Informal Feedback Rather Than Performance Measurements
– User Centred Evaluation in Scrum Projects

Marta Lárusdóttir¹, Åsa Cajander², Jan Gulliksen³

¹Reykjavik University, Menntavegur 1, 101 Reykjavik, Iceland, tel: +354 5996200, fax: +354 599 6200, e-mail: marta@ru.is

²Uppsala University, Department of Information Technology, Box 337 SE-751 05 Uppsala, Sweden, tel. +46 70 425 0786, fax:+46 18 511 925

³KTH – Royal Institute of Technology, Lindstedsvägen 3, Plan 6, 10044 Stockholm, Sweden, tel. +46 08 7909 538

Agile software development methods are nowadays becoming more common, but this does not mean that usability aspects are automatically catered for. Especially including user centred evaluations in Scrum projects are challenging for usability professionals. This paper presents results based on an interview study with 21 industrial usability professionals in five professional roles in Scrum projects. Various informal user centred evaluations are conducted. Evaluations generally occur throughout the project and some also precedes the project. The user centred evaluations conducted typically are empirical qualitative and seldom quantitative. The most common purpose for the evaluations is gathering feedback on design solutions. The business analysts mainly use one approach that is asking users about their opinion whereas the other roles use several user centred evaluation methods. Usability professionals often use informal feedback rather than performance measurements. Based on these findings usability professionals are encouraged to use various user centred evaluation methods in order to provide feedback to the three major phases of user centred design, the context of use, the user requirements as well as the design of the software. Researchers are encouraged to study the use of empirical qualitative evaluation methods in Scrum projects and describe new ways of conducting qualitative evaluations.

Keywords: User centred evaluation, usability, user experience, Scrum, human centred activities, agile development
1. Introduction

Usability of software is more important than ever before as those systems impact the lives of an increasing number of people. Before leaving home for work in the morning many people have already snoozed their alarm clock in their mobile phones, checked their mail, sent several text messages and read the latest news through an app. At work most people spend much time in front of computers to do their job, at the same time as they use several other software systems in parallel. It is a fact that most of us use technology on a daily basis to a greater extent than ever before, and software systems are integrated in all aspects of society. The usability of software systems therefore has become more crucial with time, and affects work as well as our spare time. Further, the software systems we use need to be efficient, work satisfactory and support our needs and goals. In this context user centred evaluation employed in systems development practice to ensure usability of software systems is of greater interest than ever, and is the focus of the study presented in this paper. The objective is to gain an understanding of industrial usability professionals' work with user centred evaluation in the most commonly used software development process Scrum, to be able to improve the user centred evaluation activity. Those improvements will extend the usability of the software systems generally and in the end impact people and their daily lives.

Working in software development projects is complex as there are many different aspects of quality of the systems that need to be taken into consideration. Previous research studies have pointed at large difficulties with software development projects generally, and project failures are far too common. However, even though it is well known that software development is expensive and time consuming the situation has generally improved compared to the 1990’s (Rubinstein 2007). The decision to start software development projects is often made based on anticipation of increased efficiency in the business, increased sales and other benefits for the stakeholders. When customers have acquired software to be developed, they are usually eager to see the end results as soon as possible. However, it needs to have satisfactory quality (Boehm 2006). Improving the process of software development, to make it more efficient, is therefore important, while maintaining the focus of quality. More efficient software development may also contribute to managing the costs in projects in a better way.

Around 20 years ago iterative and incremental software development processes were introduced to better manage risks and to acknowledge that it is difficult to fully specify systems beforehand (Boehm 1988). Agile processes for software engineering have evolved as a category of iterative and incremental processes to address the perceived limitations of the more established, plan-driven approaches for software development and have shown their value in quickly developing reliable software (Beck 2001). The most widely used agile processes in software industry today are XP and Scrum (Salah and Petrie 2009). In Scrum, the projects are split up in two to four weeks long iterations called sprints, each ending up with a potential shippable product. In Scrum self-organizing and well compounded teams are heavily emphasized, typically with six to eight interdisciplinary team members (Schwaber 1995).
One of the claimed benefits of using agile development is that customers’ needs are taken more into account than when developing software based on more traditional processes (Schwaber 1995). However, involving customers through the role of the product owner like it is being done in Scrum, does not mean that the real end-users are involve in the development. Nor does it adequately address their usability needs (Singh 2008). In an extensive literature survey on the integration of the usability needs into agile processes from 2010, the conclusion was that the end-user aspects have not yet been sufficiently included in the agile development processes (Sohaib and Khan 2010).

One of the most influential effects that the field of Human Computer Interaction (HCI) has had on software development in practice is the creation of a specific role for usability (Cajander et al. 2006, Boivie et al. 2006). A survey conducted in 2004 shows how usability designers, provided the right position, tools and resources, and provided the right attitude from the organization could make a huge difference in the development (Gulliksen et al. 2004). However, an in-depth interview study with usability designers working as consultants or in in-house development organizations shows that many usability designers feel isolated from the development team and experience the lack of actual influence in the development projects (Boivie et al. 2006). Proving and improving the usefulness of usability activities in the development processes is a key to success and one of the basics for the usability professionals for actually being able to make a difference (Gulliksen et al. 2006). Hence, examining how usability professionals are able to act to achieve the goal of maximizing usability within the remedies of software development processes such as Scrum is of great value to understanding the intended structure of the development process.

This paper focuses on gaining understanding on usability professionals’ experiences of applying user centred evaluation with the Scrum development process. The overall research question in the paper is: How do the usability professionals apply the user-centred evaluation methods? Furthermore, it is studied, if the different usability professional roles conduct evaluations in various ways and what the results of the evaluations are used for by the various roles. With this understanding, we are able to summarise the lessons learned and suggest improvements of the user centred evaluation activity, which will extend the usability of the developed software. In the context of practical Scrum projects we put forward the following research questions:

1. What methods do usability professionals use to evaluate their product?
2. Does the evaluation methods usage vary among the different professional roles?
3. What is the purpose of the evaluation?

There is much previous research on user centred evaluation, but the study presented in this paper is unique in its focus on analysing the usage of evaluation methods according to the roles of the usability professionals and how each of the professionals uses the results of the evaluation. Despite of the number of papers on user-centred design in agile development no detailed qualitative studies on the actual influence of user-centred evaluation in Scrum projects have been made before.
Following we will go through the theoretical background of user-centred evaluations and on Scrum, we present the methods used and the results from the interview study focusing on the evaluation methods used in relation to the different roles. Then the results are discussed and lesson learned for the integration of user centred evaluation in Scrum projects in practice are presented.

2. Theory and Background

In the following sections we will go through theories and existing research on user centred evaluation, user centred design processes, the agile development process Scrum and how usability activities have been integrated into agile software development.

2.1. User Centred Evaluation

User centred evaluation is a critical activity in software development to get feedback on how well the product fits the users’ goals and needs in the context of use (ISO 2010).

The type of information gathered in the evaluation has been evolving through the years. About 20 year ago, major emphasis was on gathering information on usability problems, which are flaws in the interface that cause problems for users (Lazar et al. 2009). Parallel to this, the emphasis was also on measuring usability in a quantitative way by measuring effectiveness, efficiency and satisfaction defined by the ISO 9241-11 standard (ISO 1998). Since around 2005, user experience has gained attention, where more subjective measures are emphasised (Hassenzahl and Tractinsky 2006). User centred evaluation can be used to gather information on three major factors, usability problems, usability and user experience.

Various methods have evolved, that describe the process for evaluating user interfaces, including how to plan evaluation, conduct evaluation and present results, parallel with the change in focus on what information is gathered in the evaluation (Woolrych et al. 2011). Recently a collection of 96 evaluation methods for evaluating the user experience of software was presented by Vermeeren et al.(2010).

