the attempt to implement Personas in 2001, we also started to approach the new company goals focusing on metrics for aspects of the system development process. In our case these were focused on usability. Therefore, an evaluation tool was developed by the head of the interaction design group at Symbian (at that time the parent company for UIQ), Patrick W. Jordan, together with Mats Hellman from the Product Planning User Experience team at UIQ Technology, and Kari Rönkkö from Blekinge Institute of Technology. This evaluation tool was the first prototype of UIQ Technology Usability Metrics (see (Winter and Rönkkö 2009), chapter 4 and Appendix A). The first testing took place during the last quarter of 2001. The goal was to see a clear improvement in product usability, and the tests were repeated three times at important junctures in the development process. The results of the testing process were seen as rather predictable, and did not at that time contribute in a measurable fashion to the development process, but showed that the test method could lead to a value for usability.

The test method was further refined during 2004–2005 leading to the development of UTUM v 1.0 which consisted of three steps. The users involved in the testing could be chosen from the group of long-term users that were mentioned previously, or could be chosen from a list of people who had shown interest in participating in testing, or who had been recommended by other users. In the first step, a questionnaire about the user’s usage of a device was used to collect data that could be used to prioritize use cases and for analysis and statistical purposes. The use cases could then be decided by the questionnaire that the user had completed (in this choice Jordan’s level of functionality is visible) or decided in advance by the company, if specific areas were to be evaluated. After each use case the user filled in a short satisfaction evaluation questionnaire explaining how that specific use case supported their intentions and expectations. Each use case was carefully monitored, videotaped if this was found necessary and timed by the test expert. The second step was a performance metric, based on completion of specified use cases, resulting in a value between 0 and 1. The third was an attitudinal metric based on the SUS (Brooke 1986), also resulting in a value between 0 and 1. These values were used as parameters in order to calculate a Total Usability Metric with a value between 0 and 100. Besides these summative usability results, the test leader could also, through his/her observation during the test, directly feedback formative usability results to designers in the teams within the organisation, giving them early user feedback to consider in their improvement and redesign work.

In 2005, Jeff Winter joined the cooperative method cooperation as a new Ph.D. student, and began to study the field of usability, observed the testing process, and began interviewing staff at UIQ. At the same time, a usability engineer was engaged in developing the test. In an iterative process during 2005 and 2006, a new version of the test was produced, UTUM 2.0, which included more metrics, and new ways of interpreting and calculating the metrics. At this time, we also started to consider how the testing is related to the agile
movement and the principles of agile software (The Agile The Agile Alliance 2001), and the balance between agility and formality, and to the differentiation of usability results for different stakeholders needs (Winter and Rönkkö 2009).

The process of cooperation has been action research (Avison, Lau et al. 1999) according to the CMD methodology (Dittrich, Rönkkö et al. 2007). Action research is a “vague concept but has been defined as research that involves practical problem solving which has theoretical relevance” (Mumford 2001, p. 12). Action research is a combination of practice and theory where researchers and practitioners cooperate through change and reflection in a problematic situation within a mutually acceptable ethical framework (Avison, Lau et al. 1999). It is thus a merger of research and praxis, and on this basis it produces findings that are exceedingly relevant (Baskerville and Wood-Harper 1996). It involves not only gaining an understanding of a problem, and the generating and spreading of practical improvement ideas, but also the application of the ideas in a real world situation and the spreading of theoretical conclusions within academia (Mumford 2001). The researcher expects to generate knowledge that will enhance the development of models and theories (Baskerville and Wood-Harper 1996). The purpose of action research is to influence or change some aspect of whatever the research has as its focus, and improvement and involvement are central to it (Robson 2002, p. 215). A central aspect of action research is collaboration between researchers and those who are the focus of the research, and their participation in the process, and the terms participatory research and participatory action research are sometimes used as synonyms for action research (Robson 2002, p. 216). Action research often takes the form of a spiral process, involving observation, planning a change, acting and observing what happens following the change, reflecting on processes and consequences, then planning further action and repeating the cycle (Robson 2002, p. 217).

CMD is based on the assumption that software engineering is heavily influenced by its social contingencies. The development of CMD was motivated by a discontent with the way in which existing research approaches to software engineering and information systems addressed use-oriented software development, since they did not address the questions of how practitioners tackle their everyday work, from a shop floor point of view, or how methods, processes and tools can be improved to address practitioners’ problems, from their own point of view (Dittrich, Rönkkö et al. 2007).

CMD is a domain-specific adaptation of action research. It combines qualitative social science fieldwork, with problem-oriented method, technique and process improvement. The starting point for CMD is existing practice in industrial settings. Although it is motivated by an interest in use-oriented design and development of software, it is not specific for these methods, tools and processes. It is a three-phase framework that can be further detailed in relation to each specific project, but it can even be limited to a single phase if no improvement is decided upon. In phase 1, Understanding Practice, the research begins with empirical investigations, in order to understand and explain existing practices and designs from a practitioner’s point of view, based on their historical and situational context. The intention is to identify aspects that are problematic from the practitioner’s point of view. In phase 2, Deliberate Improvements, the results from the first phase are used in a cooperative
fashion by the researchers together with the practitioners involved, as an input for the design of possible improvements. This phase results in the deliberation of measures to address some of the identified problems. These measures are expected to improve the situation at hand. In phase 3, Implement and Observe Improvements, improvements are implemented. The researchers follow these improvements as participant observers. The results are evaluated together with the practitioners, and the concrete results are summarised for the companies involved. They build a base for the researchers to evaluate the proposed improvements (Dittrich, Rönkkö et al. 2007).

In addition to the three-phase research cycle, CMD is built upon a number of guidelines. These are:

- Ethnomethodological and ethnographically inspired empirical research, combined with other methods if suitable.
- Focusing on shop floor software development practices.
- Taking the practitioners’ perspective when evaluating the empirical research and deliberating improvements.
- Involving the practitioners in the improvements (Dittrich, Rönkkö et al. 2007).

This CMD approach has been used in cooperation with UIQ Technology AB in a large number of cycles during the period between 2001 and 2009, which have also had a profound impact on the development of CMD. For a more detailed explanation of CMD, and examples of its use in practice, see (Dittrich, Rönkkö et al. 2005; Dittrich, Rönkkö et al. 2007). The following presents an example of one of the CMD cycles that was applied in the research cooperation between UIQ technology and U-ODD.

In discussions during late 2005, we introduced the concept of the Kano model (see e.g. CQM 1993; Sauerwein, Bailom et al. 1996) into discussions together with the participants who were involved in developing the UTUM test. The following can also be read as an illustration of one of the CMD cycles applied in this study.

The first phase began in 2005 when the researcher was introduced to the company, and began his studies. He was given a workplace at the company, and spent time to become acquainted with the company and its organisation, whilst doing background studies of the area of interest. During this period, he interviewed members of staff, observed working procedures, and studied the development methodologies and project models in use at the company. At this stage, more than ten interviews were performed, and the roles included interaction designers and architects, technical writer, graphic designer, usability test leader, the head of the interaction design department, and the head of one of the development departments, responsible for Personal Information Management (PIM) applications. Concurrently, he performed literature studies in order to become acquainted with the theoretical background to the field.

The second phase was based on the work performed in phase one. One concrete example of this is that the researcher identified methods for presenting the results of usability testing as an area of interest, and had studied the theory and use of the Kano model (Sauerwein, Bailom et al. 1996). Based on this, research articles were sent to practitioners at the
company, and he organised discussion and seminars to examine how this could be of use in the usability evaluation framework that was being developed. As a result of the discussions and seminars, a new way of presenting the results of the usability testing was formulated, presenting the metrics as plots in a 4-field diagram (see Figure 5.2).

Figure 5.2 Presenting the results of a UTUM test (Rönkkö et al., 2009).

In phase three the researcher observed how the changes in presentation had taken place, took part in meetings where the results were presented to and discussed with senior managers, and discussed these changes with the practitioners, to gain an understanding of what the improvements meant in the everyday work of testing and presenting the results of usability testing. These cycles conformed to the principles given above, by using a combination of ethnographically inspired fieldwork, combined with document studies, in order to grasp and understand the practices of the workers who were involved in the everyday work of testing and designing the evaluation framework. The work was performed together with the practitioners, and the improvements were designed, tested and evaluated together with these practitioners.

5.4 The UTUM test

UTUM is one concrete result of the research cooperation mentioned above. It is an industrial application developed and evolved through long-term CMD cooperation between BTH/U-ODD and UIQ Technology. It is a usability test framework grounded in usability theory and guidelines, and in industrial software engineering practice and experience. It bridges a gap between the Software Engineering and the HCI communities. UTUM is “a
method to generate a scale of measurement of the usability of our products on a general level, as well as on a functional level”. Hornbæk (2006) states that amongst the challenges when measuring usability are to distinguish and compare subjective and objective measures of usability, to study correlations between usability measures as a means for validation, and to use both micro and macro tasks and corresponding measures of usability. Emphasis is also placed on the need to represent the entire construct of usability as a single metric, in order to increase the meaningfulness and strategic importance of usability data (Sauro and Kindlund 2005). UTUM is an attempt to address some of these challenges in the industrial environment. Due to space limitations, we cannot give an extensive presentation of UTUM here. It is presented in greater detail in (Winter 2009, chapter 4 and Appendix A). A video demonstration of the test process (ca. 6min) can be found on YouTube (BTH 2008).

Briefly, the test is performed as follows: the test is performed in an environment that feels familiar to the tester, in order that he or she feels at ease. In this case, we call the user a tester, as it is he or she who is testing the device, under the guidance of a usability expert, who we call the test leader. The tester is welcomed by the test leader, who explains the purpose and performance of the test. In the first step, the tester fills in a form that records their personal details and usual phone usage patterns. This information can be used to choose the use cases that are included in the test, or these use cases can be predetermined. The next step involves introducing the tester to the phone or phones to be tested. If more than one phone is tested, then all of the use cases are performed on one phone, before proceeding to the next phone. For each device, the tester is given a few minutes to get acquainted with the phone, to get a feeling for the look and feel of the product. The tester is then asked to fill in a Hardware Evaluation form, a questionnaire based on the System Usability Scale (Brooke 1986), which records their immediate impression of the phone. Then the tester is asked to perform the chosen use cases on the phone, whilst the test leader records details of the use case performance, and makes a judgement of the course of events. After each use case, the tester fills in another questionnaire, concerning task effectiveness, regarding the phone in relation to the use case. Between use cases the test leader can discuss with the tester what happened during the performance of the use case. In the final step, the tester fills in a System Usability Scale (Brooke 1986) questionnaire that expresses the tester’s opinion of the phone as a whole.

This process is repeated until all of the phones are tested. The data collected in the test, both qualitative and quantitative, are used to calculate a number of metrics for effectiveness, efficiency and satisfaction, which are then used to produce usability KPIs. As mentioned previously, the initial results of UTUM testing are useful in the fourth phase of an HCD process (ISO 13407 1999), when evaluating designs against requirements. The results that cannot be used in the immediate stage of development serve as input in the third phase of the next cycle in the process, when producing design solutions.

UTUM has a number of distinctive characteristics. The first two of these concern the relationship to users, and how user input is received and perceived, whilst a further two deal with software development practice, concerning organisation and method development in the company.
The first characteristic is the approach to getting user input and understanding users. Here we apply an ethnographic mindset (Rönkkö 2010). To understand the user’s perspective, rather than simply observing use, the test expert interacts and works with the users, to gain insight into how they experience being a mobile phone user. The users who help with the testing are referred to as testers, because they are the ones who are actually performing the test. The representative of the development company is referred to as a test leader, or test expert, emphasising the qualified role that this person assumes.

The second characteristic deals with utilizing the phone users’ inventiveness, and entails letting users participate in the design process. The participatory design (Rönkkö 2010) tradition respects the expertise and skills of the users, and this, combined with the inventiveness observed when users use their phones, means that users provide important input for system development. The test expert is the advocate and representative of the user perspective. User participation gives designers, with the test expert as an intermediary between them and the users, good input throughout the development process.

The third characteristic is continuous and direct use of user input in design and decision processes. The high tempo of software development for mobile phones makes it difficult to channel meaningful testing results to the right recipient at the right time in the design process. Integrating the role of the test expert into the daily design process eases this problem. The results of testing can be directed to the most critical issues, and the continual process of testing and informal relaying of testing results to designers leads to a short time span between discovering a problem and implementing a solution.

The fourth characteristic concerns presenting results in a clear and concise fashion, whilst keeping a focus on understanding the user perspective. The results of qualitative research are summarised by quantitative methods, giving decision makers results in the type of presentations they are used to. Statistical results do not supplant the qualitative methods that are based on PD and ethnography, but they capture in numbers the users’ attitudes towards the product they are testing. It is in relation to the third and fourth characteristic we have the most obvious relation to organisational metrics.

Although these four characteristics are highlighted separately, they are not simply used side by side, but are part of one method that works as a whole. Even though many results are communicated informally, an advantage of a formal testing method is traceability. It is possible to see if improvements are made, and to trace the improvements to specific tests, and thereby make visible the role of the user testing in the design and quality assurance process.

UTUM is a tool for guiding design decisions, and involves users in the testing process, since providing a good user experience relies on getting and using good user input. UTUM measures usability empirically, on the basis of metrics for satisfaction, efficiency and effectiveness, and a test leader’s observations. We have collected metrics for all of these aspects, based mainly on the industrial partner’s reliance on the ISO 9241-11:1998 (ISO 9241-11 1998), where usability is defined 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. This is probably the best-known definition of usability (Bevan
Effectiveness is the accuracy and completeness with which users achieve specified goals. Efficiency concerns the resources expended in relation to the accuracy and completeness with which users achieve goals. Efficiency in this context is related to ‘productivity’ rather than to its meaning in the context of software efficiency. Satisfaction concerns freedom from discomfort, and positive attitudes towards the use of the product.

ISO 9241-11:1998 states that it is necessary to identify goals and decompose effectiveness and satisfaction and components of the use context into sub-components with measurable and verifiable attributes when specifying or measuring usability. These metrics for effectiveness, efficiency, and satisfaction were used as key performance indicators (KPI) of product usability.

The inclusion of all of these metrics is based on the fact that it has been seen as important to consider effectiveness, efficiency and satisfaction as independent aspects of usability, and that all three should be included in usability testing (Frøkjær, Hertzum et al. 2000). According to Bevan (2006), when the purpose of a test is deciding whether a product has adequate usability, then measures should focus on the end result as defined in ISO 9241-11, and take into account the three aspects mentioned above. In a review of experimental studies, Frøkjær et al. (2000) found that many experimental studies of complex tasks account for only one or two aspects of usability, relying on assumptions of correlations between aspects of usability. However, there is some debate about whether there are correlations between these aspects, and, e.g., Sauro and Kindlund (2005) suggest that there are correlations, and propose a single score for usability metrics based on their findings. In work in progress, we are currently examining the metrics collected in a series of tests to determine whether we can find correlations between the different aspects of usability measured in the test.

If we return again to the principles and activities of HCD (ISO 13407 1999), it is in the fourth phase that the initial results of UTUM testing are used. This is the phase where designs are evaluated against requirements. The results of testing are both in the form of metrics and as knowledge possessed by the test leader. These results can be communicated in different fashions to different stakeholders, on both a design and development level, and also on a management level. They affect features of the product that are soon to be included in a version for release to the market. However, some of the results dealt with aspects of the product that are no longer possible to influence or change at this particular stage of development, and these results are seen as candidates for future use when designing and developing design solutions. Thus, it is in the third phase, which allows the results of testing “real world” design solutions to be fed into the design process, that the results of previous cycles of UTUM testing can be used, as a form of existing knowledge, allowing the use of results that were impossible to include in the development that took place when the testing was performed.

Thus, we see that UTUM is part of a HCD based development process, and gives a demonstration of quality of use, defined as the extent to which a product satisfied stated and implied needs when used under stated conditions (Bevan 1995), from the customer and end-user point of view (Winter, Rönkkö et al. 2007). UTUM became a company standard
from January 2007, and was used in all software development projects. It was also frequently requested by external clients. So far, it has mainly been used for measuring usability and quality in use, but the ambition is to adapt the tool to capturing user experience to a much larger degree than it does today. Due to the closure of UIQ Technology AB, the challenge of UX is continued within a new research project within the telecommunication branch called WeBIS, with the purpose of producing industrially viable methods for user generated innovation in ICT services (WeBIS 2008).