There are various ways of grouping evaluation methods. In a study from 2007 about what evaluation methods had been used in research projects in HCI over a period of more than two decades, Barkhuus and Rode (2007) group evaluation methods by two dimensions. The first dimension is whether the results from the evaluation are quantitative or qualitative and the second dimension is whether the evaluation is empirical, meaning if people participated in the evaluation or analytical, meaning that evaluators analyse the design and do not include users or user surrogates in the evaluation. An example of an empirical quantitative method is measuring user performance in a laboratory setting and using surveys used to gather quantitative data. Empirical qualitative evaluation could be interviewing users or gathering feedback during meetings. Qualitative analytical methods could be peer reviewing the software and quantitative analytical method could be gathering usability problems with heuristic
evaluation or analysing interfaces through the GOMS model. These categories are illustrated in Figure 1.

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g. lab studies measuring completion time, error rates and surveys.</td>
<td>E.g. interviews, observing users and gathering feedback through social media.</td>
</tr>
<tr>
<td>Analytical</td>
<td></td>
</tr>
<tr>
<td>E.g. analysis of logs and GOMS.</td>
<td>E.g. expert review or peer review.</td>
</tr>
</tbody>
</table>

**Figure 1: Subcategories of user centred evaluation**

The results from Barkhuus and Rode (2007) literature study show that empirical evaluation is conducted in over 90% of the research studies performed in 2006. Analytical evaluation and informal evaluations are conducted in 3% cases each. Furthermore, no evaluation at all is conducted in around 3% of the cases. Over 70% of the empirical studies are quantitative, where around 10% of those also include qualitative measures. The remaining 30% of the empirical research studies are purely qualitative. This means that the dominant way of evaluating in research studies is conducting an quantitative empirical evaluation.

The research on evaluation methods conclude that evaluations may be classified according to several other orthogonal dimensions, including formal and controlled as opposed to informal and opportunistic (Twidale et al. 1994). Furthermore, evaluations can be used to gather objective data on the status of the usability of the software, called summative or to gather feedback for redesigning the software, called formative. Moreover, these researchers conclude that the most common styles are summative-quantitative-controlled and formative-qualitative-opportunistic. These two can in short be referred to as formal and informal evaluation methods. An informal evaluation method hence denotes formative evaluations with a qualitative character that take place when the opportunity is given. Informal also indicates that the arrangement is without fixed rules and does not need to produce an official result.

The effectiveness of using various user centred evaluation methods has been explored by researchers since the early 1990’ (Jeffries et al. 1991, Karat et al. 1992, Desurvire et al. 1992). In these studies, the effectiveness of using a particular evaluation method was measured by comparing the number of usability problems found by using each method. One example of this is comparing the effectiveness of using one set of guidelines conducting heuristic evaluation against using another set of guidelines (Hvannberg et al. 2007). Measuring the effectiveness of doing evaluation by counting usability problems found by using each method was criticized by Wixon (2003). He states that the results from research studies on evaluation methods fail the practitioner because the studies were conducted in laboratories and not in realistic contexts.
To respond to this claim, user centred evaluation has been studied in practical settings in various ways. The usage of the empirical method think-aloud was studied in an observational study in real settings (Nörgaard and Hornbæk 2006) and the usability improvements achieved on one particular system, when various evaluation methods were used by usability practitioners (Uldall-Espersen et al. 2008, Nörgaard and Hornbæk 2006) in a case study to name two examples.

Other researchers have examined how evaluation is done in practical settings by conducting surveys and interview studies. In a survey study from Norway, the results show that about 90% of the participants conducted user testing (Bygstad et al. 2008). This is similar to the results from the literature study on the use of evaluation methods in research studies (Barkhuus and Rode 2007). One to ten users were included in the usability testing in about 75% of the cases in the survey study in Norway and more than eleven users in the remaining cases Norway (Bygstad et al. 2008). When asked about how the users were selected for the testing, 40% responded that it was representative sample of users, which means that 60% of the users in user testing are not real users but more like user surrogates.

The results from an international web based survey show that user testing was used in 70% of the cases for evaluating the software (Monahan et al. 2008). All the qualitative methods asked about, including observations, interviews, contextual design, probes and diaries, were also used for evaluation. The most influential factor for choosing an evaluation was time constrains. These results show that various methods are used The results from a survey done in Sweden show that the best rated evaluation method by practitioners is the empirical method think-aloud, were around 85% of the participants rate the method as very good or fairly good (Gulliksen et al. 2004). Less than 60% rate quantitative evaluation conducted in a usability laboratory as very good or fairly good method and around 65% of the respondents rate the analytical method heuristic evaluation as very good or fairly good method. From these results it can be seen that the rather informal approach of empirical evaluation through using the think-aloud evaluation gets the highest rating from the practitioners.

The obstacles for doing usability evaluation in practice were explored in a survey and interview study in Denmark (Bak et al. 2008). The results show that some type of usability evaluation is conducted in about 75% of the companies involved. It is not analysed in the study what form of evaluation is conducted, whether it is empirical evaluation or analytical, but by exploring the understanding of the term “usability evaluation”, the results show that around 80% of the participants describe evaluation as empirical by stating that users are involved in the evaluation. The two major obstacles found in the study were resource demands both time and money; and the developer mind-set; that is that developers find it hard to think like users. In relation to this, some participants described that the main focus of developers are on the programming aspect to write beautiful code and not so much in participating in usability evaluation (Bak et al. 2008).
2.2. Human-Centred Design Process

A common framework for human-centred design process when studying and understanding the usability practices is the ISO 9241-210 (2010) (Human-centred design for interactive systems). This standard is recently developed version of ISO 13407 (1999) standard (Human-centred design processes for interactive systems). The ISO 9241-210 standard defines human-centred design as “Approach to system design and development that aims to make interactive systems more usable by focusing on the use of the system, applying human factors/ergonomics and usability knowledge and techniques (p. 2)”. It is stated in the standard that “Human-centred design activities can be incorporated in design approaches as diverse as object-oriented, waterfall, HFI (human factors integration), agile and rapid development (p. 10)”. This human-centred design approach should therefore fit into Scrum as well as in any other development processes. Four human-centred activities are described in the ISO 9241-210 (2010) standard that shall take place during the development of interactive systems as shown in Figure 2.

![Figure 2: The framework for human-centred design according to ISO 9241-210.](image)

It is stated in the standard that these four human-centred activities are interdependent and each activity uses outputs from the other activities. This means that when an understanding and a specification of the context of use have been made, the user requirements should be described based on this understanding. Furthermore, design solutions are made based on these user requirement specifications and are evaluated.
Iterations could be made on the three other human-centred design activities according to the results of the user centred evaluation, where appropriate. User centred evaluation therefore has a vital role in the human-centred design process.

The type of user centred evaluation conducted by usability practitioners in each development phase was studied in a survey study (Venturi et al. 2006). The results show that qualitative evaluation was conducted at least once during the software development by almost 70% participants, most frequently during the design phase in almost half of the cases. Quantitative evaluation through user testing was used in at least one phase by around 40% of the participants, most frequently in the design phase and testing phase by almost 20% each. In another survey study seven methods were mentioned being used in the evaluation phase of the development (Bark et al. 2006). Qualitative evaluation through user tests was conducted by almost all the participants, and the analytical methods, expert evaluation and heuristic evaluation by more than 75% of the participants. The rating of the usefulness of the methods had almost the same pattern as the usage, so user tests were highest rated, followed by expert evaluations and heuristic evaluation. Surveys were, however, rated as the least useful method followed by cognitive walkthrough that got a bit higher rating than surveys. In the third survey study, user testing was used for evaluation in half of the cases, and for requirements gathering, understanding the context and inspiring concept creation in the other half of the cases (Monahan et al. 2008). The most influential factor for choosing a method for participants working in the software industry was time constrains. These results show that evaluation is used to support all the other three user centred activities in software development.