5.5 SPI vs. usability testing metrics

In this section, we compare the findings of Iversen and Kautz (2001) (see table 5.1) with the situation that we found in our study. We point out in which way our findings agree with their principles, but also where we shed light on new aspects of the principles. We also set our work in relation to the key organisational factors for success in SPI as listed by Dybå (2005) (see table 5.2).

<table>
<thead>
<tr>
<th>Area</th>
<th>Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>1. Use improvement knowledge</td>
</tr>
<tr>
<td></td>
<td>2. Use organisational knowledge</td>
</tr>
<tr>
<td>Organisation</td>
<td>3. Establish a project</td>
</tr>
<tr>
<td></td>
<td>4. Establish incentive structures</td>
</tr>
<tr>
<td>Design</td>
<td>5. Start by determining goals</td>
</tr>
<tr>
<td></td>
<td>6. Start simple</td>
</tr>
<tr>
<td>Communication</td>
<td>7. Publish objectives and collect data widely</td>
</tr>
<tr>
<td></td>
<td>8. Facilitate debate</td>
</tr>
<tr>
<td>Usage</td>
<td>9. Use the data</td>
</tr>
</tbody>
</table>

In the following, we discuss how Iversen and Kautz’ five areas and nine principles for successful SPI compare with our research and our implementation of usability metrics. The areas and principles are dealt with in the same order as they are listed by Iversen and Kautz; interleaved in this, we also discuss our findings in relation to Dybå’s key factors for success in SPI.

5.5.1 Area 1: knowledge

When implementing metrics, it is important to draw on different kinds of knowledge, including knowledge about software improvement and organisational change, but also knowledge about the actual organisation where the metrics program is to be implemented. To successfully implement metrics programs, members of the organisation should be
knowledgeable about the art of software metrics, software improvement, software engineering, and the process of organisational change.

Principle 1 concerns the use of improvement knowledge. In our case, to understand and approach the industrial practice we have applied the three action research phases from CMD (Dittrich, Rönkkö et al. 2007). With the help of action research (Avison, Lau et al. 1999) and ethnography (Rönkkö 2010) we identified which aspects are most problematic for the practitioners, i.e. based on a historical and situational context, from their ‘own point of view’. Results from the first phase were used by the researchers together with the practitioners involved as input for the design of possible improvements. The outcome was the deliberation of measures that address the identified problems expected to improve the situation at hand. Improvements were implemented, and the researchers followed these improvements as participant observers. The results were evaluated continuously together with the practitioners.

Principle 2 is to use organisational knowledge. Many metrics programs fail because there is a gap between work procedures as they are described, and the actual work practices, and thereby there is a lack of understanding of the organisational context. For the program to succeed, the actors must understand why they are collecting particular data, and even understand the organisational politics. This can be done by including as many of the affected employees as possible (Iversen and Kautz 2001, p. 295). This principle is connected to several of Dybå’s key factors. The first is employee participation, which is the extent to which employees use their knowledge and experience to act and take responsibility for SPI (Dybå 2005, p. 412). This was found to be the factor that had the strongest influence in predicting SPI success, perhaps because people tend to support what they have participated in creating (Dybå 2005, p. 419). The use of organisational knowledge is also connected to the factor of business orientation, which concerns the extent to which goals and actions are connected to explicit and implicit business goals and strategies (Dybå 2005, p. 411). Organisational knowledge is created through the experience and experimentation of members of the organisation, and it is important that both practitioners and researchers direct their efforts towards understanding the knowledge that is shared between groups within the organisation. This demands respect for the expertise of others, an understanding of the relevance of this expertise, and sufficient knowledge of the other group’s problems to be able to communicate about them. In its turn, this demands that the members of the different groups have had actual experience of the activities of the other groups (Dybå 2005).

This supports the focus that our research has placed on participatory design (Schuler and Namioka 1993). In our case, this has meant that, besides cooperating with interaction designers through CMD, we also targeted different organisational roles in our usability test metric efforts, i.e. this was a way of including many employees at many levels. Results from the UTUM were addressed to the following roles in the company: interaction designers, system- and interaction architects, high and low level management, product planning, and marketing.
We performed a case study to enable adjusting our work to the work practice of those stakeholders, to answer questions such as: Are any presentation methods generally preferred? Does the choice of methods change during different phases of a design and development project? Can results be presented in a meaningful way without the test leader being present? Is it possible to find factors in the data that allow us to identify disparate groups, such as Designers (D), represented by e.g. interaction designers and system and interaction architects, representing the shop floor perspective, and Product Owners (PO), including management, product planning, and marketing, representing the management perspective (Winter 2009) (see also (Winter, Rönkkö et al. 2008) for more background to this). In our cooperative method implementation, we have included reasoning about finding a balance between formality, as we have defined it previously, and organisational needs from an agile, practical viewpoint of “day to day” work.

When referring to “formal aspects” we relate to the fact that software engineers are generally trained in a hierarchic model of knowledge where abstraction, together with the production of general solutions rather than specific solutions, is the key part. The software engineer’s systematic work often aims to produce an overview as a basis for simplifying, framing and explaining suggested findings/solutions from an objective and repeatable outside point of view (see also (Rönkkö 2010): Sections 3 and 4). Improving formal aspects is important, and software engineering research in general has successfully emphasised this focus. However, improving formal aspects may not help design the testing that most efficiently satisfies organisational needs (Martin, Rooksby et al. 2007) and minimizes the usability testing effort. A consequence of using “best practice” models, which is closely connected to a concentration upon “formal aspects”, is the fact that routines can become institutionalised. Thereby, they become a hinder to the exploration of alternative routines, and can become a barrier to improvement (Dybå 2003). The emphasis we have placed on agility is a result of the concentration that we have had on examining the practical viewpoint of the study participants.

The conflict between formality and agility is connected to the ideas regarding learning strategies found in (Dybå 2005, p. 412) where Dybå states that a learning strategy deals with the way in which an organisation is engaged in the exploitation of existing knowledge or the exploration of new knowledge. This deals with finding a balance between discipline and creativity, where relying too much on existing knowledge can limit the behaviour that is necessary for improvisation, whilst too much risk taking can lead to failure through fruitless experimentation (Dyba 2000b). Dybå argues that both exploitation and exploration are necessary, involving balancing the improvement of an existing skill base with experimentation around new ideas. Both of these modes of learning were found to be connected to SPI success (Dybå 2005).

The concept of exploitation is related to the formal aspects of our testing results, where the group we have identified as Product Owners has a long-term need for structured information allowing them a clear view of the product and how it is developing over time. However, the formal view needs to be juxtaposed with reasoning based on a ‘day to day’ basis of organisational needs. The latter viewpoint has been recognised in software
engineering in the form of: (a) “agility”, by practitioners in the field; and (b) “work practice”, by sociologists and software engineers studying the influence of human aspects in software engineering work. For ‘agility’ and usability testing see (Talby, Hazzan et al. 2006; Winter, Rönkkö et al. 2007). For examples of ‘work practice’ and software engineering see (Rönkkö 2002; Rönkkö, Dittrich et al. 2005; Martin, Rooksby et al. 2007; Rönkkö 2010, Section 8). This is connected with the exploration of knowledge, which includes such factors as innovation, creativity, flexibility, and diversity (Dybå 2005). The concept of exploration also appears to be connected to the concept of employee participation.

A clarification is in order so that our intention is not misunderstood. It would be difficult to gain acceptance of the test results within the whole organisation without the element of formalism. In sectors with large customer bases, companies require both rapid value and high assurance. This cannot be met by pure agility or through formal/plan-driven discipline; only a mix of these is sufficient, and organisations must evolve towards the mix that suits them best. In our case this evolution has taken place during the whole period of the research cooperation, and reached a phase where it became apparent that this mix is desirable and even necessary.

Based on studies of work practice we find that we have managed to strike a successful balance between agility and formalism that works in industry and that exhibits qualities that can be of interest to both the agile and the software engineering community (see also Winter, Rönkkö et al. 2007).

5.5.2 Area 2: program organisation

If a metrics program is to have a significant impact, you must address the organisation of the program itself, and the incentives that should support it. This is dependent on what Dybå calls involved leadership, which is the extent to which leaders at all levels in an organisation are committed to and participate in SPI (Dybå 2005, p. 411). Dybå states, however, that management commitment need not go beyond the level of allocating the necessary resources. In our case, the commitment of the leadership is shown in several ways. One of these is the fact that the company allocated resources for participation in the BESQ research cooperation (BESQ 2008). In this particular case, the company contracted to co-finance their participation through the work performed by the interaction design team, and the staff that had usability testing and the development of the usability test framework as part of their everyday work activities.

Principle 3 is to establish a project. To increase visibility and validate expense, the metrics program should be given the status of a formal project, with requirements for planning and reporting progress and with success criteria. The introduction of metrics into projects should also be planned (Iversen and Kautz 2001, p. 296). This is also connected to the key factor of employee participation, which is the extent to which employees use their knowledge and experience to act and take responsibility for SPI (Dybå 2005, p. 412). In our case, the metrics program was, as previously mentioned, part of a research cooperation where the company financed their participation in the research environment by direct work involvement (see BESQ 2009).
Chapter five
SPI Success Factors

Principle 4 is to establish an incentive structure, so that those who report the metrics see some advantage in their participation in the program. The results of the program will hopefully benefit the employees in their daily work, but a more indirect approach can be taken, by using bonuses and awards to facilitate adoption of the metrics program. Tools and procedures should also be developed to simplify data collection, avoiding unnecessary burdens being placed on employees (Iversen and Kautz 2001, p. 297).

In our case, it was not necessary to provide an incentive structure, as the work performed by the participants was a part of their everyday planned work activities and they were positive to the research cooperation together with the researchers. We did establish procedures to simplify data collection, but as mentioned, the metrics program was manned primarily by staff whose main task was user research, so collecting metrics was part of their work tasks rather than an additional burden. These were the same people who were involved in developing the usability test in cooperation with researchers from academia. Applying a work practice perspective helped us to adequately understand and adequately address other stakeholders.

5.5.3 Area 3: program design

Just as the metrics organisation needs to be designed, so does the content. The metrics need to be organised and designed. This is also connected to Dybå’s key factor of business orientation. It is also dependent on management commitment, where the role of managers and leaders is to create a situation that is supportive of learning and SPI (Dybå 2005).

Principle 5 is to start by determining goals. Successful implementation needs clear goals from the beginning, since without these, the effort is impossible to manage, and decisions on which measures to choose will be random (Iversen and Kautz 2001, p. 298). This is closely connected to Dybå’s key factor of business orientation, and it concerns the need to focus on the alignment of SPI goals with business goals in both the long and the short term. If the SPI strategy is not kept in line with the business strategy, there is a risk that the processes simply become a burden (Dybå 2005).

In our case, we have always concentrated on adapting the metrics to the business needs and strategy of the organisation. In relation to each development project where usability was tested, prioritized use cases were decided by clients, and metrics measured based on these. The overall usability goals were set in an evolutionary fashion over a long period of time. Iversen and Kautz (2001, p. 303) experienced difficulties to create a baseline for long-term measurement of improvements, since measurements were made whilst the organisation was improving. We also found creating a baseline difficult, partly because the marketplace was constantly changing, but also because the test methodology itself was in a constant state of change. It was only recently, when the test was beginning to stabilise, that efforts were made to create a baseline for usability testing. Due to the closure of the company this did not mature to the extent we wished for.

Principle 6 is to start simple. Starting with a small set of metrics is to be recommended, as systematically collecting data to use as the basis for decision making is difficult and complex. Also, it may be found that if the right metric is chosen, it will be possible to fulfil

:: 159 ::
the predefined goals through the use of one simple measure, if you choose the right metric from the beginning (Iversen and Kautz 2001, p. 299). This is connected to the key factor of a concern for measurement, where Dybå (2005) finds that it is better to collect a few measures that are directly related to the process, rather than many measures that lead to a lack of focus and confusion about what is important. Translated to our situation, where we collect product metrics rather than process metrics, this means that we began by choosing a limited number of suitable measures for the product. These are the metrics for efficiency, effectiveness and satisfaction that are collected in the UTUM test. We started on a small scale in the year 2001. The results of the first testing process were seen as predictable, and did not at this time measurably contribute to the development process, but showed that the test method could lead to a value for usability (see (Hellman and Rönkkö 2008b) for an overview). Thereafter the metrics evolved continuously over a period of time.

5.5.4 Area 4: communication

Metrics must be interpreted, since measuring the performance of professionals is not an exact science. A metrics program that attempts to do this can suffer from negative feelings, and this can be countered by communicating the objectives and results of the program, and encouraging discussion about the validity and reliability of the results. This is related to Dybå’s key factor of a concern for measurement, the extent to which the organisation collects and utilises quality data to guide and assess the effects of SPI activities (Dybå 2005, p. 412). Emphasis is placed on using the metrics to improve the improvement process itself: the ability to measure aspects of the process, and make judgements of the process based on the measurements.

In our case, we have not measured the success of the metrics process itself. Given the focus that usability testing places on collecting metrics for the product rather than the process, this factor must be translated to our situation. Dybå found that it is necessary to feed the data back to the members of the organisation, and this has been performed in our case by presenting the results of usability testing to different stakeholders at many levels within the organisation, and by encouraging discussion of the results and their meaning. For more details regarding this, see below, regarding principles 7 and 8, on how the results were published and the debate that took place. In improving the usability testing process itself, improvements have not been based on measurement of the process, but have been made through the iterative research and development process that has taken place over a period of years.

Principle 7 is to publish objectives and collected data widely. The objectives of the program must be communicated as widely as possible, and results must be published broadly for as many relevant actors as possible. It is important to ensure that individuals are protected, and that metrics are not related to performance evaluations. Published metrics must be based on reliable and valid data, in order to support fruitful discussion. As metrics may reveal unpleasant facts about an operation, it is important to use the figures in order to improve the organisation rather than to find scapegoats for the problems that have been uncovered (Iversen and Kautz 2001, p. 300). This also has connections to Dybå’s factor of concern for measurement, where it is stated that the data must be perceived as accurate and
valid, and must be used in a non-threatening way to identify and solve problems (Dybå 2005).

In our case, all usability test results were published on the company’s intranet as a news item for all employees to read and comment on. Test results were presented to specific teams connected to the test in question. Later on the company used the UTUM test results, based on the metrics for efficiency, effectiveness and satisfaction, to calculate a Usability KPI. The quality team at UIQ was responsible for collecting data for each KPI and the usability person in charge sent the latest UTUM data to this team. The Quality Team then put together a monthly KPI report to the Management team. The data were not specifically sensitive in that they did not collect metrics about the performance of individuals. They were aimed at measuring the usability of a product at a particular time and were not traceable to any individual within the organisation. The user researcher also presented the metrics to two different groups of stakeholders: the Designers and Product Owners mentioned earlier (Winter, Rönkkö et al. 2008). For more information about these KPI’s, see Section 4 in this article, where we describe the UTUM test.

Principle 8 is to facilitate debate. There should be a forum for discussing the metrics program and its results, to clarify how the material that is collected will be used, in order to prevent the creation of myths about the material and its use (Iversen and Kautz 2001, p. 301).

In our case, there was much discussion in the company surrounding KPI’s, about what they did and could communicate. These were tough discussions, and strong opinions were often expressed, e.g. that KPI’s might provide management with false impressions that could lead the organisation in the wrong direction. Also that the ‘form’ was focused on rather than the ‘content’, and what the consequences of this might be in time-critical situations. There were also opposing opinions, that KPI’s could give the organisation a sharper focus. Most debates were about what was measured, the importance of it, and for whom the metrics were produced. And of course questions related to consequences for the staff were ventilated, i.e. what the introduction of specific measurements might mean for them and their roles in the long term.

5.5.5 Area 5: data usage

The data that is collected must be used to implement improvements or increase understanding about how the organisation works. If data is not applied, those who supply them are not likely to supply further data, resulting in deterioration of the data quality. Principle 9 is to use the metrics to gain insight into software processes, and correct the problems. If this is not done, the metrics program may turn into a procedure that merely adds to development overhead. If there are no consequences of poor results, the metrics program is unlikely to succeed, but it is important to bear in mind that software development cannot be measured precisely, and that the data must not be over-interpreted. However, recognising trends from imprecise data is still better than having no data at all (Iversen and Kautz 2001, p. 302).
In relation to this principle we found for example that adequate knowledge to interpret numbers in a critical and correct way required thorough work and relevant involvement (often it was only the test leader who was in a good enough position to make such interpretations). And since numbers actually can get in the way of communicating ‘the message’ we decided to use graphs without the use of numbers as a presentation form; graphs that give a quick summary of the knowledge experienced by the testers.