2.3. The Software Development Process Scrum

Scrum is one of the software development approaches that belong to the agile software development family (Sutherland and Schwaber 2007). Other well-known software development approaches in this family are Extreme Programming (Beck 2001) and Kanban (Anderson 2010). However, research indicates that Scrum is the most frequently used process of the agile software development processes (Hussain et al. 2009). The main features of the agile development approaches are intense communication between different stakeholders and rapid feedback based on regular delivery of working software (Beck et al. 2001). Another feature of the agile approaches is that they try to connect the field of software development and the field of business development. Moreover, they are iterative and incremental, and that they aim at being lightweight and adapt to changing requirements and circumstances. Many of the agile software development approaches were created in the 1990’s, but some of the prominent concepts can be traced back to the 1930’s and to the ideas of Lean production (Larman and Basili 2003). The agile software development approaches are described in the Agile Manifesto from 2001 that includes the core ideas of the different approaches (Beck et al. 2001).

Scrum appeared as a software development approach in 1995 and was presented by Jeff Sutherland and Ken Schwaber (1995). Today, it is a widely used software
development method, and some claims it to be the de facto standard for agile development. Scrum is a simple framework with a few ceremonies, roles and artefacts (Sutherland 1995). The roles are called Product Owner, Scrum Master and Team (Schwaber and Beedle 2002). The Scrum team is self-organising and works independently. The theory prescribes daily Scrum meetings where the team meets and plans the work during the day, and where the tasks are distributed in the group. The work in the team should be guided by collaboration and communication. The Scrum process is illustrated in Figure 3.

![Scrum Process Diagram](image)

**Figure 3: An overview of the Scrum process.**

The Product Owner has the responsibility to represents the needs and ideas to the other stakeholders including the customer. In this work, the Product Owner writes so called user stories and manages them in a document called the Product Backlog. User stories are a standardised way to express the users’ needs and the business value in sentences with a specific structure. The Product Backlog contains the requirements for the system to be built, including the user stories. Parts of the Product Backlog may be technical, and parts may be more human oriented.

The different iterations in Scrum are called sprints, and it is recommended that these are about two to four weeks long. The sprints are planned in sprint planning meetings, and the requirements to be addressed in each sprint are defined in the Sprint Backlog. By the end of each sprint, the Scrum team should have completed a potentially shippable product to release. During the end of each sprint there is a demo session showing other stakeholders what will be delivered in that particular sprint. At this point Scrum recommends that the team collect feedback from the stakeholders to adapt the system to the users’ needs in subsequent sprints. The work conducted before the actual Scrum project has started is often referred to as Sprint Zero or as a pre study.
The team and the Product Owner define an exit-criterion of the items in the Product Backlog called Definition of Done. It describes when the item is complete and can be delivered to the customer, and the text might include descriptions of tests to be done.

The Scrum board is a physical or electronic task board used by the team members to manage the different tasks on the Sprint Backlog and what to do during the different sprints.

A series of observational studies of agile development teams have been conducted (Robinson and Sharp 2010). It is concluded that customer collaboration differed between the teams depending on the application and the organizational context of development. The technical practice “is not just any (‘warm and fuzzy’) collaborative, communicative and co-ordinating activity that is acceptable but the detailed work, intimately connected to the technical that our analysis has revealed. The creation of working software is a socio-technical enterprise.” (Robinson and Sharp 2010).

Even though Scrum is not a strict process there are many constrains implied on conducting usability activities while using it. The sprints are short and potentially shippable product should be made in each of them. Studying how usability activities need to be adjusted to Scrum is important.

2.4. Usability Activities in Agile Software Development Approaches

There are a few research studies that focus on how usability activities are conducted with agile approaches. The results from a study with five IT professionals show that most of them felt that combining usability activities with Scrum added value to the product (Hussain et al. 2009). This result was further confirmed in a survey with 92 practitioners world-wide where almost all participants concluded that the combination of the two methods added value to the product and increased usability (Hussain et al. 2009).

Three project teams were examined in a field study to investigate the use of usability activities within agile development in one particular organization (Chamberlain et al. 2006). The results show that all projects had some degree of design before the coding started. It is also noted that the project, where usability activities and agile principles were well integrated, had the least problems when delivering the product to the customers.

The adaptations to usability activities that usability practitioners did when changing from traditional development process to agile development in one organisation is described in a paper by Sy (2007). The study was done in one particular company. Their results show that the new variation of the user centred design methods produce better-designed products than the older versions of the same methods. One of the positive experiences was that the communication in the agile development team did narrow the gap between uncovering usability issues and acting on those issues by incorporating changes into the product.
An observational study was conducted of a mature Scrum team in one large organization, and their interactions with the user experience designers working on the same project (Ferreira et al. 2011). The results show that cooperation between the agile developers and user experience designers was achieved through ongoing articulation work. Evaluations on the design from the user experience designers were done in the team of developers, where the team analysed the design, often immediately after the user experience designers had sent it to the team. Then the developers did sent feedback to the user experience designers or met face to face with them to give feedback. The results show that developers and user experience designers made decisions only within their disciplines, which give the indication that the co-operation in the Scrum teams is not as much as expected.

Themes that are perceived by user experience practitioners to be highly influential in the success of integrating user centred activities and agile approaches were examined in an interview study (Kollmann et al. 2009). The ten user experience practitioner had various roles such as user experience architects, user experience consultants, information architects and interaction designers, and they worked in different organisations. The results show that the practitioners wanted to have the design ready one sprint ahead the implementation, if possible. Sometimes this was not possible because of time constrains. Furthermore, it is stated that lightweight approaches to user involvement are important. Many of the participants accessed end users quickly through user groups or panels. All participants agreed that rather than having no users at all, they would accept using colleagues or a smaller number of users. One of the conclusions is that participants strongly expressed the necessity to be flexible, practical and sensible (Kollmann et al. 2009).

The key principles and main attitude behind the user centred design approaches should carefully be considered when tailoring the Scrum process to fit the needs of the project. Our conclusion is that human-centred activities could fit well with agile development, but that there is a need to adapt the user centred design methods to fit in the agile framework. However, to be able to do that we need to find out in what way human-centred activities are used by practitioners of agile methodologies today, and why.

In this paper, we study on how user centred evaluation is conducted in the agile process Scrum to be able to suggest improvements to the evaluation activity both for usability professionals and researcher with the aim of extending the usability of software used in the future.
3. Method

In this chapter the method used for the interviews will be described, the participants background and the data analysis method.

3.1. The Interviews

Our qualitative study included 21 semi-structured interviews where an interview template was used and questions were adapted in accordance with the organizational role of each informant and their usage of evaluation methods to get rich information from the informants on the particular methods that they had used. Questions to the informants were related to their background, and their experiences working with Scrum and usability. The interview template included the following themes:

1. Background
2. Experience from using Scrum and from using user centred design
3. Experience from integrating user centred design into Scrum
4. Role of users
5. Non-functional requirements
6. Design methods
7. Sprint zero
8. User centred evaluation
9. Definition of done
10. Responsibility for usability
11. Final remarks

The themes on the interview template were all covered in all the interview, but not always in the same order, because the informants were asked to explain their ways of working, so the interview would be more like a in depth conversation. Most interviews were conducted on site in the informants’ organisations and lasted for about one hour. Most interviews were conducted by two researchers interviewing one respondent, where one researcher acted as a conductor and the other as a note taker. One interview was conducted through Skype with a web camera. The interviews were recorded. Due to technical problems, however, three of the interview recordings were incomplete. All recorded interviews were transcribed verbatim. The quotations provided in the text are not always verbatim, but sometimes slightly rephrased in order to be more readable and representative. Furthermore, we refer to our informants as males despite their real gender when describing their comments.

3.2. The Informants

The informants were found through personal contacts, a presentation at an HCI interest group, social media and suggestions from informants already interviewed. The people who participated in the interviews all worked with Scrum and were interested in integrating usability activities into the Scrum process. We have chosen to group the
roles of the informants in five categories based on a description of the responsibilities for usability professionals described by Mayhew (1999), complemented by a fifth role described by Schwaber and Beedle (2002):

(1) **Usability expert role** included informants that had extensive knowledge of HCI. Their main job task was usability activities. They had the authority to decide on usability issues for the project. Typically, they were not members of one particular development team.

(2) **Interaction designer role** included being one of the members of a team in the Scrum process and the main activities were to design and develop the user interface.

(3) **Business analyst role** whose main job task was to analyse the requirements of the software during pre-studies.

(4) **Developer role** included being a team member developing and testing the software.

(5) **Scrum manager role** had the Scrum Masters role, which are responsible for the project management of the Scrum team or the Product Owner role, which is responsible of specifying what is needed in the software from the customers’ viewpoint and prioritize these needs.

In Table 1 an overview of the informants’ background is given.

**Table 1: Overview of the informants’ background**

<table>
<thead>
<tr>
<th>Categories of job roles</th>
<th>No</th>
<th>Gender</th>
<th>Age</th>
<th>Industry experience</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability experts (UE)</td>
<td>5</td>
<td>4 male, 1 female</td>
<td>32 - 46</td>
<td>3 - 12 years</td>
<td>3 advanced education, 2 research education</td>
</tr>
<tr>
<td>Interaction designers (ID)</td>
<td>7</td>
<td>6 male, 1 female</td>
<td>31 - 55</td>
<td>2 - 30 years</td>
<td>1 no university degree, 2 basic university degree, 4 advanced education</td>
</tr>
<tr>
<td>Business analysts (BA)</td>
<td>4</td>
<td>1 male, 3 female</td>
<td>40 - 42</td>
<td>12 - 15 years</td>
<td>3 basic university degree, 1 advanced education</td>
</tr>
<tr>
<td>Developers (DE)</td>
<td>2</td>
<td>1 male, 1 female</td>
<td>29, 30</td>
<td>1 - 5 years</td>
<td>2 advanced education</td>
</tr>
<tr>
<td>Scrum managers (SM)</td>
<td>3</td>
<td>2 male, 1 female</td>
<td>30 - 33</td>
<td>3 - 8 years</td>
<td>2 advanced education, 1 research education</td>
</tr>
</tbody>
</table>

Each of the informants were marked by two letters explaining their role and a number, so the first informant categorized as a usability expert got the marking UE-1. The two letters are given in brackets in Table 1. Detailed information of each informant’s background is given in Table 4 to Table 8 in Appendix A. In total there were 7 females and 14 males, their age reached from 29 – 55 and they had been working from 1 to 15 years in the industry. Their education varied from having no university education, basic university degree which was typically Bachelor, advance degree which was typically a Master degree to having a research education or a PhD degree.
The informants worked at 14 companies in various organisational contexts. The main types of organizations were product development and consulting companies, or a combination of those. Some of the companies were international, having employees worldwide. The number of employees reached from 8 to 12,500.

3.3. The Data Analysis

Data from different interviews were compiled and rigorously analysed by two researchers working together to ensure quality. Thematic analysis, as defined in (Ezzy 2002), was used as an analytic approach where data was reviewed, organised and analysed in order to identify themes in an iterative manner. The respondents’ statements were categorized into codes in a software program for data analysis. First, a scheme of detailed codes was made and a couple of interviews coded with that coding. Then, the codes were made more general so that similar subjects would fall in the same code. The final version of the code scheme contained 13 codes and all the interviews were coded according to those. Some codes represent a set of inductive constructs while others are rather a set of predefined ideas. At this stage interpretations were made and discussions included search for alternative understandings and interpretations to statements made by the informant. In this work and in the subsequent writing up of the data the quality criteria defined by Klein and Myers (1999) were used to ensure quality.

4. Results

In this section results from the study are presented and summarized in three sections. The first section describes what user centred evaluation methods are used by each professional role, as defined in subchapter 3.2. In the second section results on the different purposes of the evaluations described by the informants are given. Finally the results are summarised in the third section of this chapter.

4.1. User Centred Evaluation Conducted by each Professional Role

As described in the background there are various types of user centred evaluations, see chapter 2.1. The information from the informants was analysed according to how many conducted empirical or analytical evaluation were conducted and if they conducted quantitative or qualitative evaluations as described in Table 2. Furthermore the evaluation methods that are used for these types of evaluation are also described in Table 2. The data was analysed according to what role the informants had as defined in subchapter 3.2 above.
Table 2. User centred evaluations conducted by the professional roles

<table>
<thead>
<tr>
<th>Professional Role</th>
<th>Type of Evaluation</th>
<th>Empirical Qualitative Evaluation</th>
<th>Analytical Qualitative Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Evaluation method</td>
<td>Measuring user performance and surveys</td>
<td>Observing Users</td>
</tr>
<tr>
<td>Usability experts</td>
<td>N</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Interaction designers</td>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Business analysts</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Developers</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Scrum managers</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>N</strong></td>
<td><strong>21</strong></td>
<td><strong>13</strong></td>
</tr>
</tbody>
</table>

There are several interesting things that can find in Table 2. Only four informants conduct empirical quantitative evaluation by measuring users’ performance or doing surveys with 300 to 500 participants, two of the usability experts and one Scrum manager that previously worked as a usability expert. Almost all of our 21 informants conduct empirical qualitative evaluation, where the most common way is to ask user about their opinion, done by 18 informants. Other empirical qualitative methods used are observing users conducted by half of the informants and receiving qualitative feedback from user surrogates conducted by one third of the informants. The informants get feedback from users in some form or the other, like through user forums, blogs and discussion groups. These are new and informal ways of receiving feedback from users. Some of these evaluations with users are done during the development, and some after delivering part of the system in a sprint.

All the usability experts and both the developers use inspection methods for receiving analytical qualitative feedback, but none of the business analysts do so. All the usability experts and the business analysts conduct evaluation with real users, but none of the business analysts and only one usability expert involves user surrogates. Half of the other roles ask user surrogates in some cases and almost all of them do also ask real users.

In the following further information will be given on the evaluation methods used by each usability professional role. Further detail on what methods each person uses can be seen in Table 9 in Appendix B.

*Evaluation Conducted by the Usability Experts*

An interesting finding is that the usability experts use at least three or four different approaches to evaluate their systems. All of them do some kind of empirical evaluation. Two usability experts conduct quantitative empirical evaluation by measuring user performance in a formal user centred evaluation and one uses questionnaires to gather quantitative data with around 300 users. One informant asks external expert company to
He comments: “I’m not doing it myself but we let someone else do it, and then we can sit and watch.” The evaluation is often done on a running high fidelity prototype and sometimes even on the actual system. The informant stresses that user testing should be done before implementation starts. Furthermore, he states that it should be done every now and then on the actual product “to make sure that what has been done is actually in line with the detailed requirements”. Moreover, user performance measurements are most important when you have “the big picture in place” and when you can see how the system works. He states: “When you have the big picture you can test a number of different things at the same time” (UE-2).