In our case there were two main channels and target groups for implementing improvements, i.e. the previously introduced groups of stakeholders designated as Product owners and Designers. As previously defined, PO included management, product planning, and marketing, representing the management perspective, whilst Designers were represented by e.g. interaction designers and system and interaction architects, representing the shop floor perspective. The distinction between these groups was connected with finding a balance between formality, and organisational needs from an agile, practical viewpoint of “day to day” work. Using the term formal in contrast to agile does not imply that we see agile processes as being informal or unstructured, but because the word “formal” is more representative than e.g. “plan driven” to characterise the results of testing and how these results are presented to certain stakeholders.

In relation to the first group, the PO, data usage is focused on comprehensive documentation consisting of spreadsheets containing the formal side of metrics and qualitative data. It is important to note that we see this formal element in the testing as an increased use of metrics that complements the qualitative testing method side. Not the other way around. Metrics back up the qualitative findings that have always been the result of testing, and open up new ways to present test results in ways that are easy to understand without having to include contextual information. They make test results accessible for new groups. The quantitative data gives statistical confirmation of the early qualitative findings, but are regarded as most useful for PO, who want figures of the findings that have been reached. There is less pressure of time to get these results compiled, as the most important work has been done, and the critical findings are already being implemented. The metrics can also be subject to stringent analysis to show comparisons and correlations between different factors.

In relation to the second group, the Designers, the high tempo of software development in the area of mobile phones makes it difficult to channel meaningful testing results to the right recipient at the right time in the design process. To alleviate this problem, we have integrated the role of the test expert into the daily design process. Directly after or during a period of testing, the test leaders meet and discuss findings with this group. This can take place even before all the data is collated in spreadsheets. In this manner the test experts are able to present the most important qualitative findings to system and interaction architects within the organisation very soon after the testing has begun. It was also shown that changes in the implementation have been requested soon after these meetings. An advantage of doing the testing in-house is to have access to the tester leaders, who can explain and clarify what has happened and the implications of it – just in time - when that information is desired by these actors. And the results of testing that is performed in-house
can be channelled to the most critical issues, and the continual process of testing and informal relaying of testing results to designers leads to a short time span between discovering a problem and implementing a solution.

Table 5.3 Summary of findings

<table>
<thead>
<tr>
<th>Area</th>
<th>Confirmation or New Knowledge</th>
<th>Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>C</td>
<td>We confirmed that it is important to use the knowledge within the organisation to improve and evaluate the program, to adapt to the environment and the situation at hand</td>
</tr>
<tr>
<td>Organisation</td>
<td>C</td>
<td>Our study confirms the importance of allocating resources, and in particular we found that it was an advantage that resources were dedicated to the program, minimizing the burden of “extra” work</td>
</tr>
<tr>
<td>Design</td>
<td>N</td>
<td>SPI needs clear goals from the beginning, but we also found the importance of adapting to a rapidly changing world by adopting a flexible approach. This means that there must be a readiness to change the goals during the ongoing process.</td>
</tr>
<tr>
<td>Communication</td>
<td>N</td>
<td>It is not only important to base decisions on measurable phenomena – it is also important to include the knowledge that is based on intuition and experience, and that it was the test leader who was most suited to communicating this knowledge.</td>
</tr>
<tr>
<td>Usage</td>
<td>N</td>
<td>We also found that it is important to have both agile and formal results, to satisfy the needs of different stakeholders. The formal results support the agile results and ensure that testing results are accountable in the organisation</td>
</tr>
</tbody>
</table>

The main findings detailed above are summarised in table 5.3, showing in which areas our study confirms the findings of Iversen and Kautz, and where our material extends their findings with new knowledge. For each area, a confirmation is marked with a C in the second column, whilst new knowledge is marked with an N. This is discussed in more detail in the next section.

This concludes our comparison of the findings of Iversen and Kautz, the discussion of Dybå’s key factors, and the findings from our research. The comparison is made here to help the reader assess the applicability of the principles for companies attempting to implement software metrics programmes and usability programmes. We proceed with a discussion of what we have found, primarily where our study gives new insights into the principles, and what we can learn that may help organisations improve testing and improvement efforts.

5.6 Discussion and conclusions

Beginning with a comparison of the principles listed by Iversen and Kautz, and the results of our research, there are a number of findings that are interesting in light of the situation that we have presented above.

:: 163 ::
Iversen and Kautz found that an organisation must adapt its culture to practices and a way of thinking where decisions are based on measurable phenomena rather than intuition and personal experience. This was also found in the case of UIQ. However, having said this, our experience shows that the test leader is a primary source of important knowledge that is based on intuition and personal experience. Most importantly, we found that the metrics based on these factors are just as relevant in the metrics program as the other types of metrics.

The success of a metrics program was found to be most likely in a situation where measurement does not place a particular burden on the practitioners, and where incentives for participation were offered. In our case, there was a test leader who was assigned to the collection, analysis and presentation of test results, who could request further resources when necessary, and the testing work was considered to be an integral part of the development process, and was included in estimates made when planning project activities. This meant that the metrics activities were not seen as an extra burden on top of the everyday work. The focus on measuring the product rather than the process means that others within the organisation are not burdened by collecting information or through the use of the information. This is important to consider when designing and implementing both testing programs and improvement processes, emphasising the importance of ensuring that measurements are performed by specialists who are well versed in the structure and operations of the organisation.

Successful implementation of SPI is said to need clear goals from the beginning or the effort will be difficult to manage and decisions on which measures to choose will be random. Our experience illustrates another aspect of this, and shows that we must accept that the dynamics of the marketplace, where rapid change is becoming the norm, make it difficult to achieve stability. There is a collision here, where those who deal with processes apply their lessons to the product, implying that improvements in the process automatically lead to improvements in the product. We find that improving the product requires a degree of flexibility, and suggest that this may be applicable within the field of SPI. We have found that it is important to follow and adjust to a rapidly changing world by adopting an agile approach to the design and performance of the testing or improvement program.

It is important to continually improve and evaluate the metrics program. This has always been an integral part of the development and operation of the UIQ metrics program, and has taken place through discussions and workshops together with researchers, practitioners, management, and even outside companies. It is crucial to adapt the principles to the environment and the situation at hand. The work that has been done here, leading to the success of the metrics program, and its adoption within the everyday development activities shows that this has been a successful strategy in our case.

Concerning the use of the data, this has been one of the main focuses of our recent research. We have found it to be of the greatest importance that the data, the results of the testing, are used in ways appropriate to the needs of the recipients. This demands both agile results, for example in the form of direct feedback from the test leader to a designer or developer, and formal results presented in different ways for different groups of
stakeholders. The formal side of testing is necessary as a support for the agile side, and ensures that testing results are accountable within the organisation, thus raising the status of the agile results. This balance was not discussed amongst the principles listed by Iversen and Kautz, although it is connected to Dybå’s discussion regarding learning strategies and the exploitation or exploration of knowledge.

We find that the test leader is a key figure, since this is a person who owns the requisite knowledge to understand and explain the results and their meaning, and can work as an advocate of the user perspective. It is vital to find the right people, who can function as test leaders, analysts, communicators, and advocates of the user perspective. Within usability and user experience testing it becomes apparent that the test leader needs to possess traits other than those demanded in the engineering and technical areas. Here, the focus is not directed towards understanding and discovering faults in software; it is instead directed towards understanding human beings, and how to best treat them to get them to reveal their attitudes and feelings towards and understanding of our products – things that they themselves may sometimes be unaware of – and even discover why they feel and act the way they do. We have for example found that some of the successful test leaders have a background in healthcare or the social services.

To summarise the concrete lessons for product focused usability work that can be learnt from the discussion above, which we suggest may also be applicable to SPI, we find that it is important to:

• ensure the presence and participation of specialists who are well versed in the structure and operations of the organisation,
• adopt a flexible approach to the design and performance of the programme, to adapt to changes in the environment,
• adapt the way results are presented in ways appropriate to different groups of stakeholders, to support the emphasis that different groups place on agile or formal results,
• find suitable people to work as key figures when collecting metrics, when analysing and communicating results, and who can act as advocates of the user perspective.

To conclude, this work shows the role usability testing can play in both product and process improvement. Just as SPI has been found to be a useful vehicle for improvements in software engineering, so is usability testing found to be a useful vehicle that is well in accordance with the principles of SPI and is simultaneously in line with a focus on business orientation. This is a fact that emphasises the usefulness of usability testing, which is an area that has been neglected in the area of software engineering. As such, we can make a contribution to the field of software engineering.

However, there are still areas that need to be addressed, and one of these is the need to create a structured approach to user-centred design processes. Despite the similarities between usability testing and SPI, usability testing is not equivalent to SPI. In the same way as there is a need for SPI, there is also a need for usability process improvement. As previously mentioned in Section 3, it is important that international standards for usability should be more widely used, and that they should be used to improve usability capability.
within an organisation. The standards that exist provide a framework for how user-centred design should be practiced, and can have a serious impact on HCI and usability practices (Bevan 2001). Earthy et al. (2001) claim that the implication of ISO 13407 and ISO 18529 is that there is a professional responsibility for software engineers, system engineers and usability professionals to adopt good practice as defined in ISO 13407 and ISO 18529 as their baseline.

Taking the lessons learned from our study, and relating them to usability process improvement based on these international standards may lead to new insight in how to include these lessons in the detailed practices specified in the standards. Finally, the study presented above is an illustration of how two paradigms can interact, by showing how the metrics could be tailored to allow professionals schooled in a qualitative tradition to interact together with professionals schooled in a quantitative engineering tradition, and that it is possible to balance these two traditions that are often seen as conflicting.

Acknowledgments
The authors would like to thank the anonymous reviewers and the editor for their helpful advice and suggestions. This work was partly funded by The Knowledge Foundation in Sweden under a research grant for the software development project “Blekinge – Engineering Software Qualities” (BESQ 2008) and the Vinnova Foundation in Sweden under a research grant for WeBIS (WeBIS 2008), dealing with method development, mixing innovation and usability.
Chapter Six

Article Two
Chapter Six

Identifying organizational barriers - A case study of usability work when developing software in the automation industry

Under review at Journal of Systems and Software

Jeff Winter, Kari Rönkkö, Mikko Rissanen

Abstract

This study investigates connections between usability efforts and organizational factors. This is an important field of research which so far appears to be insufficiently studied and discussed. It illustrates problems when working with software engineering tasks and usability requirements. It deals with a large company that manufactures industrial robots with an advanced user interface, which wanted to introduce usability KPIs, to improve product quality. The situation in the company makes this difficult, due to a combination of organizational and behavioural factors that led to a “wicked problem” that caused conflicts, breakdowns and barriers. Addressing these problems requires a holistic view that places context in the foreground and technological solutions in the background. Developing the right product requires communication and collaboration between multiple stakeholders. The inclusion of end users, who fully understand their own work context, is vital. Achieving this is dependent on organisational change, and management commitment. One step to beginning this change process may be through studying ways to introduce user-centred design processes.

6.1 Introduction

Product quality is becoming one of the dominant success criteria in the software industry. One of the challenges for research is to provide the industry with the means to deploy quality software, allowing companies to compete effectively (Osterweil 1996). Quality is multi-dimensional, and impossible to show through one simple measure, and research should focus on identifying various dimensions of quality and measures appropriate for it (Osterweil 1996). Even though this was written in 1996, it is still valid today, when companies are competing in a rapidly changing and competitive global market.

Usability is one of the important quality factors. However, despite the fact that our study focused on usability, we do not present details concerning the theories and methods of usability work. The idea behind this article is not to discuss the technicalities of usability,
even though usability has been the vehicle which we have focused upon. For more information regarding usability, see e.g. (Nielsen 1993; ISO 9241-11 1998; Dumas and Redish 1999; ISO 13407 1999; Bevan 2001). Furthermore, we see that the situation that we have uncovered is probably not limited to the problems of performing usability work, but is representative of many types of settings, where complex issues are dealt with in complex organizations, and could therefore be of interest in many different situations.

The context of this study is a large multinational industrial company, with roots in a traditional engineering field, and extensive experience of working with industrial machines. One branch of the company produces and markets industrial robots and robot systems. The company acknowledges the importance of incorporating concepts of usability in their design and development processes, to improve the quality of their product. The entrance point to this study was a desire to find ways of measuring usability on the basis of Key Performance Indicators (KPIs). Given our background in industrial cooperation, and our focus on usability and user experience, we were invited to perform a case study together with the corporate research department. The study was intended to result in new knowledge of how the industrial organization currently works with usability, and should lead to new theories of the problems involved in such work, and ideas that could facilitate this work.

Questions pertaining to the usability of robot systems are important to study. They deal with the changing conditions for traditional engineering organizations, where it becomes necessary to change the way of thinking when designing and developing products, in order to remain competitive in a rapidly changing and competitive global market. For a producer of robot systems in today’s market, it is important to design systems that can be used by domain experts but not robotics experts (Scholtz 2003). To study this field, practical real-world observations are important, as is the inclusion of experts from multiple domains in research efforts (Goodrich and Schultz 2007).

In the context where we have performed the study, the area of operations is designing and producing industrial robots, so the study deals indirectly with human-robot interaction (HRI), which is the study of teams of humans and robots, and the ways they influence each other (Scholtz 2003). HRI is concerned with understanding and shaping interactions between one or more humans and one or more robots. Essential components of HRI are evaluating the capabilities of humans and robots, and designing the technologies and training that produce desirable interactions (Goodrich and Schultz 2007). One of the primary goals of HRI has been to investigate natural means by which humans and robots can interact and communicate (Dautenhahn 2007). HRI is related to human-computer interaction (HCI) and human-machine interaction (HMI), but differs from them since it concerns systems that have complex, dynamic control systems, which can exhibit autonomy and cognition, and which operate in real-world situations (Fong, Thorpe et al. 2003). It is important to consider the interactions that humans will have with robots, and what software interface architecture and user interface can accommodate humans (Scholtz 2003).
In this case, we investigate the areas of robotics and HRI from the perspective of the organization and organizational processes where the HRI is designed and developed. It is impossible to understand transformations occurring in the nature of work and organizing without considering both the technological changes and the institutional contexts that reshape organizational activities (Orlikowski and Barley 2001). Therefore, it is important to focus on the situation within the organization. We must understand how human and organizational factors affect the execution of software development tasks, as many studies show how behavioural factors have a greater impact on software productivity than tools and methods (Curtis, Krasner et al. 1988). Orlikowski and Barley (Orlikowski and Barley 2001) stated that “IT researchers have yet to ask how institutions influence the design, use, and consequences of technologies, either within or across organizations”. This study looks more closely at these aspects. The primary intention of the work at hand is thus to study the processes and communication within the organization, with particular focus on the role played by usability. This work has been performed as a case study (Yin 2003). The data in the case study has been analysed in a grounded theory (GT) approach. (Robson 2002).

The study illustrates a situation where the company successfully produces and sells a complex product that technically is at the front of the market, where adding features has been a strong selling point, and where usability is in line with other products of similar complexity and ability. There is however a growing understanding that usability is important for future sales, where a successful product has a good balance of price and quality, which includes both features and usability. The initial focus of our study concerned the introduction of KPIs, but we did not drive the process as far as finding ways to formulate candidates for these KPIs. However, we do show a number of areas where breakdowns occur in the organization, and see that a combination of different factors means that the company has difficulty in reaching the goal of formulating measurable usability KPIs. The factors involved combine to create what Rittel and Webber (1973) call a “wicked problem”. To address these problems, we find that it is important to take context into account when designing and developing the product, and that a process of user centred design (UCD) may be a way to address the identified situation. However, we also see that in order to reach the stage where UCD is possible, and to address the issues that we have found, would require a shift in perspective within the organization. There must be an expansion of the understanding of the importance of usability, a commitment to making usability a central factor, and incentives for improving usability. There must be an overview of what is important to focus efforts upon. This overview could lead to useful KPIs, and a system of incentives to work with these KPIs, and a situation where the importance of UCD pervades the organization. However, in the situation that exists today it is difficult to create this overview, because of the factors that we detail below.