All the usability experts gather qualitative empirical feedback by observing users using the system or a prototype of the system, through informal field studies, informal think-aloud or observing users using paper prototypes. Four usability experts receive feedback through interviews with real users or surveys and all of them get feedback from users in an informal way, like informal field studies, informal think-aloud or observing users using paper prototypes. One usability expert gets feedback on the design from a user surrogate that was also in the development team for the software. One informant describes that they receive feedback from users through informal field studies because he thinks that obtaining information that way gives almost as much information as empirical quantitative evaluation. He says: “I prefer just to hang around to be there and see how they work with our system today, what we can improve, easily, what are the big shortcoming in our system and how can we improve them” (UE-1).

One informant describes that they gather qualitative empirical data by observing two or three users using paper prototypes of the system. This informant explains why they do not have more users involved by saying: “It is difficult to motivate people to participate in user centred evaluations on paper prototypes. If the system had been more elaborated and closer to delivery, then it would be easier to motivate users to participate in the evaluations. I think that is the biggest struggle that we have is getting people motivated” (UE-5).

This informant describes a new way of receiving feedback from users through blogs. In their company the usability professionals did reply to users’ blogs more and more the last couple of years, and the users have included comments about user experience in the blogs. The informant remarks: “And that is really great because we get the input and we don’t have to ask for it” (UE-5).

All the experts do analytical qualitative evaluation with developers either through discussions on the design in the team or by doing inspection evaluation both by walking through the design in the team or by doing inspection evaluation by themselves.

**Evaluation Conducted by the Interaction Designers**

All of the interaction designer conduct qualitative empirical evaluation. One of them conducts quantitative empirical evaluation by using a survey with 500 participants but none of them by measuring users’ performance.
Five interaction designers observe users in various ways. One of them observes users using paper prototypes while solving tasks and interviews them afterwards. Another observes users using the system solving tasks and gets feedback from them on a business conference and two other observe users in their real settings. The fifth informant describes a quite informal evaluation process, where the users participate in user group meetings or workshops as they call it. They have about twenty-five users that are assigned to a group, of which ten to fifteen people attend in a meeting every three weeks. The informant describes that this constant change of participants at the meetings has both drawbacks and advantages. They do not want “beginner feedback at all times” since they know that the majority of the users are intermediate. Moreover, they are going to make the design of the system available online to receive feedback from all users. The users themselves requested the online system. This is also done in order to prepare the users for the new system, and to make “them feel as participants in the process too” (ID 3).

Almost all the interaction designers do get feedback from users either through interviews, group meetings or questionnaires. Sometimes this information is gathered in relation to observing users. One informant describes this approach in the following way: “We sat in our own room at a conference and we brought in users and let them test the products. Afterwards, we interview them and basically we just watch when they run the application and talk about it” (ID-4).

It is interesting to note that all except one of the interaction designers do inspection evaluation or evaluation with user surrogates. Two of them observe user surrogates and two get feedback from user surrogates through discussion, furthermore two designers get feedback from the team through discussion and one interaction designer has done inspection evaluation. One informant describes that they do not evaluate with real users, but with a sponsor paying for the product, who tests the system. The informant describes the evaluation in the following way: “we had a lot of good feedback, but we had no real users”. The evaluation with the sponsor is quite informal (ID-4).

**Evaluation Conducted by the Business Architects**

The main evaluation method for the business architects is asking users for their opinion. Three of the four business architects get feedback from users through group meetings, which they call workshops and one gets feedback right after delivering a subversion of the system.

One informant describes how they conducted the workshops by saying: “We had one workshop for each department where we discussed goals and problems with the existing system and the existing way that they work today. After that we started talking about how they want to work according to their goals” (BA-3).

One informant describes that they present a working version of the system after the first sprint, and the users are given one week to test the system and to give feedback. So the feedback is delayed one sprint. In the meantime the project continues with other items from the backlog. When they deliver a version of the system to the end users after
every sprint, they always get the feedback: “Now I understand what you were talking about”, the informant explains. In this way the users get a real experience of the system, as compared to whiteboard drafts. These whiteboard drafts do not give the same feeling or understanding of the system, according to the informant. Furthermore, the informant says that users that are not experienced in participating in systems development projects give little feedback on simple prototypes in group meetings called workshops but give better feedback on a running system. Experienced users give much more feedback, and also give feedback on paper prototypes and early drafts (BA-1).

Only one business analyst informed on gathering empirical qualitative information by observing users. This informant describes it by saying: “The users came to visit us to test the system. So we checked, ‘are the field labels good?’ and ‘do they understand what to do?’” This was usually done in a very iterative manner, an interaction designer did mock-up prototypes of the system, he and the informant evaluate with users, and then change the mock-ups and evaluate again, so the same persons did the requirement analysis, the design, the evaluation and the redesign during the development of the system (BA-2).

None of the business architects did empirical quantitative evaluation, get feedback from user surrogates nor conduct qualitative analytical evaluation.

**Evaluation Conducted by the Developers**

One of our developers gets feedback from users in the end of the project. This informant explains that the evaluation of the system was planned as a part of each delivery of the system, but that this was changed and that it was “Scrum turning into a waterfall”. They do some internal testing in the development team during the project, but the user centred evaluation and testing with users occurs only at the end of the project (DE-1).

This developer evaluates the system both by inspection and by observing a user surrogate. He explains by saying: “At some point we showed him how things were going like you can do this and he said ‘yes that’s good’ and ‘this should be different somehow” (DE-1).

The other developer only reports the usability issues that he finds while testing other things in the software. None of the developers gathers quantitative empirical feedback nor observe users.

**Evaluation Conducted by the Scrum Managers**

Each of the Scrum managers uses three different methods for evaluating their software. One of the Scrum managers described that they conduct empirical quantitative evaluation by measuring user performance in think aloud tests around twice a year. The other two Scrum managers get feedback from users both through observations and interviews. Two Scrum managers observe user surrogates and two informants used inspection, were one mentioned heuristic evaluation and the other used a more informal inspection. One informant explains that he does inspection evaluation quite often, by explaining: “Usability inspections are required almost several times each week where I
am asked to verify user interfaces. So I am doing them all the time and try to follow the principles that we have set up for this kind of interfaces” (SM-2).

4.2. *The Purpose of the Evaluation*

Evaluation is mainly used to get feedback for the three other human-centred activities defined in the ISO 9241-210 (2010) standard, which are the context of use, user requirements and design. The results on what is the main purpose for conducting the evaluation; that is what information the evaluators wanted to gather in the evaluation are described in this subchapter. An overview is given in Table 3. Furthermore, it is described how the results from the evaluation are presented to the practitioners.

**Table 3: The purpose of the evaluation**

<table>
<thead>
<tr>
<th>Professional Role</th>
<th>N</th>
<th>Feedback on Context of Use</th>
<th>Feedback on User Requirements</th>
<th>Feedback on Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability experts</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Interaction designers</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Business analysts</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Developers</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Scrum managers</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>21</td>
<td>12</td>
<td>16</td>
<td>19</td>
</tr>
</tbody>
</table>

It can be seen in Table 3 that almost all the informants gather feedback on the design of the software they are developing, 16 gather feedback on the user requirements and 12 on the context of use. In the following it is described how the feedback is gathered in each of the human-centred activities.

*Feedback on the Context of Use*

Just over half of the respondents mention that they gather information on how the users are working. One informant uses both questionnaires and interviews to gather this information and one interaction designer observes and interviews users in the field during user centred evaluation near the end of the project. Three respondents use workshop meetings to get feedback from users on the context of use before the actual implementation of the product begins. The remaining respondents use one method each namely field studies, informal user observation, inspection evaluation and observation of a user surrogate to gather this information.

*Feedback on the User Requirements*

A majority of our informants gather information on user requirements by evaluating prototypes with users or user surrogates both by observation or interviews. The design of the prototype can be preliminary showed in sketches, mock-ups or paper prototypes or more detailed. Some informants mention that this is done every now and then in the
One informant uses questionnaires to ask users to prioritize the requirements during workshop meeting with the users. In one company, an idea forum is used to gather information on user requirements. The informant describes it by saying: “On the web site you can sign up to be a product adviser, so you are more like involved in the product development. The users say, I need this and I need this and then we have to ask why” (SM-3).

One part of describing user requirements is describing the usability criteria by stating the level of the usability for the software the professionals want to reach. Several informants described different approaches on how to include usability criteria in the Scrum process. It is clear that there is not one way of doing this in practice, and that the challenges for the usability professionals are numerous. Only four of our informants say that a precise description is given of the usability criteria, in eight cases a broad criterion is stated and five informants said that no usability criteria was described. One of the usability experts describes his experience in stating usability criteria by saying: “As soon as you write anything broad or qualitative it doesn’t mean anything or it gets lost. When you have something that’s very clear and measurable, it has direct impact on usability, like the booting time can’t be more than five seconds. These kinds of things are very clear and obvious and that you can easily put in there and make sure that they are enforced. Then we don’t have to care as much. Somebody else enforces them and all the team members are working towards reaching those goals” (UE-4). Hence, the usability expert perceives that writing measurable usability criteria makes his work easier; he does not have to be as alert, because more people are involved in enforcing those goals and working towards those. Given this argument is surprising that only four of our informants describe usability criteria as measurable goals.

**Feedback on the Design**

All of the informants gather information from evaluations as input to redesign of the system. Some of the informants use several kinds of evaluation methods each, for example surveys, interviews and observations to gather input for redesign. Some of the informants evaluate prototypes before the actual development begins, some evaluate also in every sprint, whereas others evaluate running prototypes every once in a while during the development.

One of our informants describes the importance of doing evaluation before the actual implementation starts by saying: “Most evaluations are done in the strategic part of the project. What I would say is that in the end of the sprints it’s more about confirming that what has been done is actually in line with the detailed requirements. I think it is more important to do evaluation before production and then anytime when you have a decent chunk of functionality to evaluate. Then you can evaluate maybe a number of things at the same time” (UE-2).

Two of our informants discussed that there is not much time in Scrum to do evaluation or to gain a holistic overview of the system. One informant discussed that there is little time in Scrum to back up and look at the whole picture; in Scrum “the eagle view is missing” he stated. He explains: “A developer is too busy working on a
yellow label describing one task, so usability is not in the picture.” The developers’ concern is much about details like “I want this button red instead of blue”, but not about the work flow or how it actually is to use the system, he explained (ID-7).

Furthermore, one of our informants describes that they had no requirement from the customers to do evaluation. He explains: “It was like they were in a hurry. Their only requirement was that we had to finish the project and everything should look the same, so we had no requirement of making any user centred evaluation” (ID-1).

**Presenting the Results from the Evaluation**

Some informants explain that it is not clear how the Scrum team should describe the results from user centred evaluation. Sometimes the evaluation results are used to describe a new user story in the sprint backlog, and sometimes the problems found are described as amendments of a current story. Two of the informants describe that usability problems presented as new user stories are more likely to gain the developers interest than when presented as the same user story that needs improvements, because in Scrum much emphasis is on finishing the user stories. Describing amendments to an already defined user story means that it can’t be marked as finished.

When a developer has defined that some functionality is implemented, the usability expert checked in some cases if it had usability problems. The expert is wondering: “Should the developer be allowed to say it’s done or should the usability expert have time to walk through it before the developer says it done?” The developer might perceive the functionality to be done, whereas the usability expert says: “No, no, no, you’re half way there”. The usability expert wonders: “Should I allow the developer to say it’s done and I can put it back and say that we have to have this fixed, or should it be another post-it note describing a new user story for an improved version?” (UE-4).

**The Multiple Purpose of the Evaluation**

Many informants describe that the different evaluation methods are used to gather feedback that contributes to all four human centred activities. This is often the case especially when conducting workshop meetings. This indicates that in practice the four different human centred activities presented in ISO 9241-210 (2010) are integrated and seamless in relation to each other during the workshop meetings. However, it is still noticeable that all the evaluation methods used during the workshop meetings in Scrum are of an informal character as described in subchapter 4.1. The experienced professionals in the interviews have realized the necessity to capture qualitative feedback through evaluation in many different ways, and it is noticeable that they are quite innovative in finding ways to channel the comments from different stakeholders. The usage of new forms to capture comments, such as blogs or other social media was positive and interesting, for details see subchapter 4.1.
4.3. Summary of Results

To summarize, we relate the main results to the research questions. The first research question is: What methods do usability professionals use to evaluate their product? The results from this study show that:

1) **Various Empirical Qualitative Methods are Used**
   Almost all the informants use qualitative empirical types of user centred evaluation. The interviews clearly indicate that conducting informal empirical qualitative evaluation is a success factor, see subchapter 4.1

2) **Quantitative Evaluation is Rarely Conducted**
   Only three out of 21 informants do measure user performance quantitatively where a large number of users take part in the evaluation factor, see subchapter 4.1. In these cases the user performance measurements were conducted once or twice a year. Furthermore two informants used surveys with between 300 to 500 participants. None of the informants conducted quantitative analytical evaluation.

The second research question is: Does the evaluation methods usage vary among the different professional roles? Our results show that:

3) **Various Evaluation Methods Used by Different Roles**
   The usability experts use a combination of at least three or four different approaches to evaluate their systems, the interaction designers and Scrum managers all use three approaches except one interaction designer that does not evaluate at all. The business analysts mainly use one approach which is asking users about their opinion. One developer uses three approaches and one developer only evaluates analytically.

All the professional roles involve real users and all roles except business analysts involve user surrogates in some cases, see subchapter 4.1. The third research question is: What is the purpose of the evaluation? Our results show that:

4) **User Centred Evaluation is Conducted for Several Purposes**
   The informants evaluate prototypes, i.e. paper prototypes and sketches, to a very large extent in all the human centred activities to gather feedback from the users or user surrogates, see subchapter 4.1. Hence these evaluations relate and contribute to understanding and specifying the context of use, specifying the user requirements and to producing design solutions.

5) **User Centred Evaluation Often Conducted Early in the Project**
   We conclude from our interview study that user centred evaluation is often conducted early in the Scrum projects. Sometimes it is conducted in phases before the actual implementation starts to gain understanding for the UI design
of the software. The informants describe this as a successful approach, see subchapter 4.2.

In the following these main results will be discussed.

5. Discussion

This article focuses on analysing the usability professionals’ experience from conducting user centred evaluation in the Scrum development process. More specifically the starting point of the analysis have been the classifications of user centred evaluations presented in section 2.2, the usability professional roles described in subchapter 3.2 and the usability activities explained in the ISO 9241-210 (2010) standard. Hence, we have studied what methods are used, how they are used by the different usability roles and how the results from the evaluation influence the other three human-centred activities. In this chapter, we have chosen to discuss the more notable results from the study presented above in subchapter 4.3.