In the following, we discuss our methods in section 2, and then in section 3 present a detailed illustration and summary of the problems that we uncovered during the study. This is followed in section 4 by a discussion of the causes of these problems, in relation to work
from a number of related areas, and we discuss possible solutions to some of the problems. The article ends with Section 5 with conclusions and some ideas for future work.

6.2 Methods

The study was performed as an interview-based case study. A case study focuses on a particular case, taking the context into account, and involves multiple methods of data collection; the data can be both qualitative and quantitative, although qualitative data are almost always collected (Robson 2002, p. 178). A case study approach allows the retention of characteristics of real life events, including organizational and managerial processes (Yin 2003). Although the study has been performed in a limited part of a large organization, and it could therefore be claimed that it lacks relevance in a larger context, it has been informed by an ethnographic approach (Rönkkö 2010), which provides an analytic focus that allows a focus on the types of challenges that are common for many software development projects, in complex design and development situations where many stakeholders are involved (Rönkkö, Dittrich et al. 2005).

The material in this study was collected in a series of semi-structured interviews (Robson 2002), which took place between November 2009 and March 2010. The interviews were both preceded and followed up by a long process of industrial cooperation, including discussions, meetings and workshops. The interviews were performed by Jeff Winter, hereafter referred to as JWI, and Mikko Rissanen, referred to as MRI.

6.2.1 The Industrial Organization

The industrial organization, ABB, is a vast group of individual technology companies that operate in almost all of the sub-domains within the power and automation markets, for instance robotics products and supervisory control systems for production plants of all sorts. MRI belongs to ABB Corporate Research which is a research organization within the ABB Group, the umbrella organization. ABB Corporate Research is thus a unit that provides short and long term research support to individual ABB companies (here referred to also as business units) that own the products and businesses, such as the target of this case study. Corporate Research itself does not own products or develop them directly. Rather, it investigates new product and business ideas, technologies, methodologies and acts as the link to academia, with the aim of ensuring the whole company’s competitiveness in global markets. Scientists and technology specialists at Corporate Research work closely with the business units in collaboration projects and bind the latest knowledge from universities to ABB-specific demands as needed.

6.2.2 Meetings and Interviews

The project began with an initial meeting at Blekinge Institute of Technology (BTH). Representatives of ABB’s Corporate Research unit contacted Kari Rönkkö and subsequently visited the Use-Oriented Design and Development (U-ODD) research group (U-ODD 2011) headed by Rönkkö at BTH. The purpose of the meeting was to discuss U-
ODD’s research methods and results, and more specifically to discuss their previous work in the area of usability and usability testing (see e.g. Rönkkö, Winter et al. 2009). This first meeting led to a general agreement to find areas for future research cooperation. MRI was our main contact at ABB Corporate Research and specifically its software research programme Industrial Software Systems. His background within the company, first as R&D scientist and later as principal scientist and coordinator for usability and human-computer interaction research, meant that he had extensive experience of performing research within the company, targeting various business units, systems and products by investigating various software technologies, product development methodologies and organizational factors, often in collaboration with universities (see e.g. Aleksy, Rissanen et al. 2011; Andersson, Bligård et al. 2011; Azhar and Rissanen 2011; Björndal, Rissanen et al. 2011; Kume and Rissanen 2011). Thus, MRI acted as an important source of knowledge about the company and its development and history.

By September 2009, a specific plan had been formulated, and JWI, together with Rönkkö, travelled to the industrial site for discussions about the background to a potential project, the goals and expectations of the industrial and academic partners, and practical matters about performing a study. This resulted in an agreement to cooperate in performing a case study that would examine problems and possibilities of developing usability Key Performance Indicators (KPIs). In November 2009, plans for the project were finalised, a unit within the company was selected for study, and JWI travelled to the industrial site to prepare for the coming interviews.

An initial meeting took place to discuss general conditions and expectations. JWI was introduced to the industrial site, and he presented some of the results of his previous research to members of the Corporate Research unit, and other representatives of the company. He was given access to a workplace at the site, and had access to company documents to help him understand the structure and operations of the company in general, and the industrial unit where the case study was to take place. After two days on-site, the process of interviewing began. The interviews were performed in accordance with a jointly written research plan, initially formulated by JWI and then validated and adjusted together with MRI. The plan specified how the interviews should take place and what they should focus upon.

Ten interviews were performed, at two different sites, on five different occasions. The interviews are listed below, with details of when and on which site the interview took place, and who participated. To maintain anonymity, the names are changed, and the sites for the interviews are simply listed as Site 1 and Site 2. For nine of the interviews (all except the second interview) at least one of the researchers was on-site together with the interviewee/s. Both of the researchers were on-site for the first interview, which allowed them to observe the interview process at first hand, and decide how future interviews should be structured and performed. In interviews three to ten, where only one researcher was on site, the second researcher participated via telephone. The call was made via a
speaker phone, so that all of the participants could hear what was said and could talk to one another. Interview 2 differed, in that, whilst both of the researchers were at one site, the interviewee was at a different site. This interview was performed via a speaker phone, with a research manager at the parent company, who was responsible for a project dealing with usability KPIs.

The interviews took place as follows:

1 & 2. 11th Nov. 2009, at Site 1. Present on-site: JWI, MRI. Sandra (System owner), Martin (Project manager and Principal scientist). Telephone: Michael (Research program manager, Switzerland).

3, 4 & 5. 4th Dec. 2009, at Site 2. Telephone: MRI. Present on-site: JWI. Sam (Senior Specialist/Architect), Harry (Program manager), Carl (Senior Consultant):

6 & 7. 7th Jan. 2010, at Site 1. Telephone JWI. Present on-site: MRI. Ben (Product Manager), George (Support Manager), James (Support Engineer, Training Coordinator).


10. 29th Mar. 2010, at Site 1. Telephone JWI. Present on-site: MRI. Mike (Vice President, Marketing and Sales), Andrew (Sales and support), Lars (Sales and support).

6.2.3 Interview Procedures

With the exception of interview two, all of the interviews followed the same general structure. As previously mentioned, the study was performed as a case study. In case studies, data are often non-numerical, so they are often referred to as qualitative research. However, they can be based on any mix of quantitative and qualitative evidence; they can even be limited to quantitative evidence (Yin 2003, p. 14). Case studies can be characterised as “flexible design” research (Robson 2002), where flexibility means that the research design can evolve during data collection. In our research, the processes of interviewing and analysis influenced one another, as they took place in parallel. However, with the exception of interview two, all of the interviews followed the same general structure. Interview two followed a different structure since it served a different purpose: to gain the management’s view regarding the importance of developing measurable usability KPIs. With the exception of interviews 2 and 3, all of the interviews were performed in Swedish, and the quotes given in the following material were translated to English by JWI.

The selection has been partly a convenience sample, thereby including those who have been able and willing to participate. The first participants were chosen by the researchers from the corporate research department, from their knowledge regarding suitable candidates. Furthermore, we have chosen participants based on suggestions from other interviewees, who suggested suitable candidates who could extend our study. As shown in the list of interviews in the previous section, we have tried to gain a broad cross-section of
the organization, from product management to developers, and the roles (as listed above),
range from a Vice President of marketing and sales, to program, product, and support
management, to “shop floor” representatives for development, integration and support.

During the interviews, both researchers were active and could ask complementary
questions, or ask respondents to clarify or develop their answers to certain questions. To
begin, the participants were shown a PowerPoint (Microsoft) presentation, giving the
background to the study and introducing the problems the researchers were studying, and
some of the questions that they hoped to answer. The participants were guaranteed
anonymity, and ensured that it would be impossible to identify them through the case study
material, or publications related to it. It was also explained that nothing would be published
without the permission of both the participants in the study, and the company.

When interviewing proceeded, the researchers began by explaining who they were, and
which roles they held in the study, and gave some more general information about the
purpose and structure of the research. They asked for permission to audio record the
interview, and explained that the material would be transcribed verbatim, to be used in the
analysis process. The participants were assured that the study material would be treated
according to ethical guidelines: that the recordings and transcripts would remain in the
possession of the researchers, and would be treated in a way that would preserve the
anonymity of the participants.

Seven of the interviews were audio recorded, whilst three were not. All of the recordings
took place at site 1, and the interviews were then fully transcribed. At site 2, the first
interview took place with a member of staff at management level, and he suggested that the
interviews not be recorded that day. The explanation given for this was that the researchers
could choose between recording, which would probably mean that they were likely to give
the official on-the-record version of things, and not recording, which meant that they could
give a more informal and impromptu version of the situation. Given this, it was decided not
to record the interviews that day, but extensive notes were taken during the interviews.
These notes were then written in the case study logbook the same evening, whilst the
material was still fresh, and the material was also commented in the logbook at the same
time.

The general structure in the research plan was used in all of the interviews except interview
two, which had a different purpose. However, although the general structure and focus of
the interviews remained the same throughout the process, the analysis process meant that
the researchers became sensitised to different issues, and these insights influenced the
questions that were asked, the way they were asked, and the data that was collected.

Researchers kept in touch during the interview period. They did this via telephone and e-
mail, in order to discuss ideas, findings and preliminary results. After the ninth interview,
they held a longer telephone meeting, to talk about preliminary findings, and to discuss
how the continued work of the analysis of the case study material should be done. This
meeting was transcribed and used in the analysis process. For further details of the analysis process, see the following section.

Results were presented when Rönkkö and JWI visited the industrial site on 2nd June 2010. At this meeting, the initial results of the case study were presented to and discussed with the main partners in the research project, other members of the Corporate Research group, and other stakeholders within the company. This meeting resulted in further input to the analysis process, and plans for continued cooperation. As a concrete result of our cooperation, we were co-organisers, together with Sanjay Tripathi, ABB Corporate Research, India, of the first Workshop in Industrial User Experience (WIndUX 2011) at the India HCI 2011 conference (WIndUX 2011).

6.2.4 Analysis Procedures

The case study material was recorded in a research logbook containing transcriptions and notes taken when interviewing, which was the main part of the case study database. This collected all of the information in the study, allowing for traceability and transparency of the material, and reliability (Yin 2003). The logbook contained marginal notations of thoughts and ideas concerning themes and concepts that arose when reading or writing the case study data.

In a study of this kind, where the methods that generate qualitative data are a substantial part of the study, serious attention must be given to how the data are analysed (Robson 2002, p. 456). In our study, the field studies have been informed by the underpinning assumption in ethnographic approach (Rönkkö 2010, Section 2) and analysis has built upon techniques taken from grounded theory. Grounded theory (GT) is both a strategy for performing research and a style of analysing the data that arises from the research (Robson 2002, p. 191). GT studies empirical events and experiences, and a GT researcher follows up hunches and potential analytic ideas about them (Charmaz 2010). The iterative approach found in grounded theory where data analysis fuels data collection, which then leads to further analysis, is a feature of the analysis of most qualitative studies (Robson 2002, p. 487), and many analyses of flexible design studies have been influenced by grounded theory.

The material has been analysed in a fashion inspired by constructivist GT. Constructivist GT differs from traditional GT as constructivists view data as being constructed rather than being discovered, and view their analyses as being interpretive, rather than objective reports (Morse 2009). Objectivist GT assumes the discovery of data, that data are separate facts from the observer, and that the data should be viewed without preconception. Constructivist GT, however, assumes multiple realities, and multiple perspectives, which means that the data are not separate from the viewer or the viewed, but are constructed through interaction (Morse 2009). Charmaz contrasts the Constructivist GT approach with the way that Glaser and Strauss (Glaser and Strauss 1967) talk about discovering theory. In their approach there is an assumption that the theory emerges from data separate from the observer, whereas Charmaz assumes that neither data nor theories are discovered. Instead,
since we are all a part of the world we study and the data we collect, our theories are constructed through our past and present interactions with people, practices and perspectives (Charmaz 2010). According to Charmaz, Constructivist GT is profoundly interactive, in both data collection and analytic processes. It adopts the inductive, comparative and open-ended approaches of earlier GT approaches, but also includes the abductive approach when checking and refining the developments of categories. Abductive reasoning arises from experience, leads to creative but logical inferences, and creates hypotheses for testing these inferences, to arrive at plausible theoretical explanations. Thus, when a GT researcher encounters a surprising finding, he or she should consider all theoretical ideas that could account for it, gather more data to put the ideas to the test, and adopt the most plausible explanation (Charmaz 2009). The fact that we were inspired by constructivist GT led us to include the areas of related work that we discuss later in this text. It meant that we gathered theories and knowledge from a number of areas that were connected to the situation that we observed, but which otherwise might have remained outside the scope of the study. Thus, our approach has affected the analysis of the situation, the discussion of the situation, and the conclusions that we have reached.

The data in the study was discussed several times together with research colleagues. This meant they were given access to anonymised transcripts of the interviews. We held meetings and discussions after they had read the material and formulated ideas about the meaning of the data. This was done as a form of triangulation, to cross-check the interpretation of the results. The group discussed the data, and the thoughts and findings of both the researchers and their colleagues. The structure of the material was drawn on a whiteboard, to illustrate the contents of the data, and the connections between different elements. These discussions affected the continued work of analysing and collecting the data, the choice of areas of related work to look at more closely, and the interpretation of the case study material.

Thus, the processes of data collection and analysis have been iterative. At each step in the analysis process, the earlier material has been read and re-read, to see if support can be found for the new ideas and theories. At the end of the process, a theory has been formulated that is grounded in the empirical material.

We have worked in an interpretive research process, and must therefore have other notions of rigour that reflect the same sorts of quality notions that are used in positivist research, which discusses results on the basis of objectivity, reliability, internal and external validity. We have performed our study so that it complies with the quality standards as defined by Gasson (2003), which are Confirmability, Dependability/Auditability, Internal consistency, and Transferability.

The main issue concerning confirmability is that findings should represent the situation being researched, rather than the beliefs of the researcher, and interpretive GT research focuses on reflexive self-awareness, in an attempt to guard against influences and prejudices. Subjectivity is dealt with through a process of reflexivity, where we retain
awareness of ourselves as part of a social context, and the fact that we apply biases and prejudices to the collection, analysis, and interpretation of data. This is in line with the ethnographic approach (Rönkkö 2010), with its focus on “members’ point of view” that has influenced our research approach. Thus, we affect the course of the study, and the choice of the material that we see as being related to our case, but this is always borne in mind during the research process, whilst attempting to understand reality as it is experienced by the participants.

The main issue regarding Dependability/auditability is achieving consistency in the way in which a study is conducted, and it is important to show the use of clear and repeatable procedures. The procedures whereby data are collected and analysed must be recorded so that others can understand them. A research journal must be kept, where records are kept of all data collection and analysis, and we have done this through maintaining a log book where the study is recorded in detail, from the initial stages of the research process, through the interviews and the analysis of the material to the collection of related work that has informed the analysis process.

To achieve internal consistency, it is necessary to explain from which data and how the theoretical results are derived, and whose perspectives these constructs reflect. The source data and the process whereby it is collected should be described in enough detail to demonstrate a fit between the theory and the data. In this paper, we do this by discussing the research procedures that were followed, by showing the instances of issues that were described in the interviews, showing the connection to theories within the area, and then relating this to the results that we present. This is in line with Charmaz (2010) description of theory, which states that Constructivist GT is an abstract understanding of the case at hand. It can start with sensitizing concepts that address power, global reach and difference, to finally arrive at a theory, based on inductive analysis, which shows connections between local worlds and larger social structure.

The core issue of transferability is to which extent a researcher can claim general application of a theory, and this depends on similarities or differences in the context where the theory is to be applied. We discuss the context of our case in our conclusions, and there we discuss the scope of our theory.

The empirical material and the related work that led to the formulation of the theory are discussed in the following sections.

6.3 Results

The long term focus of the business unit is on globalisation, efficiency and quality. According to one of the product managers, his focus was on maximising the number of users of his product, and to differentiate their product from their competitors’. Usability was seen as being a part of that focus. However, views of what usability actually is in this context, and its importance, were found to be varied.
Therefore, the primary intention of the study at hand was to examine the role of usability in the organization, and was part of a management drive to find ways of measuring usability on the basis of KPIs. However, the case study material showed that there are many factors in the situation that we studied that make problems of the development of the systems that were discussed into what we see as a wicked problem. This is partly on the level of how to work with usability questions, but is also apparent on many organizational levels.