5.1. Various Empirical Qualitative Methods are Used

The usability professionals working in Scrum projects used various types of empirical qualitative evaluation methods, often done in an informal way including few users to give quick feedback to the development team. This is also done at the very core of the Scrum systems development process, as one of the major advantages of the process is perceived to be that it “shortens the feedback loop between customer and developer” (Schwaber 2004). However, it is noticeable that the majority of the professionals interviewed get feedback from users and acquire knowledge related to the human-centred activities in an informal manner without considering it as evaluation. Very few informants name the methods they are using, such as heuristic evaluation or the think-aloud method, when describing their evaluation. Their description is on how they do the evaluation and for what purpose, but not on what particular methods they use. They actually do not use the word “evaluation” but they talk about obtaining feedback on their work. They see evaluation as a formal and structured activity that needs much time for preparing and conducting the evaluation, and analysing and reporting the results. There is definitely a need for more low cost empirical methods that fit well in the Scrum process. Usability professionals could use the artefacts, which are already in Scrum to structure their evaluation, like suggested by Beyer et al. (2004). The descriptions of user stories could be used as a basis for the users’ tasks used for the empirical evaluation for example.

Some of the informants describe that usability activities are especially well suited in Scrum, because Scrum is a dynamic, informal process based on oral communication and teamwork rather than on written specifications and criteria for the system, as presented in subchapter 4.2. However, the informants describe that the
informal power structure that emerge due to the informal character of Scrum has consequences on how user centred evaluation is conducted. Informal evaluation methods have many advantages, as presented for example by Twidale (1993). In formal quantitative evaluation all participants perform the same tasks under as similar conditions as possible to gather reliable data. This type of evaluation is appropriate when the user goals and the tasks to achieve those are well known (Rohn et al. 2002). Furthermore, formal quantitative evaluation fits well, if the motivation for the evaluation is to measure the usability of the system in a systematic way. However, quantitative evaluation is limited in the ability to gather data on true user tasks, task flows, user profiles and context of use, especially if the user goals and tasks are not clearly defined. Furthermore, informal evaluation has the benefit that the users are relaxed and sometimes are able to explore the system on their own, they also have time to make interpretations on various levels. Additionally, informal evaluation leaves room for the discussing the user experience, reflections or intuitions about a design. To meet this need two new methods for qualitative evaluation were suggested recently by Riihiaho (2009). The main weakness of informal evaluation is that often only a few users are included, so the evaluator can’t be really sure that the results are reliable enough to base further work on. Further research is needed on what the actual benefits and weaknesses are when conducting qualitative evaluation in software development practice.

Our results are in contrast with what types of evaluation have been used in research studies according to Barkhuus and Rode (2007). In their study about 70% of the empirical studies from 2006 are quantitative and 30% qualitative. It is clear from these results that practitioners in Scrum projects and researchers in HCI do not choose the same types of evaluation when conducting empirical evaluation. We can conclude that simple observations with users participating and other informal empirical methods can give rich information for redesigning the software. This type of evaluation should be studied further in research studies. Actually, very few research studies have examined the effectiveness of evaluation methods during the last ten years, as confirmed by Barkhuus and Rode (2007) and Woolrych et al. (2011). Studying the evaluation types that are used in industry would be of great value for the usability professionals and for educational purposes.

5.2. Quantitative Evaluation is Rarely Conducted

When working on usability activities you need to have the “Big picture in place” like one of our informants described it, to be able to evaluate the usability. This is hard in Scrum because the deliveries after each sprint are minor. Therefore, evaluating the user performance in a quantitative way in each sprint is difficult because there are only slight changes after each sprint. Many of our informants describe that they do not have time to measure user performance because the sprints are short only 2 to 4 weeks and the emphasis is on delivering implemented software after that period to the customer. The user performance could be evaluated for example every three months, but there is no natural timing for it defined and often there is no time for it.
Others have concluded that conducting quantitative evaluation by measuring user performance is time consuming and scientifically rigorous and therefore it does not fit well in the agile spirit (Gould et al. 1988). This was also found in a study on Scrum projects in Iceland (Larusdottir et al. 2010), where the developers explained that they did not have time to do user performance measurements because of the short iterations in Scrum. Furthermore, one of the main hindrances for doing user centred evaluation generally in a recent study from Denmark (Bak et al. 2008), was resource demands. Furthermore, the most influential factor for choosing an evaluation was time constrains found in an international study (Monahan et al. 2008). The participants thought it is too time consuming to evaluate and it is expensive to get external consultants.

A study from Norway describes that over 90% of the participants perform empirical evaluation by conducting user testing (Bygstad et al. 2008) and in about 75% of the cases one to ten users are included in the testing, which could mean that the evaluation is rather informal. This is similar to our results that almost all of our informants carry out some type of empirical testing and much of the evaluation is informal. In the study from Norway, the results are not analysed according to if the evaluation is quantitative or qualitative, nor if it is formal or informal.

Two of our informants conduct quantitative empirical evaluation by using questionnaires. It is known that making questionnaires is time consuming (Lazar et al. 2009) but when the questionnaire is designed many users can be reached and analysing the results does not have to be time consuming. This could be one approach for practitioners to gather quantitative data with in their time constrains.

None of the informants conduct quantitative analytical evaluation. In a literature study from Barkhuus and Rode (2007), where evaluation types in research studies were examined, the results show that analytical evaluation is not used much in research project either.

5.3. Various Evaluation Methods Used by Different Roles

Our results show that the use of evaluation methods is different between the various usability professional roles. Usability experts and Scrum managers are the only roles that do quantitative evaluation and the business analysts mainly use one approach which is asking users their opinion. All the professional roles involve real users and all roles except business analysts involve user surrogates in some cases.

All roles except the business analysts use a combination of approaches to evaluate their software. The results from a study were a combination of analytical method, heuristic evaluation and empirical method, think-aloud evaluation show that combining these methods, did give more extensive results from the think aloud evaluation, than when only the think aloud method was used (Frökjær and Larusdottir 1999). However, one can also argue that finding usability problems alone does not make a better system. The real impact of the evaluation on the usability of the software is what really matters to the professionals (Uldall-Espersen et al. 2008, Law 2006).

Involving real users in evaluation has been recommended for over twenty years, e.g. by Grudin (1991). Our results show that almost all of our informants do so, even
though the evaluation is often informal and many of our informants use various ways of involving users. Some of our informants sometimes ask user surrogates to take part in the evaluation. This has been recommended for example by Lieversley and Yee (2007), if real users can not take part. There is no substitute for involving the actual users, like pointed out by Beyer (2010). The user surrogates sometime do not know exactly how the users’ tasks or the users’ needs are. Practitioners should be encouraged to ask real users to take part in evaluation whenever possible.

One can note that Scrum is a process made by developers for developers, where their needs and work are the main focus. This basic value of Scrum affects the integration of user centred evaluation in Scrum, as for example in one of the interviews where the informant describes “I interfere their work to help” (UE-1). One consequence of this focus on the system developers and their work could be that the persons having the various usability roles do not really see where they fit in the Scrum development process. Some researchers have suggested new roles in the Scrum process, for example Singh suggests having two product owners, one focusing on usability and the other on more conventional functions (Singh 2008). The responsibilities for the different usability roles need to be studied further in research studies.

5.4. Various Evaluation Methods Used for Different Purposes

Almost all the informants use qualitative empirical types of user centred evaluation. The informants evaluate prototypes, i.e. paper prototypes and sketches, to a very large extent in all the human centred activities to gather feedback from the users or user surrogates. Hence, these evaluations relate and contribute to understanding and specifying the context of use, specifying the user requirements and to producing design solutions. Using a variety of methods for evaluation was also examined in a study by Monahan, et al. (2008). Their results show that the qualitative methods, observations, interviews, contextual design and probes and diaries were all used for evaluation by practitioners to gain insight into the different phases of software development. This is similar to our results. This type of evaluation often gives rich insight into redesign alternatives, but because often only a few users are included in the evaluation the practitioners do not have as strong basis for their decisions as if they had used more formal ways of evaluating their software with more users.