Figure 6.1 is an illustration of the situation that became apparent during the analysis of the interview material. The “warning signs” show areas where problems became apparent in the case study material. The “obstacles” illustrate breakdowns in communication or other obstacles that lead to problems. These problems, obstacles and breakdowns are discussed in greater detail in the following section.

As can be seen, there are several types of stakeholders involved, including several layers of company management, developers, support staff, end-users and integrators. This section looks at all of the elements in Figure 6.1, and provides examples from the interviews to show the background to this description. We separate the illustration into its component elements, and provide illustrations for all parts of the larger picture.

6.3.1 Integrators Create Solutions Out of Reach of ABB

This description is based on material from the interviews, in combination with data from previous studies performed within the company. Integrators can be external to the company, or in-house. They specialise in tailoring robots and robot systems for end-user
organizations and end-users. They have a number of goals, both internal and external. They are responsible for obtaining contracts, and ensuring that the robots are so well adapted to the customers’ operations that they receive no support calls afterwards. Their primary tasks are installing a robot cell at a customer site, and gaining a formal acceptance from the customer. Their intention is to satisfy customer needs, thereby making customers happy, and building a good reputation for the company. The main physical output of the integrators’ work is a functional robot cell, which runs first at the integrator’s workshop, and finally at the customer site.

![Figure 6.2 The role of the integrators](image)

Integrators begin by gathering requirements, which are used to write a technical specification draft. Then they buy the necessary components, such as conveyor belts or smart cameras, and buy a robot from a manufacturer. They program the robot, and make a customised robot cell, which can be tested with representatives of the end-customer. A manual is written, the robot cell is finalised at the customer site, and the robot is acceptance tested. Finally, end-users are trained, and the robot is fine-tuned and optimised at the customer site. The integrators handle any support calls that may follow.

A number of problems exist in the integrators’ work. Firstly, the initial requirements that are produced can be inaccurate. This may be because the customers have limited knowledge of robots, but also because the environment at the customer-site is unspecified, and may even contain limitations that can stop the deal. An optimal solution is hard to prove and produce, before actual production. In many cases, micro-level customisation is needed to fit the robot to the customer’s environment. Customer visits requiring travel are time consuming and expensive.

In the interviews, it became apparent that integrators can make far-reaching changes to the interface that ABB has no control over. In interview 1, Sandra said that integrators create their own human-machine interface and that “…we create an interface that lets them… short-circuit our interface.” This can mean that the end-users never see the interface that ABB designs. There is a risk that “…they can make unwise decisions that make the product less efficient to use…” and therefore, it is important that the integrators really understand the system. This can lead to problems for ABB, and in interview 9, Charles said that “…the end-user sees that it says ABB on the robot, and if we have an integrator who doesn’t do his job, it is ABB who ends up with the bad-will.” He also said that he had seen situations
where they have lost orders because an integrator had done a poor job, even though it was not the robot itself that was at fault.

![Figure 6.3 Conflicts between levels of the organization](image)

**6.3.2 Conflicts of Interest Between Departments**

We can see how differing interests can create obstacles between management and other levels in the organization. For example, product management may be reluctant to implement some directives that come from higher management. This is often for financial reasons. An example of this was that product managers could be reluctant to allocate funds to a problem that had been prioritised at a higher level in the organization, when they saw more pressing problems to deal with, and where the budget steers the work that can be performed. In interview 7, James said “...if I am going to be a project leader, and Research or Head Office come and say that I have to pay lots of money for some external resource to come and say what I should do [...] then I am going to say right away, no way! I would rather put the money into something of my own.”

This type of problem was mentioned again in interview 7, where James said “It’s a question of budget. [...] the product division and the service division don’t have the same bosses… and that’s something that you find throughout the company, and that is more or less apparent depending on how well managers communicate.” This leads to a situation where e.g. the product department might create functions that cause extra work for the service department, but where the service department has no extra resources to deal with the problems. In this situation to ensure that service is satisfactory means “…there must be different models for allocating resources.”

**6.3.3 Usability Problems Filtered Out**

Different stakeholders have their own ideas of what the root of a problem is, and what the solution must be. Their choice of problem formulation, and problem solution, depends on their own viewpoint. In interview 1, Sandra said that “There’s never any rigour, any stringency, and it’s very person-dependent, and we have the division in X-town, [...] and they have a bit different culture to what we have here.” These differences affect the formulation of the problem space and the solutions offered.
As a consequence of this, when the development organization plans and performs their work, it is often the case that usability problems are left unaddressed, even though they often appear to be relatively simple to address. One reason for this is the fact that issues that are reported, via the support organization, or through other channels, are prioritised by a Change Control Board (CCB). Given the number of issues that must be dealt with, usability issues often end up towards the bottom of the list of prioritised items. Other issues are seen as being more important for the use of the product. Development resources are scarce and must be channelled into the areas that are most likely to give a return on investment. The resources that are available, usually in the form of man-hours, set the bounds for what can be done, and usability issues often fall outside of these bounds. One of the reasons for this is that they are often seen as “added value”, and not a core feature, without which the product cannot be sold.

In interview 7, James said that they take up requests and issues that have been collected by the support organization, but that discussions in the CCB often get to be “…very hard and very technical, or whatever you would say, this is the problem, how can we solve the problem, OK… next problem, how can we fix that, and so on. When that happens, it is easy to lose sight of the wider perspective.” James also said that “Small changes never get made. Things have to have a certain dignity before they are dealt with. If you want to change some small thing that would increase usability, it never gets done. It is so small that it is rejected right off.” This is seen as a result of the fact that “…the process is so gigantic from the case where someone has a problem till there is a solution in place.” Things don’t get done because “It isn’t worth the process… things disappear into a void.”
Chapter Six
Identifying organisational barriers

Figure 6.5 Usability problems are left un-dealt with

A further complication is that using feedback from the shop floor is hard to manage. Sam said in interview 3 that “There is a matrix of users, from those who know what they want and know how to say it, to those who don’t know what to say and don’t know how to say it. The dangerous ones are those who don’t know what they want, but do know how to say it, since they can influence things extremely negatively. The ones who know what they want but don’t know how to say it are also difficult to deal with.” He said that observation would give much more than collecting user feedback.

There is also a reluctance to let end-users influence the way the product is developed. In interview 4, Harry, a program manager and member of a Change Control Board, said that meeting individual customers was not desirable, since that could give the customer the impression that they would develop solutions for their particular problems. This was not seen to be possible, since many of the problems are too specific, and not general enough. Fine tuning products to particular needs is not feasible, since these small changes have no impact when it comes to marketing. The marketing department cannot make a case that extensive changes have been made in a product on the basis of a number of small changes.

Another factor that influences what gets done is the fact that the organization has a strong background in traditional engineering values. This affects the way in which problems are discussed and solved. In interview 3, Sam talked about the fact that the company has until now been mechanical-technology intensive. The developers within the company have been mechanical developers, and it is only recently that they have seen a division between software and mechanical developers. Sam was of the opinion that it was still difficult for developers to work at ABB, particularly when compared to a pure software company, since the company is still influenced by the traditions of mechanical engineering.

Since even small changes in a complex system can potentially lead to far-reaching and unforeseen consequences in the system as a whole, there is a reluctance to deal with small
issues, including usability issues. In interview 3, Sam said that, although some departments were interested in making usability improvements, the general attitude was “...if it is up and working, don’t tamper with it.” In interview 7, James said that there are examples of known faults that create frustration for end-users, and that never get fixed “...it isn’t even like they are hard to fix; it’s just the organization that doesn’t think it is important to solve.” However, there are also valid reasons for not making small changes, because “...everything affects everything else, and that’s why testing is so important, since... it’s actually someone’s life that is in danger if something goes wrong there [...] That’s why, before you make a change, you have thought it through, and planned it and worked on it and so on, so you don’t dare make small changes.”

In interview 3, Sam said that “…design takes place up-front, and they try to see ‘is this going to be good and easy to use?’ The concepts have to be established early on, as it is difficult and expensive to change things later on in the process, especially when the system is already launched.” He used a simile that he returned to several times during the interview. It is like building a house; if it is really bad, the people don’t move in, or move out soon after they have moved in. Some people live in the house a while and then want extensions or additions. Depending on the way that you have built the house, it can be difficult or even impossible to alter the house according to the wishes of the tenants. This is comparable to developing the system, where the architecture should be designed to allow alterations or additions as long as possible without having to demolish the house and rebuild from the beginning. The implication of this is that many of the functions that exist are given, since they are built into the system from its foundation, but the art of changing the system is to find the changes that mean that the tenants stay in the house.

However, despite the architecture, things are rarely simple to deal with. In interview 8, Oscar said that, when dealing with changes “...usually, it isn’t a case of just doing something in our unit, in the graphical interface, it is most often more holistic, that you have to do something in the steering system too. As a rule, lots of different partners are involved.” In interview 5, Carl, a developer, also said that it was problematic to fulfil a user’s wishes without analysing whether it is a good idea to fix things, and as a developer, it was difficult to see the whole picture. He said that he was “…out on one of the branches fixing things; I don’t have time to see things holistically.” He was of the opinion that the simplest solution is to fix something out at the far end of a branch. You don’t want to take too great a risk and end up doing something that no-one has asked for.

Since there is no holistic view of the problem or the consequences of choices, then choices made affect the product in unforeseeable ways. This means that when issues are “filtered”, usability problems often fall through the filter, and disappear.

6.3.4 Lack of Muscle

There is an understanding of the fact that even small issues are important, and actors within the support department expressed the desire to have a stronger influence regarding what gets to be done when dealing with issues, even small issues. Through their contact with
end-users and integrators, they gain knowledge of problems that are experienced in the field, and in some cases, they would like to ensure that the issues that they have knowledge of are taken care of. Issues which may be regarded as small issues in the eyes of some stakeholders are understood by the support organization as being important. Many customers report these and complain about them. In this light, even a small GUI bug fix that would only take a few hours to resolve would make a big difference to the customers as seen through eyes of the Support department. However, they describe a situation where they do not have sufficient power within the organization to ensure that these issues are dealt with.

![Figure 6.6 Support cannot influence what gets done](image)

The support organization has limited means of influencing the decisions taken within the CCB, and they express a desire to have their own budget that could be used to “pay” the development organization to take care of the issues that they see as being important. In interview 7, James said “We suggest changes every day, which we have experienced, in meetings with program managers, but we don’t have the… it is a question of prioritising things, with other functions and requirements.” He continued by saying “A while ago, we discussed earmarking a certain number of hours in every project for support… that support had a certain number of hours to use in development projects, so that support could say that this case is so important that I don’t care about your priority lists, I want this done at once.”

In interview 1, Sandra talked about previous experience of working with usability issues and said that her role was to “…work with these things from the point of view of a purchaser. And be able to influence things, because it often turned out that […] you come back with an evaluation of a beta with, you know, shelf metres of ideas for improvements, but it isn’t always those that… (get done) […] but if you are on the other side, as a purchaser, you could be able to, like, order beforehand things that would get left out otherwise.” However, this does not seem to have left any lasting changes in the way usability is worked with.

James said that although the support organization does have a voice, it is hard to make that voice heard “…we don’t own any products, we don’t have a pot of money to affect the product, and it’s a case of selling in to the product manager, or program manager, good
ideas… It’s up to the individuals who are at the table, to sell in our ideas, and get them priority, new functions, above all.”

6.3.5 Lack of Two-Way Communication with the User Community
There is a lively user community/forum on the Internet, where end-users of the product can discuss problems and their solutions. This is another area where a lack of resources makes itself felt. Previously, the support department has been active within the support forum, but is no longer able to keep up this function. They saw this as being an important way to be visible within the user community and to gain knowledge of user problems. The knowledge gained could be used to suggest solutions for end-users, and to find ideas for improving the functionality of the product. However, they now find that they do not have enough resources to be active within the forum/community, and see how they miss valuable opportunities for product development and improvement.

![Figure 6.7 The support organization cannot reach the user community](image)

The web forum had been running for years and customers were active in presenting their problems and improvement ideas, but James seemed to feel under-resourced in comparison to the potential. In interview 7, he said that “…we have our web pages and forums and suchlike… but the processes around them, and who is responsible for it all, we aren’t so mature in how we deal with it all”. On the subject of user feedback on software issues, James said that “…we just don’t have the resources to take care of everything. Or to keep it going the way we would like to.” He said that this was also under-developed on the hardware side, that there was a lack of organizational maturity, and that they would find it difficult to deal with the extra work that getting the feedback would involve. Nevertheless, getting remote feedback would allow them to “…look at their scenarios, and see what they are talking about, and get a picture of how the land lies. And it would be much more cost efficient.” If this was possible, it would allow them to spend the time actually getting things done, and would allow them to choose “key partners, the ones who use our PC products or robots to their limits, and we could have workshops with them, where they could tell us
about their troubles... it would be very rewarding. But still, setting up something like that means that you have to have the resources available.”

Efforts have been made to collect data automatically, by building a support tool in the software, so that when a problem made itself known, the support tool would report it to support, and transmit the data from the system, so that the customer did not need to do so. However, they found that “…it demanded too many resources – from the development organization and from us, to deal with certain types of data and so on. So even if we built it in, there was no organization to receive these things. There was no organizational maturity. When we came with all these cool files... there was no-one who could take care of it or deal with it.”

In interview 6, Ben said that, since the product is in the whole lifecycle of a production system “…my problem is that I have a large number of different types of users.” It is difficult to communicate with end-users, since although there is a register of people who have purchased their product, they do not know who the end-users are, since the customer is not necessarily a user. “When you purchase the product, we register an e-post address, but that could be the user, it might be the user’s boss, it might be a purchaser, it might be an integrator […] it could be anyone.” A solution for this could be to collect information via Internet. However “…it’s tough […], we have like, a... we have a tendency to have a one-way web, and a web that moves forwards… rather slowly.” There are a number of tools that are available for use, but they do not contain the tools that they need, and sometimes they find that “you have to like, go head-to-head with... with the web police, as I call them.” The one-way web means that they can’t reach their customers. In interview 2, Sam also mentioned cases where the web policies in operation in the company meant that it was harder to reach customers with information.

Figure 6.8 Communication problems

The interviews showed that in general there are problems of communication between the support organization and the end-users. There is a desire for close two-way communication between support and end-users, but there are organizational obstacles that hinder this communication. James said that, in his support role, he has to filter out the information that he receives from end-users, but that “…in a modern age, end-users want to be able to
submit their opinions, or points of view, or wishes or whatever, straight to developers”. He only has a small contact area with users, and said that “…end-users can’t necessarily be bothered getting in touch with me, so that I can then get in touch with the product departments”. Also, he said that even if they could include the community, there must still be people who can “…give help and answer questions, filter things, and send them to the right person, and so on. And at the top level, they don’t want to allocate these resources… There aren’t enough of us to manage this kind of quick response.” He said that “…from the beginning, our goal was that we should be active in the forums, now we are more… we have the role of watchers.

There are many ideas for improving the communication, including allowing support to communicate with end-users directly via the product itself. This is already technically possible in many ways, as the product can already collect and send data to the organization, but it appears to be the case that there is a lack of resources that would allow these solutions to be developed. In the situation that exists today, there are no routines for analysing the data that can be collected, few possibilities to develop new ways of collecting and using the information, or for developing the processes for communication, which are currently regarded as immature.

6.3.6 Lack of User Knowledge

The primary stakeholders who are external to the organization are end-users and purchasers of robots and robot systems. These stakeholders differ greatly, both in their knowledge and experience of robots, and in their requirements. They include actors within the automotive industry, and other “heavy” industries, where there is a long tradition of the use of robots and robot systems, but they can also include actors with little experience or knowledge of robots. These inexperienced actors are becoming more important in the marketplace, as the company attempts to expand the market for their products, by finding new customer groups. In interview 10, Mike said “…what we see as being the future is all of those small customers, rather than the traditional approach to the car industry, the volume purchasers.” However, in interview 10, Lars said that “…we are very technology driven, and actually have very complicated systems, and you can do a lot more with our systems that with lots of our competitors. But that means that they are complicated, too. The more possibilities you have, the more you have to learn to understand the system.” They mentioned the fact that small customers do not necessarily see the cost benefit or appreciate the possibilities “… they cost money. And they… maybe they don’t see the benefits of them when they are purchasing. They would earn from it in the long run, but when they are buying, they don’t see it.” The same theme arose in interview 9, where Charles talked about customers who have a low competence level of software in general and robots in particular. “…they start from scratch as far as robots are concerned. […] It’s a big step for them, technically […] these customers are anxious about investing in such complicated equipment. We have competitors who have a more homogenous software offer […] and we often lose orders because customers feel that it’s easier to buy from them, because they have more integrated
solutions.” He said that these customers “…know nothing about robots, and find it difficult to find competent staff… this is a big problem within the engineering industry.”

Obtaining and understanding information from end-users is difficult. In interview 7, when asked how they obtained information about what end-users think about the product, Ben said that “…we have no systematic way of doing it at the moment […] I am going to start with customer satisfaction surveys. […] to periodically go out and gather in information from our users, about what they think about our product, that we can use to improve it.”

This was not just to gather information about usability, but about all sorts of factors. He said that they had done some such surveys, but that “…we have never really managed to do so much with the material that we have collected, maybe not even really understood it…”

In interview 8, Oscar expressed the fact that, when they are developing the product “…it is important to remember, that we hardly ever use it the way it is used in reality.” In development they test on small systems, and perform simple actions, whereas customers “…have, like, hundreds of IO signals, and huge robot systems, and such like […] we get indications that things don’t really work… when you use it with loads of IO signals, for example.”

There is a high risk when working with usability that the company develops the things seen by the developers as important, rather than what is important for the end-users. In interview 1, Sandra described how they do not know what the end-users really need. When developing a product, they begin to work in the situation that already exists, and this affects what gets developed. It was described as “…we always come along with a huge rucksack of functionality, that has to be led over to the next product, and in the end […] maybe we are just fine-tuning, polishing the interface and the logic of some function endlessly […] and testing the interface, and using someone who is fantastic at creating interfaces, but if that functionality isn’t what gets used…” In this way, they can end up developing the wrong things, “polishing the interface” because they think that they are developing what the end-users want, but if it then isn’t what end-users want, then all of the work is to no avail.

In interview 7, James said that it is the end-users who “…are going to have the most opinions. They are going to experience problems that we can’t even imagine, because their
scenarios are unique every time.” In interview 1, Sandra also said that the developers do not understand the end-users, and the end-users do not understand the developers. The end-users do not use the functionalities that the developers are concerned with. She said “it’s important to talk to the customers out in the real world… it’s pretty cool, because I found that the people who use our products don’t understand what I am asking about, the details, and I don’t know what they mean, because they often [...] use our products in a way that I don’t… I don’t understand how they do it. I don’t know our products to 100%, how they are used, but I know some of the finer points, and tricks, that they don’t know.”

This was confirmed again in interview 4, where Harry said that developers and end-users think in different ways. Developers are given a lot of responsibility, and they build in extensive functionality that lets the end-users do things that many of them are not capable of. The end-users have enormous possibilities to adapt the system, but this has greatly increased the complexity of the product. In interview 6, Carl also said that it is easy to develop cool solutions and functions that there is no actual demand for, whilst at the same time it is easy to miss simple things that are actually needed.

In interview 7, James said that “There are probably lots of areas where we assume that the end-users understand things, or work in a particular way, and in lots of cases, perhaps things take much longer, or are much more frustrating than we realize. It isn’t certain that we have the right idea of how the end-users experience the products.” When they look at the support errands that are submitted they find that “… we get lots of things, but I wouldn’t say that we understand everything about how the customers experience the product”.

According to Sandra in interview 1, “If we are going to measure what is important […] then we must know it […] that this is what is most important, not for us, but for them. The ones who use it. If we can’t formulate it, why should we do something? […] We are a large organization, and can be, like, too far from – and I won’t say the end-user, but even from the user.”

Thus, we see that there is a lack of knowledge of who the end-users are, how they work, and what their needs are. This lack of knowledge means that the solutions that are developed today are not adapted to the actual needs of the end-users. To deal with this is important for the end-users of today, and is even more important if the company is going to adapt to the demands made by tomorrow’s end-users.

6.3.7 Summary

The situation that we have detailed shows some of the organizational barriers that make it difficult for the company to reach the goal of formulating and introducing measurable usability KPIs. Here follows a brief summary of the main points for each specific area above, to aid the reader in gaining an overview.

In section 3.1 it is visible that 1) Integrators can make far-reaching and unexpected changes to the interface designed and produced by ABB. The integrators themselves experience
problems producing satisfactory solutions, due to a high degree of uncertainty regarding what is most appropriate for the customer.

In section 3.2 it is clear that 2) Divergent interests on different organizational levels create conflicts regarding the use of scarce resources, and departments can create work for another department that has no resources to deal with it.

In section 3.3 it is demonstrated that a consistent view of the root causes of problems experienced by end-users, or their solutions, is lacking. The lack of consistent view has the effects that 3) Resource limitations force a prioritisation of tasks, and the fact that usability issues are often regarded as minor problems results in their being left unaddressed, as other issues are seen as more important and pressing, and likely to lead to revenue. Furthermore, 4) the product is so complex that small changes have potentially unforeseen effects on other parts of the product, so the consensus is often to change as little as possible. These decisions are often based on judgements made from an engineering perspective that prioritises functionality rather than usability. In fact 5) there is an expressed reluctance to allow users to influence the product, since user feedback is hard to manage, and 6) there is low understanding of what users actually need, or the importance of taking this into account. This means that many usability issues are filtered out and disappear.

In section 3.4 it is visible that there are some stakeholders, for example in the support department, who understand that even “small” issues are important for the end-users. However, 7) the actors who understand end-users are in the position that they do not have the power in the organization to push the changes that they see as important.

In section 3.5 it is visible that many of these issues are discovered via contact with the end-user community, via the channels that exist and are active today. But even here, 8) there are both economic and systematic barriers that make it difficult to maintain the level of contact that is desirable or to even to deal with the information that is collected today.

In section 3.6 it is demonstrated 9) there is a wide range of end-users, and there is a lack of knowledge of who the end-users actually are. They range from the traditional users with long experience of using complex robot systems, to newer users with little experience or knowledge. These newer customers are seen as important for the future of the product, but the complexity of the product can make things difficult for these users, and even motivate them to choose simpler products. As already mentioned, it is difficult to gain information about the end-users’ opinions, and 10) there is insufficient knowledge of how the product is actually used in the end-users’ context. The product is used in ways that the developers cannot begin to imagine. Users and developers do not understand one another, and think in different ways, leading to a situation where “cool” functionality can be developed on the basis of incomplete knowledge, whilst important things can be missed.

To summarise, here is a list of identified obstacles hindering introduction of KPIs for usability:
1. Integrators can make far-reaching and unexpected changes to the interface designed and produced by ABB

2. Divergent interests on different organizational levels create conflicts regarding the use of scarce resources, and departments can create work for another department that has no resources to deal with it

3. Resource limitations force a prioritisation of tasks, and the fact that usability issues are often regarded as minor problems results in their being left unaddressed

4. The product is so complex that small changes have potentially unforeseen effects on other parts of the product, so the consensus is often to change as little as possible

5. There is an expressed reluctance to allow users to influence the product, since user feedback is hard to manage

6. There is low understanding of what users actually need, or the importance of taking this into account

7. The actors that understand end-users are in a position that they do not have the power in the organization to push the changes that they see as important

8. There are both economic and systematic barriers that make it difficult to maintain the level of contact that is desirable

9. There is a wide range of end-users, and there is a lack of knowledge of who the end-users actually are

10. There is insufficient knowledge of how the product is actually used in the end-users’ context

The background to the case study was an interest within the company to introduce measurable KPIs for usability, and to find candidates for possible measures. As shown in the above, the interview material illustrates many of the problems that make defining and introducing such KPIs difficult. The situation that we have studied is found to be complex, in its mixture of technical, organizational and social factors, and this complexity is seen to lead to many problems and breakdowns. We find that there are many organizational factors, both internal and external to the company, which lead to the situation that exists today. However, this is not new knowledge. The types of problem that exist at ABB are not new and unique problems, but have been in existence for a long time, and although much research has been performed on investigating and remediating these types of problems we find that they are still in existence. Existing research on organizational and social factors will be presented below, in order to shed light on and place our findings in a wider context.
organizational context. Our findings concern usability efforts, which have not been widely discussed in terms of the organizational and social challenges below, whereby we see the importance of contributing here.

6.4 Discussion of Results
In this section, we look at theories regarding the importance of organization and organizational change in the field of systems and their development, and compare those with the ABB situation. The theories, which are chosen from areas that we were led to through the constructivist GT approach used in the study, are taken from several different academic discourses: This approach is also supported by Fuggetta’s view that it is important to reuse experiences gained in other communities, such as organizational behaviour research (Fuggetta 2000). We also look at the material from a design perspective in order to see what might be required to remedy some of the problems that are apparent in our case. One important aspect here is to understand the nature of the problem that we are confronting. We need to understand how the identified problems are related to contexts and how to approach them, whereby we introduced the concepts of wicked problems, collaboration over coordination, and User Centred Design, to explain and address the situation.

6.4.1 Organizational factors are important
The complex process of developing software was discussed in Fuggetta’s roadmap for software process (Fuggetta 2000). He claimed that developing software is a collective and creative effort, and that the quality of the product is dependent not only on the procedures used to create and deliver it, but also on the people and the organization. Our case exhibits ways in which the institutional processes and cultures within the company shape the way in which the product is designed and developed. In many cases it is the organizational influence, rather than the dynamics and limitations of the technology itself, that decides what gets to be done, or conversely does not get to be done. By showing the structure and complexity of the organization, we show some of the connections between the ways in which the technology is designed and developed. Iivari (2006) discusses the way in which organizational culture affects usability work, and states that the cultural context can reinforce and advocate some aspects of usability work, and deem other aspects unimportant. Iivari shows that many studies suggest that culture plays an important role in affecting user involvement in organizations, and suggests that culture is a factor that is intertwined with organizational change efforts to introduce user involvement into organizations. In 2001, Orlikowski and Barley discussed how important it is to take into account the role of the organization when analysing IT (and by extrapolation other) systems (Orlikowski and Barley 2001). They state that there is a lack of knowledge in this area, and that “IT researchers have yet to ask how institutions influence the design, use, and consequences of technologies, either within or across organizations”. They show how studies have taken place at a “lower” level of abstraction, and have not looked at how
regulative processes, normative systems, and cultural frameworks shape the design and use of technical systems. They claim that understanding and guiding techno-social developments requires knowledge of technological systems, social processes and their interactions. There is a need for research that concerns both technology and the social dynamics of organising; that embraces the importance of understanding the role of the human agency as embedded in institutional contexts, as well as the constraints and affordances of technologies. Our results show that this is still relevant today related to organisations’ efforts to improve usability.

Work published by Curtis et al (1988), regarding the problems of designing large systems, also shows that the types of problem that exist at ABB are not new or unique problems, and neither were the problems new when their work was published. They state that “…these problems have survived for several decades, despite serious efforts at improving software productivity and quality”. They characterise the problems as behavioural, and claim that these problems affect software productivity through their impact on cognitive, social and organizational processes. They maintain that the effect of these behavioural factors on software development is much greater than the effects of tools and methods. They also maintain that, since tools and methods that aid individual activities often fail to provide benefits that justify their impact on the team and organizational factors that affect design processes, it is important to understand how behavioural factors affect the execution of software development tasks. The most salient behavioural problems, in terms of the additional effort or mistakes that were attributed to them, were: the thin spread of application domain knowledge; fluctuating and conflicting requirements; and communication bottlenecks and breakdowns (Curtis, Krasner et al. 1988).

When we compare our results with those of Curtis et al, we see that their description corresponds well to the findings in our case, where a large proportion of the problems are based on insufficient knowledge of the expanding domains where the robots are used, and of the users’ characteristics and needs. This leads to a situation where requirements are hard to specify, and once specified can be found to be in conflict with one another, or impractical to implement because of dependencies within the system, or organizational limitations. We see situations where communication breakdowns occur, and a number of factors that lead to bottlenecks in the design and development processes. The problems that were mentioned in the interviews were behavioural problems rather than problems associated with tools and methods and our case is an illustration of the way in which behavioural factors still make it difficult to design and develop systems.

Uldal-Espersen et al (2008) argue that most studies of usability work disregard the organizational setting where the work is situated. However, they state that real-life organizational settings are important when studying practical usability work, and that projects focusing on usability must be based on genuine interest on the side of the stakeholders, including management, users, IT representatives and usability experts. Thompson et al (2010) describe a situation where, as practitioners, managers of design and
UX teams, they find it difficult to position the work of team members, and that roles are often misunderstood. Other disciplines, such as product management and development see the UX work as unnecessary, or even threatening. They experience a situation where, when deploying UX resources, the corporate culture is unready to accept them, and the role of UX manager becomes more that of a change manager. It was found to be difficult to communicate the UX vision and to engage the other disciplines. Rohn and Wixon (2011) also point to a number of what they call persistent questions, such as the fact that there is a lack of clearly defined challenges and opportunities that UX teams face or will face, to which extent design and user research are accepted and influential parts of company leadership, and whether there is a commonly accepted vocabulary to describe tools and methods, both within the teams and for business and technical leaders. These descriptions are illustrative of a situation that we recognize from this case study, where the usability workers found it difficult to make an impact in the organization, and where the resources available for usability work were often cut back, since there was no clear view of the importance of the usability work within the projects.

Curtis et al claim that developing large systems must be treated as a learning, communication and negotiation process. They say that “…three capabilities that we believe must be supported in a software development environment are knowledge sharing and integration, change facilitation, and broad communication and coordination” (Curtis, Krasner et al. 1988). They found that conflicts among requirements often resulted from differing needs of customers, changing needs of individual customers, changes in underlying technologies or competitors’ products, or from misunderstandings regarding the application domain. They even resulted from internal sources, such as marketing, corporate politics, and product line management. Design teams often negotiated to reduce problems and limit requirements to those that could be implemented within schedule, budget, and time restraints. All of these types of mechanisms, such as conflicting requirements, internal conflicts and breakdowns, and adaptations intended to reduce and limit problems and requirements, were found in our case.

The problems discussed by Curtis et al emerged from processes at one level in a layered behavioural model but affected processes at several levels. They operated through different mechanisms and required different solutions. Their model encourages thinking about projects on multiple levels of analysis. It advocates an extension of the evaluation of software practices from individuals, to teams and projects. The picture that they paint fits in well with our findings of conflicts, breakdowns and barriers originating in different parts of the organization, which then spread to and affect other parts of the organization. These problems cannot be dealt with on an individual level, and to understand and treat the mechanisms that lead to the problems, it is necessary to approach them by taking a more holistic view of the structures and processes involved, than is taken today.
6.4.2 The existence of a Wicked Problem

Curtis et al found that the problems that are experienced in large system development projects were not experienced in smaller well-understood applications with stable and complete specifications, and they characterise these problems as wicked problems. Problems of importance are seen as being ‘wicked.’ (Coyne 2005) The concept of wicked problems originates in the work of Rittel and Webber (Rittel and Webber 1973), who discussed the fact that the professional’s job has traditionally been seen as solving problems that can be regarded as “tame”, meaning that they are definable, understandable and consensual. However, they maintain that a shift has occurred in the types of problems that professionals are called upon to solve. Unlike in the natural sciences, societal problems are not definable and separable, and have no solutions that are findable. They are ill defined and rely on political judgement for solution. The term wicked is used “in a meaning akin to that of ‘malignant’ (in contrast to ‘benign’) or ‘vicious’ (like a circle) or ‘tricky’ (like a leprechaun) or ‘aggressive’ (like a lion, in contrast to the docility of a lamb).” (Rittel and Webber 1973) Wicked problems are only loosely formulated, and there is no ‘stopping rule.’ They can be redefined and resolved in different ways over time, and their formulation depends on the viewpoint of those who present them. There is no ultimate test of a solution to a wicked problem. Solutions can only be tested in a practical context, and the solutions are not easily undone (Coyne 2005).

The concept of wicked problems has been referred to in many cases. Keating et al discuss the ways in which systems engineering approaches have successfully addressed well-bounded problems, and how engineering disciplines have traditionally been directed at solving clearly defined problems, with clear goals, that can be attacked from an established body of theory and knowledge. However, these conditions are no longer the rule, and increasing information richness, and higher degrees of complexity lead to a need to address a situation where complex systems problems are emerging. These environments are characterised by ill-defined and potentially tacit goals that are value-laden and that may be impossible to define explicitly (Keating, Rogers et al. 2003). Denning describes wicked problems as messes, which are seen as being worse than problems (Denning 2007). Problems are difficult situations where there is an expectation that solving some difficulty will end the problem, whereas messes are complex, and there is often no agreement on whether the mess consists of a single problem, or the convergence of many problems (Denning 2007). This description is applicable also in the case of ABB. In the “list of identified obstacles hindering introduction of KPIs for usability” found above (hereafter this will be referred to as the list) increased information richness and higher degrees of complexity relates to point 4, 5, 6, 9 and 10; “Ill-defined and potentially tacit goals that are value laden” relates to 1, 6, 9 and 10; “expectation that solving some difficulty will end the problem, whereas messes are complex” relates to 2 and 3. Some of the signs of a mess are the existence of a disordered condition, the absence of an obvious problem statement, a lack of apparent causal relationships, a feeling of confusion or of being stuck, discord and controversy (Denning 2007).
According to Denning & Yaholkovsky (Denning and Yaholkovsky 2008), messes are often found to be non-technical in origin. This is certainly so in our case, where it is not the technology in itself that leads to the problems that we have detailed, but the behavioural problems that we have discussed above. They maintain that messes are best addressed through collaboration. Collaboration is a stronger form of working together than coordination (meaning a system of regulated interaction for people and objects to fulfil their goals), or cooperation (which means playing according to a set of behavioural rules). In collaboration processes, solutions or strategies are created through a process of group interaction. The complexity of collaborative design arises from a need to manage large amounts of information, and understand design decisions that have determined an artefact’s long term evolution (Fischer 2001). Collaborative design leads to work-products that are enriched by multiple perspectives of the participants and the discourses that result from the collaboration. In our case, the factors that lead to the existence of the wicked problem lead to a situation where collaboration, as it is defined above, becomes problematic. The mechanisms of coordination obviously exist, and the normal design and development processes work in a way that allows the product to be successfully developed and marketed. However, the breakdowns and barriers that exist make collaboration difficult. This is supported by the work of Easterbrook (Easterbrook 1995), who maintains that collaboration is difficult. Beyond the work necessary to make progress on the task itself, work is also necessary to maintain relationships between partners and negotiate the nature of the task. Two key factors are to what extent collaborators need to develop and maintain a shared understanding, and how collaborators deal can with conflict (Easterbrook 1995). Given the situation that we have shown to exist at ABB, where there are a number of non-technical breakdowns and obstacles to cross (see points 1, 2, 3, 5, 6, 7, 8, 9 and 10 in the list), there are obviously many barriers to introducing a collaborative way of working.

The work of Arias et al (Arias, Eden et al. 2000) also deals with breakdowns and wicked problems. They discuss them mostly in terms of symmetries of knowledge, rather than the general behavioural problems that we have shown to exist in this case, but there are many parallels to our case and their results can be used to draw interesting parallels. They claimed that the major changes that have occurred in HCI have been at the levels of design, systems, technology and media, where actions and changes can take months or even years or decades. When working at these levels, they state that the relevant theories would probably be grounded in social and organizational themes. This is in accordance with the discussion above, and is clearly relevant to our study, where we are dealing with a technically complex product that has developed and evolved over a long period, and where we have seen that many of the problems are based in organizational factors, and in communication breakdowns between individuals and departments. They claim that these types of breakdowns are important to take into account as they see that the main activity in designing complex systems is that participants instruct one another (Arias, Eden et al. 2000). The view of the domain expert as the sole source of design knowledge does not recognise the fact that all stakeholders can contribute. Knowledge is distributed between
stakeholders, and complex design problems, such as the ones in the situation that we have studied, require more knowledge that any single person can have. Therefore, different stakeholders are required to participate, communicate and collaborate. However, as we have already discussed above, factors in our case make the processes of collaboration difficult, and even communication and participation, particularly on the part of end-users, is hard to bring about. Bringing together different points of view is important, since by doing this, it is possible to create a shared understanding that can lead to new insights, ideas and artefacts. However, communication breakdowns often occur, since different stakeholders use different norms, symbols and representations (Arias, Eden et al. 2000). Applying this knowledge to our case would provide some explanations for the situation that exists today, and would suggest the importance of creating ways for stakeholders to communicate more effectively, in particular with end-users.

However, they also maintain that design problems are wicked, and the design context is characterised by change, conflict and multiple stakeholders. Coyne states that “Wickedness is the norm. It is tame formulations of professional analysis that stand out as a deviation” (Coyne 2005). Problems of designing for complex situations are well known, and have been in existence for a long time. In the field of Computer Supported Cooperative Work (CSCW), for example, Easterbrook (1996) stated that “Software products succeed not because they are designed to suit their users, but because the users can adapt to suit the software”. CSCW design is hard, because there is no framework for understanding how group members adapt to cope with coordination breakdown and conflict. In a case such as ours, where there is considerable complexity, given the complexity of the development process (2, 3, 5, 6, 7, 8 in the list), and the fact that the integrators affect the product offered to the end-users, it is hard to understand how the end-users adapt to the product. It is also hard to know how to make the product adaptable in the most suitable way (1, 9 and 10 in the list). However, we see that even wicked problems are solved, although we cannot find optimal solutions to them, and this is often done in a “satisficing” manner. Satisficing is the process whereby humans, given all of the alternative possibilities, still make decisions and solve problems. Instead of looking for optimal solutions, decision-makers set feasible goals and use decision methods that look for good, or satisfactory solutions, instead of optimal ones as rationality might suggest (Atwood, McCain et al. 2002). Hughes et al (1994) discuss the way in which design is, at best, a ‘satisficing’ activity, often dealing with ‘wicked’ problems, and that this involves being governed by, and influenced by an interplay of political, moral as well as technological considerations.

6.4.3 The importance of context in design

It is often impossible to reach consensus in design processes, where the best that we can hope for is an informed compromise, since there is a symmetry of ignorance, where crucial knowledge for problem resolution exists as tacit knowledge in the heads of individual stakeholders (Arias, Eden et al. 2000). In the context of our case, as long as there is a situation where breakdowns occur and communication is problematic, and where there are
difficulties getting and using input from key stakeholders such as end-users, then it is
difficult to create a situation where it is possible to produce the product that satisfies the
largest number of end-users. Arias et al (Arias, Eden et al. 2000) claim that solving design
problems requires that the owners of the problems are put in charge, thereby promoting
direct interaction that involves people in the decisions that affect them. All stakeholders can
make important contributions and should be turned into designers and co-developers, to
bring their important perspectives to the design process. Being a consumer or designer is
not binary; there is a continuum ranging from passive consumer to designer, and problems
arise when, for example, someone wants to be a designer and is forced to be a consumer. In
our case, this would for example highlight the importance of obtaining and using end-user
feedback, allowing the end-users to become stakeholders who can become more active as
designers. All of these things have been found to be difficult today. In agreement with
Olsson and Jansson (2005), we believe that addressing usability problems in highly
qualified work settings requires involving users to a much greater degree than today, and
much earlier in the development process. They found that types of problems addressed
when developing for a highly qualified work setting were such that they hindered the users
from achieving their goals efficiently in relation to their work activities. The work activities
themselves were often complex, demanding that the users plan and make important
decisions, and collaborate and exchange information with others. They state there are
business goals that drive the organization, and that the goals for computer systems are
heavily influenced by assumptions about how people work and how the organization
functions. Different individuals have different goals, and so the implications for the design
of work and technologies differ, leading to the situation where system goals are not always
the same as the users’ goals. This is reflected in the situation that we found at ABB.

ABB is in a situation where robot systems have become complex systems, and there are
parallels to the work of Keating et al (Keating, Rogers et al. 2003), who, in their work on
System of Systems Engineering (SoSE) discuss the way in which systems that once were
developed as stand-alone systems to address singular problems can no longer be regarded
as isolated. Traditional Systems Engineering has placed context in the background, and
therefore finds it difficult to cope when addressing areas with high levels of ambiguity and
uncertainty, and problems with a high degree of contextual influence. SoSE attempts to
deal with some of these problems, but there are still challenges, including working in a
situation where focus is placed on technical problem solving, which we have found to be in
focus in our study (e.g. 3 and 4 in the list). In reality, it is the contextual, human,
organizational, policy and political dimensions that shape the decision space and solutions
for the technical system problems, and this requires the ability to design and develop across
the traditional boundaries placed by program budgets, operating philosophies and different
objectives. All of these factors that have been found to be problems in our case. Design
cannot be restricted to what have been regarded as stable contexts and isolated systems, and
context must be placed in the foreground, whilst technical solutions are in the background.
Chapter Six
Identifying organisational barriers

Here, we see the importance of context, and this focus is also reflected in the work of Orlikowski and Barley (Orlikowski and Barley 2001), who state that technologies are both social and physical artefacts. Most engineering requirements can be met with multiple designs, so all technologies represent a set of decisions made by specific designers. Accounts of technological change need hybrid explanations that weave human actions and choice with the functions and features of technology and the contexts of technology use, in a way that attends to the micro-dynamics of situated practice.

So far, we have seen the presence of behavioural problems, leading to the existence of a wicked problem, the importance of the roles played by multiple stakeholders, and the importance of context when addressing the problems of designing complex systems. This focus on the importance of context and practice, highlighting the role of end-users as designers, lead us into a discussion of the way in which design processes can be managed.

6.4.4 The Need for User Centred Design

Although they address knowledge management (KM), the work of Fischer and Ostwald (Fischer 2001) is interesting to consider from the point of view of the design and development practices that we have studied at ABB. They discussed KM from a design perspective, where the goal is to enable innovative practice at an organizational level, by supporting collaboration and communication among knowledge workers within and across domains, which we have seen as problematic in our case. They see workers as reflective practitioners who struggle to understand and solve ill-defined problems. They show that users are more concerned with doing their work than with the system itself, but the system must be an adequate tool for their work. Users see whether or not the tool fits their needs. Since other specialists do not use the tool to do their own work, they cannot have this knowledge, and this should be reflected when the system is designed. The needs of workers are unique, since they result from efforts to understand a problem. The context of problem solving dictates the information demands and the context for learning. The need for learning comes from the work, and the learning takes place in the context of the work situation. This is very important to consider in our case, where the interface that is designed must be made adaptable for users, allowing them to tailor it to their own needs, whether they be simple or advanced. Doing this needs knowledge of the context where the tool is to be used, and as we have seen, this knowledge is difficult to gain and to use if it is gained.

Although Fischer and Ostwald discuss in terms of computers, their findings can be relevant in the case of robots, which can more and more be considered to be related to computers and computer systems. Their KM perspective goes beyond individual perspectives and requires stakeholders to learn new relationships between practices and attitudes, which we have found to be difficult in our case (e.g. 1, 6, 7, 9 and 10 in the list). There is a need for an environment that supports social interactions among communities of practice and communities of interest. However, this does not mean that the needs of the individual must be superseded by the interests of the group. Individuality is important, and communities gain their strength from the creativity and engagement of the individual. It is important to
understand and maintain the balance between the individual and the community, and “Collaborative systems will not work in a noncollaborative society” (Fischer 2001). Thereby, they highlight the need for including the individuals in the design processes, and over a period of a few years, this is one of the areas that the HRI community is beginning to become aware of.

We have already seen in our case study that the engineering approach at ABB has been successful at producing robot technology, but that it is no longer found to be adequate to approach the new challenges that they face, where usability is becoming a key factor. One widely spread way of working with usability problems is by engaging usability specialists, and to some degree that has been done at ABB, but even this kind of usability work is lacking in some areas. Friess (2011) highlights the problem that even usability experts have trouble translating issues found in testing. It was shown that many findings from usability testing sessions do not make their way into reports, and those findings that are in usability reports can differ substantially from what was uttered in the testing sessions. Consistency and continuity varied greatly between sessions (Friess 2011). This illuminates a situation where developers cannot be expected to understand and deal with issues that are found, when even usability experts find this to be difficult. This points to the importance of including users at all stages.

One way of approaching this area is through a process of User Centred Design (UCD). Fong describes how communication between humans and machines is usually mediated by an interface, which although it can offer great power and flexibility, comes at the cost of high learning. Other interfaces, for example menu based ones, may be easier for novice users, since they do not make assumptions about what the user knows (Fong, Thorpe et al. 2003). This correlates well to what is found in our study, where experienced end-users can make use of the advanced functionality built in to the interface, whilst the same interface can become an obstacle for novice users, which can lead them to choose a simpler alternative product. This fact highlights the need to take into account the needs and skills of the end-users when designing the product. However, work analysis and user involvement are required to develop common ground, but the UCD principles of early and continuous focus on users, and understanding of users and their tasks, are easily neglected when different development activities compete (Olsson and Jansson 2005), which we observed clearly in our study. According to Gulliksen et al (Gulliksen, Boivie et al. 2006) the design process must be cooperative, dependent on the input of several disciplines and the users. They state that UCD is essential, and that the usability professional is vital in the process, since this role drives the UCD process. This role has the potential to fill the gap between analysis, evaluation and construction, and fill the gap between UCD and software engineering. Usability experts need to be involved continuously in the development process, since they can make the greatest impact when they are involved in projects on a day-to-day basis, including the construction phase when many usability issues need to be dealt with. However, they should not simply become substitutes for user participation.
HCI is becoming more and more interested in the design process, and there are many interpretations of what design is, but good design usually implies a sense of fitting a purpose that can be verified empirically, and the contribution of satisfaction, and similar concepts, will be a topic of increased interest in the future (Karat and Karat 2011). Design is a collaborative process between users, designers and developers, where the expertise of every group is important for the outcome. This approach requires cooperative methods, and the design process is a process for creating knowledge and common ground (Gulliksen, Boivie et al. 2006). Uldall-Espersen also highlight the fact that usability work must involve a multitude of stakeholders, and their study showed how a usability improvement process benefitted from bringing different areas of expertise into play, by applying different techniques, and through close cooperation between people with different types of expertise. The approach taken in their study has important similarities with UCD, where design teams in iterative processes create designs based on studies of real users, tasks and needs, and can respond to changes in the surroundings (Uldall-Espersen, Frøkjær et al. 2008).

Keinonen (2008) states that socially relevant design must be related to the needs of people. UCD of ICT is born from a need to transform complicated technologies into real-world applications. UCD was previously concerned with optimising the design of HCI, but is now concerned with a wider range of interests, aimed at matching interactive technologies with human-human communications, organizational requirements, and social and emotional perspectives. Gulliksen et al (2006) show that attitudes held by the people involved in the process are important, and are crucial for results. If the attitudes of developers, program managers, management and clients are not conducive to UCD and usability, then there is little point in introducing usability and UCD. They state that usability needs a UCD process, and discuss the importance of having a user-centred process, where usability must be tightly integrated. They found that moving from a technically oriented process to a user-centred approach requires more than a few usability activities.

Including users in design and development is important for many reasons, including gaining domain knowledge in complex contexts, and to gain credibility in usability work. Chilana et al (2010) found that usability experts struggle in gaining credibility with management and persuading developers to make changes based on user data, and these problems are made worse when domain knowledge is lacking. Complexity leads to a situation where more effort is needed to communicate with design team members, managers and developers. Usability experts also found it difficult to make usability an integral part of software development instead of an overhead. Particularly in the case of developers, who were sceptical about usability recommendations, motivating changes was difficult when usability experts lacked domain expertise. The more domain experts were involved, the less credibility became an issue. In our previous work on usability, we have found that there is a general need for usability process improvement, and that UCD, which can have a serious impact on HCI and usability practices, can have an important role to play (Winter and Rönkkö 2010). In UCD, the role of the human, who was once reduced to an operator, part of an information processing system, is replaced by a more active and
holistic contributor (Keinonen 2008), and based on the results of our study, we believe that this is necessary to resolve some of the problems that we have uncovered, and this is a movement that is gaining focus in the area of HRI.

Regrettably, much usability thinking, based on quick and cheap usability testing, has removed emphasis from the holistic view of the users, and has allowed teams to fragment usability into features supplemented by fanciful use cases derived from these use cases, which became apparent in the material in our study. User experience is supposed to be defined by realistic tasks that are meaningful to users, described in terms that they would understand (Wixon 2011). However, this cannot be done without the domain knowledge that only users can provide, and UCD could avoid this problem. Factors connected to why users buy or use products can be expressed in terms of what the user can do, and how easily they can do it, and this approach is captured by usability engineering metrics and goals. However, these metrics and goals must extend beyond a list of functionality and features to a more holistic product concept that reflects value. This kind of evaluative construct can only be measured through user research (Wixon 2011).

In the HCI field, there has been a move away from studying small groups of users who mainly used desktop computers, to groups of users looking for productivity tools, to studying the role of computers in entertainment (Karat and Karat 2011). Thus, HCI practice has changed radically, and it has been found that HCI practitioners must work in multidisciplinary teams, and coordinate with other areas in the company. Multiple perspectives are valuable, and Karat and Karat found that, as usability specialists, they needed to expend a great deal of time identifying and recruiting users to work in the HCI efforts, but that working with truly representative users is crucial in performing high quality usability work (Karat and Karat 2011).

As early as 2002, Adams claimed that the HRI community ought to approach HRI development from a human factors perspective, rather than an engineering perspective (Adams 2002). This is important to consider, given the focus we have seen in our study, where the engineering perspective has been prevalent. Adams maintains that the user’ perspective must be included throughout the design and development process in order to develop good interfaces. The HRI community should build upon results produced within the human factors community, and include users in the design and development of robot interfaces. Since an initial consideration in UCD is to examine the needs of users, it is beneficial to implement UCD in the HRI processes (Adams 2002). In 2003, Scholtz cited Hill, who also states that research teams in HRI should include human factors practitioners in multidisciplinary fields, and that HRI is more than simply creating a smart interface for the user. It is important to consider the skills of both humans and robots, and develop systems that allow both actors to fully use their skills (Scholtz 2003). However, it takes time to change paradigms, and in 2005, Adams (Adams 2005) still claimed that that HRI must move beyond the prevailing engineering approach to a UCD process. She maintained that it was only recently that the HRI field had become aware of the need to apply UCD
and other standard human factors techniques. There were still few efforts at implementing UCD, and it was difficult to gain access to real users, and this difficulty of has been one of the key factors of our study.

ABB has an extensive background in the area of mechanical engineering, and in recent years, there has been an increasing interest regarding the role of engineering in HCI. Since 2006 there is an Engineering Community SIG at the CHI conference. The engineering community sees that the role of engineering is becoming more important as the HCI field matures, and they are concerned with improving the economics of operations and improving safety to life or property. However, they find that they are still at the stage where they are discussing key issues, such as what is the importance of engineering in the greater HCI community, and are concerned with finding ways to develop positions that address these issues. They see the importance of HCI professionals making an impact in engineering projects, in order to realize the potential impact of CHI (Butler 2010). In 2011, the discussion at the workshop still concerned the serious challenges that faced the role of the engineering community in the larger CHI community. At the workshop, great emphasis was placed on practical matters, and on summarizing the issues and objectives of the workshop, and defining the key issues for 2012, where one of the issues was how HCI can work better with software engineering, which is the project context where the value of the field will usually be realized (Butler and Brooks 2011).

As previously mentioned, the case study began based on a desire to formulate and introduce measurable KPIs for usability. During the study, we discovered issues that make it difficult to formulate and agree on suitable candidates, due to obstacles and breakdowns in the organization, even though a number of candidates for KPIs could be created on the basis of the material in this study.

We have now traced the theories related to our case study, beginning with the fact that many of the problems that we have observed have their basis in organizational or behavioural problems. Given the complexity of the situation, we conclude that a wicked problem exists, that cannot be solved using the methods and techniques that are currently in use, and that attempts to introduce usability KPIs are not likely to solve the problems that they are aimed at. However, we do see that usability efforts, which have their basis in an attempt to understand the context and everyday work of the actual practitioners, are important, given the importance of including context in design processes. We also see that the only way to see and understand the context, and work practices that exist there, is to work in a process of UCD, which naturally includes the participation of actual users. However, we have also seen that there are a number of obstacles that must be overcome before this can be realised.

6.5 Conclusion and Further Work
In this final section, we present our work based on our focus on and discussions concerning usability, which was the main area of interest in the study. However, we see that the
conditions that exist reflect a situation that could be equally relevant for many other factors. Thus, the results are interesting as a general illustration of the problems of working with complex questions within complex organizations.

We have presented how problems occur when working with combinations of engineering problems and usability requirements in a complex organization. In a competitive world, there is often a consensus that it is new features and new products that are sold to customers. Each new feature can be a competitive advantage, as something that adds value or increases the price. Both adding value and increasing prices tend to increase profits. However, usability is only an added value which is hard to formulate, and is therefore a feature that is hard to sell as such. All products in the markets are said to be “usable” or “easy-to-use”; therefore, new features gain more attention by default. This leads to a situation where usability issues are often poorly formulated and understood, and there are few mechanisms that lead to satisfactory solutions. From a marketing point of view, it is too easy to just say “our product is easy to use” although proving it is difficult and would take a substantial effort. It is more cost effective to create a positive impression about usability by good marketing than by other means, whereas the functions promised must be found in the product, at the risk of litigation.

The context of this study is a large multinational industrial company, with roots in a traditional engineering field, and extensive experience of working with industrial machinery. The context is affected by the history and traditions of the company, which mean that the organizational structures and cultures that exist are still rooted in the successful strategies that have previously worked well for the company. However, changes in technology and in the marketplace mean that demands for usable products are forcing changes in the products that are produced and marketed, and in the way these products are used. This means that these structures are being challenged by the changes that are occurring, and in order to produce products that compete, it becomes necessary to change the view of what constitutes quality, and how this quality can be achieved. The business unit that was the object of study produces and markets industrial robots and robot systems. Regarding quality, the product is technically at the forefront of the market, and as far as usability is concerned is in line with other products of similar complexity and ability. So far, the engineering approach that has been followed has been successful in business terms when applied to clients within the traditional engineering background. However, the market has changed to include also smaller actors with diverse backgrounds and there is a realisation that usability is becoming a quality indicator, that improving usability is important and valuable, and the company was committed to finding ways of working with it. We believe that the situation that we describe is representative of other large organizations that are facing the same challenges of moving from a situation where changes in technology and markets force changes in deep rooted assumptions about how products should be produced and what constitutes quality.
In this study, we show that a combination of problems made it hard to work with usability KPIs in the way that the company hoped. Firstly, there are inherent problems that can be characterized as behavioural, and which are a result of organizational factors. We see how these organizational factors caused conflicts, breakdowns and barriers. The resulting problems were found in many different parts and on several levels of the organization. They affected other parts of the organization, and were compounded by the kinds of communication problems that are common in complex organizational structures. These characteristics led to problems when attempting to decide upon usability KPIs. Addressing these problems requires a holistic view of the structures and processes involved including a view where context is placed in the foreground whilst technological solutions are in the background. However, gaining that view is difficult. These behavioural problems, combined with the complexity of the marketplace, where there are uncertainties regarding domain knowledge and user needs, especially when expanding to new domains or sub-domains, lead to a situation where a wicked problem occurs. Such problems cannot be addressed through the engineering approach that has so far been used so successfully to target a market consisting mainly of engineering clients. However, the market has changed and today there is a need for a new approach targeting a much more diverse group of clients including non-engineering actors. This is difficult to realize in practice if adhering to the old approach targeting engineering companies.

The problems that are encountered are not based in the technology itself, but have their foundation in the behavioural problems. The above mentioned engineering approaches that have prevailed so far have been directed at solving clearly defined problems with clear goals, but these conditions are no longer the rule, as the complexity of the environment increases. In the design and development processes involved, the knowledge that is needed is distributed between different stakeholders, including end-users, who are required to participate, communicate and collaborate. These processes are characterized by change, conflict and multiple stakeholders, and the behavioural problems involved make participation, communication and collaboration hard to bring about, whereby reaching consensus becomes a problem.

Reaching a consensus amongst stakeholders is nonetheless of great importance. Solving design problems requires letting all stakeholders bring their contributions to the design process, in essence taking context into account. However, the fact that the situation is characterized by breakdowns and barriers makes it hard to get and use input from key stakeholders, whereby it becomes hard include context, to reach consensus, and to produce the product that will satisfy the maximum number of customers. To reach a situation where this is feasible would require changes in the organization and organizational culture, and as we have previously discussed, organizational change is a long and complicated process that is difficult to guide and manage. Still, we believe it to be the case that it is only if these changes occur that it will be possible to reach a state where usability work is seen as a natural and important part of the design and development processes, where incentives are created for working with usability, and where it becomes possible to work with usability
KPIs that different stakeholders can agree upon. This change process must have backing from management, but should, according to the principles of PD, involve staff and their representatives from all levels of the organization.

As discussed above, developing a complex system such as this requires us to deal with ambiguity and uncertainty, which requires us to place context in the foreground, whilst placing the technical solutions in the background. This is the opposite of what has the organizational culture has traditionally been focused upon. To actually achieve this requires the active participation of end-users, allowing us to create designs that combine human actions with the features of technology, to attend to the dynamics of situated practice. Only by doing this, we believe, is it possible to achieve consensus and thereby produce the best possible product.

Thus, to design the product that best serves the end-users requires knowledge of the context where it is used, and as we have seen, it is only the end-users who have that intimate knowledge. The end-users have unique needs, as they strive to understand and solve the everyday problems they are faced with, and they are the only ones qualified to judge whether a product fits their needs. This calls for a design environment that supports interactions between stakeholders, and the inclusion of end-users in the design processes. This need is something that the HRI field is only now becoming aware of, but which has become very clear in our study. One way of approaching this is through UCD, which is born from the need to transform technologies into applications, and the HRI community is becoming more aware of the fact that UCD is an important component. UCD sees the user as an active contributor who brings their own perspective to the design situation. Given the importance of context mentioned above, we believe that the contributions to be made by end-users are necessary to surmount some of the problems that we have found in our study. The inclusion of end-users is vital, to ensure that there is a balance between the skills of both humans and robots, to allow both actors to fully use their skills. However, although we see this as necessary, we still find that there are many barriers to be overcome before this becomes possible to realise in practice – in other words, we are still in a situation where we cannot reach the consensus needed to produce the best possible product. We believe that this cannot be done until UCD becomes a natural part of the development processes, allowing us to overcome some of the organizational barriers to usability work.

To summarize this discussion, we find that working with usability KPIs in the kind of organization that we have studied is difficult, and that achieving a situation where this is natural would require cultural and organizational changes that are likely to be far reaching and difficult to manage and achieve. However, we believe that to succeed in the long run with the type of product development in the kind of marketplace that we have studied requires that the company to make these changes.

To do this will require management to devote resources to actively encourage the change processes, and these change processes must involve staff from all levels of the organization, working not only with change processes but with the creation of KPIs. In
In order to create an understanding of the importance of usability, there must be a focus on developing KPIs that reflect the importance given to usability, and there must be incentives for meeting the goals set in the KPIs.

In order to create KPIs that reflect the actual user needs, it is important to know more about the users, and how they use the products in their everyday working life, and this requires mechanisms for user research and user participation. These methods do exist, but are still not widespread in the area that we have studied.

Most importantly, we believe that finding ways of involving different types of actual end users, with an ultimate goal of making UCD an integral part of the design and development process, is the most important factor to work with, and believe that this could be used to start the process of change that is necessary to work with effectively with relevant KPIs. Furthermore, we see that the process of UCD is vitally important for designing and developing products that can compete in a changing marketplace.

To conclude, this study is an investigation of the connections between usability and organizational factors. This is a field of research which, although it has been subjected to research, e.g. in the field of HCI as can be seen in the material above, still appears to be somewhat under-researched and discussed. It is however an important area, given the complexity of companies and the need to develop successful products. Beyond being a graphic illustration of the situation in a complex organization, this work demonstrated that there is a clear need for organizational change that can lead to a situation where it is possible to work with usability, and that one way of achieving this could be through working with a process of UCD. This is in accordance with the discussions that are beginning to take place within the HRI community, where the importance of UCD is becoming clear. We do find that the existence of a wicked problem means that there are no obvious paths to follow today to alleviate the situation that exists today, but although a wicked problem can, by its nature, never be fully solved, we believe that working with these factors can be a way of working towards a satisficing solution. Thus, work still remains to be done to actually solve some of the issues that we illustrate here, to find methods of including users in a collaborative process, in a fashion that is suitable for this kind of complex organization, working with a diversity of users in a rapidly changing field.

Beyond looking to the fields of HCI and UCD, it should be possible to look at the fields of participatory design and action research, to see in which way users can be included in these processes. Thus a future study may be to return to the organization, and through an action research process implement change by working closely together with and users and members on different levels of the organization.
References


References


References


References


:: 213 ::


:: 214 ::
References


:: 215 ::
References


:: 216 ::


:: 218 ::

References


:: 219 ::
Achieving product and process quality are among the central themes of software engineering, and quality is an important factor in the marketplace. Usability and user experience (UX) are two important quality aspects, particularly for interactive products. To achieve usability means producing products that let users do things in a satisfactory, efficient and effective way. To develop products with good UX, means going beyond usability, in ways that are still not clear to us.

Achieving good usability and UX is hard. This thesis is concerned with organizations working towards these goals. This research has revolved around understanding and improving the processes by which technology is designed and developed, and understanding the demands and expectations users have. It is about how companies can and do develop products with good usability and UX, and what stops them from working towards this as efficiently as they could. The usability and UX challenge has been viewed from the viewpoints of quality, organizations, and institutions, with a focus on participatory design, user-centred design and wicked problems.

This research can be characterised as empirical research performed over a period of seven years, in close cooperation with industrial partners. The research was performed using multiple data collection methods to create constructs and shape theory. The field methods have ranged from being a participant observer, to performing interviews and holding workshops with members of the participating organisations. A case study approach was initially used, but focus soon moved from case study methodology to a closer focus on grounded theory, and finally the focus shifted to constructivist grounded theory.

This thesis contributes to the field of software engineering in several ways. Usability has a long history within software engineering, human computer interaction, and design science, but the different discourses within the fields have meant that communication between the fields was problematic. The research in this thesis has moved between the different fields, contributing to bridging the gap between the areas.

The thesis provides an illustration of how usability work actually takes place in different types of companies, from a developer of operating systems for smartphones, to a global engineering company, which knows that it must find ways of working with, and measuring, usability and user experience. It gives concrete knowledge about the way in which companies can work with usability testing, and how they can provide information to satisfy the information needs of different stakeholders.

This thesis also provides a discussion of the state of UX today, taking up the problems that stop industry making use of the definitions and theories of UX that exist.

Thus, it gives an illustration of the different factors in product design, development and sales, from dealing with organizational factors to satisfying user needs, that all make usability work such a rocky road to navigate.

ABSTRACT

Achieving product and process quality are among the central themes of software engineering, and quality is an important factor in the marketplace. Usability and user experience (UX) are two important quality aspects, particularly for interactive products. To achieve usability means producing products that let users do things in a satisfactory, efficient and effective way. To develop products with good UX, means going beyond usability, in ways that are still not clear to us.

Achieving good usability and UX is hard. This thesis is concerned with organizations working towards these goals. This research has revolved around understanding and improving the processes by which technology is designed and developed, and understanding the demands and expectations users have. It is about how companies can and do develop products with good usability and UX, and what stops them from working towards this as efficiently as they could. The usability and UX challenge has been viewed from the viewpoints of quality, organizations, and institutions, with a focus on participatory design, user-centred design and wicked problems.

This research can be characterised as empirical research performed over a period of seven years, in close cooperation with industrial partners. The research was performed using multiple data collection methods to create constructs and shape theory. The field methods have ranged from being a participant observer, to performing interviews and holding workshops with members of the participating organisations. A case study approach was initially used, but focus soon moved from case study methodology to a closer focus on grounded theory, and finally the focus shifted to constructivist grounded theory.

This thesis contributes to the field of software engineering in several ways. Usability has a long history within software engineering, human computer interaction, and design science, but the different discourses within the fields have meant that communication between the fields was problematic. The research in this thesis has moved between the different fields, contributing to bridging the gap between the areas.

The thesis provides an illustration of how usability work actually takes place in different types of companies, from a developer of operating systems for smartphones, to a global engineering company, which knows that it must find ways of working with, and measuring, usability and user experience. It gives concrete knowledge about the way in which companies can work with usability testing, and how they can provide information to satisfy the information needs of different stakeholders.

This thesis also provides a discussion of the state of UX today, taking up the problems that stop industry making use of the definitions and theories of UX that exist.

Thus, it gives an illustration of the different factors in product design, development and sales, from dealing with organizational factors to satisfying user needs, that all make usability work such a rocky road to navigate.