The use of various methods was also examined in a study by Venturi et al. (2006). Qualitative evaluation was the conducted at least once during the software development by almost 70% of the participants, most frequently during the design phase, in nearly half of the cases. Quantitative evaluation was used in at least one phase by around 40% of the participants, most frequently in the design phase and in the testing phase. This study was conducted eight years ago and the results were not analysed according to the development process used. In our results the quantitative evaluation is not as common. This could indicate that especially in the agile development approaches evaluation is conducted more qualitatively than in other processes. It could also indicate
that the type of evaluation approaches used by practitioners has been changing during the last decade.

5.5. User Centred Evaluation Often Conducted Early in a Project

We conclude from our interview study that in the integration of user centred evaluation in Scrum the evaluation often moves to phases before the actual implementation starts. In this section we will discuss and interpret the reasons for this change and its possible consequences for the usability of the product.

Generally, requirements gathering and systems analysis is not included in agile methodologies (Beyer et al. 2004). During the requirements gathering or pre-study work described by our informants, the focus of evaluation at that stage is on obtaining an understanding of the whole product from a holistic perspective including for example context of use and an understanding of the user requirements. Several of our informants have described evaluation at this stage by saying that “the big picture is needed” or “the eagle view is important”. Other researchers have also concluded that when integrating agile development and human-centred activities the design team focuses on “Just-in-time design” where the usability professionals focus on a few new features at the time (Sy 2007). Due to this implication the usability professionals in our study prefer to conduct some human-centred activities before the project in order to address usability from a more holistic perspective. This is also the method used in the organisation described by Sy (2007) where “Sprint zero” (i.e. before the project begins) contains specified human-centred activities to understand the context of use for example. This has also been recommended by other researchers, for example by Williams and Ferguson (2007) and Detweiler (2007). They recommend that activities related to UI design should be performed before the actual implementation starts.

Noticeable is that a consequence of this move of human-centred activities is that these activities become less iterative in nature, and that even though the systems development as such is produced iteratively in sprints the human-centred activities often occur at the very beginning as in a waterfall model, like concluded by Sy (2007). The results from an interview study with interaction designers and developers show that conducting the user interface design before the actual implementation starts is commonplace in agile development (Ferreira et al. 2007). Our results also concur with that of Chamberlain et al. (2006) as they also conclude that some of the design is already set before the actual Scrum project starts.

In the agile development it is vital that user centred activities are also conducted through the whole software development process, because the user interface design can change much during implementation. Processes for integrating the design, implementation and evaluation of the user interface into the Scrum process has been suggested by Sy (2007) and most recently by Silva da Silva et al. (2011). Both these processes suggest that the user interface design is done one sprint in advance the implementation and evaluation is done one sprint after the implementation. This way, the usability professionals get longer period to work on the various tasks and for involving users in the development.
5.6. Lessons Learned

This study has revealed several important aspects of the use of user centred evaluation in Scrum projects. To conclude we describe the lessons learned from the study and state our recommendations to usability professionals and researchers in the field based on the results so far:

(1) Our informants use many informal ways to include users in their evaluation. Much of the evaluation is qualitative. Usability professionals should be encouraged to conduct this type of evaluation whenever possible to maintain their knowledge of the user aspects. Conducting qualitative empirical evaluation whenever possible has also been suggested by Illmense and Muff (1999). User testing whenever possible, has been recommended by Najafi and Toyoshiba (2008) and Wolkersorfer et al. (2008).

(2) Usability professionals should try to find new ways of including real users in their evaluation. Even though they can only include a few users because of time constrains they should do so, because some information from the users is better than no information.

(3) The majority of our informants use various methods each to conduct user centred evaluation in a successful way. We recommend that usability professionals evaluate both by obtaining feedback from users by using qualitative and quantitative evaluation depending on the resources available and through conducting analytical evaluation. By using various ways of evaluating the software the practitioners gain a better overview of the users’ perspective. This has been suggested by others, for example by Beyer et al. (2004).

(4) Usability professionals should be encouraged conduct user centred evaluation to gain insight into the three major phases of user centred design, the context of use, the user requirements as well as the design of the software. Many of our informants describe doing evaluations before the actual implementation starts as being successful. This has also been recommended by other researchers, for example by Sy (2007), Williams and Ferguson (2007) and Detweiler (2007). They recommend that activities related to UI design should be performed before the actual implementation starts.

(5) Researchers should be encouraged to study the use of empirical qualitative evaluation methods in the Scrum projects so this type of evaluation will be easier for the usability professionals. These methods are often used by our informants but have not been studied much by researchers, so it is hard for the usability professionals to know how effective these methods are.

(6) Researchers should be encouraged to describe new ways of including empirical qualitative evaluation in the Scrum projects so this type of evaluation will be more effective for the usability professionals. The new methods could for
example be based on the existing artefacts in Scrum, like suggesting a way to
describe user tasks for user observation based on the user stories in Scrum.

To be able to increase the focus on usability in Scrum projects in practice the above
mentioned lessons learned provide important insight into the use of user centred
evaluation in practice. These lessoned learned can help usability professionals that want
to improve the usability of the end result. Furthermore these insights could guide
researchers to focus their studies on qualitative evaluation methods in agile projects to
examine the effect each method has, how the methods can be improved and what new
methods could be suggested.

The study of user centred activities in the Scrum development practice is an
interesting area as it has the potential to impact software development practice and the
usability of the systems built. Therefore, we aim at further studying how the Scrum
software development process could be tailored to better incorporate usability activities.
More specifically, we aim at studying how the understanding of the context of use, the
user requirements and the design of the user interface are conducted in the Scrum
development practice in accordance with the ISO 9241-210 (2010).

6. Acknowledgements

We would like to thank COST Action IC0904 called Twintide for the financial support
that they have provided. Also, we would like to thank all informants in the study that
took their time to participate in the study. Moreover, we would like to thank associate
professor Anders Jansson at the department of information technology at Uppsala
University in Sweden, Christina Dörge at the Carl von Ossietzky University in
Oldenburg in Germany, adjunct professor Niklas Hallberg and Elina Eriksson at the
Royal Institute of Technology in Sweden and Gudmundur Valsson, project manager at
Sudurorka in Iceland, for valuable comments.
7. References


Sutherland, J., 1995. Business object design and implementation workshop. In: *Addendum to the proceedings of the 10th annual conference on Object-oriented programming systems, languages, and applications (OOPSLA)*. Austin, Texas, United States: ACM Press, 170-175.
A. Appendix

Table 4. Usability experts’ background information

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender, age</th>
<th>Experience from work in industry</th>
<th>Organisational context</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>UE-1</td>
<td>Male, 34</td>
<td>3 years</td>
<td>Product development for a focused user group. About 30 people employed</td>
<td>PhD in HCI</td>
</tr>
<tr>
<td>UE-2</td>
<td>Male, 46</td>
<td>12 years</td>
<td>Product development and consultant organisation. 6000 employees worldwide</td>
<td>PhD in HCI</td>
</tr>
<tr>
<td>UE-3</td>
<td>Male, 34</td>
<td>10 years</td>
<td>Product development company with some consultant services, with over 100 employees.</td>
<td>Master in computer engineering and cognitive science</td>
</tr>
<tr>
<td>UE-4</td>
<td>Male, 32</td>
<td>3 years</td>
<td>Product development for a focused user group. 2000 employees worldwide</td>
<td>Licentiate in Human Computer Interaction</td>
</tr>
<tr>
<td>UE-5</td>
<td>Female, 33</td>
<td>4 years</td>
<td>Product developers, 170 employees worldwide of whom 30 in Stockholm</td>
<td>Master Degree in Media Technology Science</td>
</tr>
</tbody>
</table>

Table 5. Interaction designers’ background information

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender, age</th>
<th>Experience from work in industry</th>
<th>Organisational context</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID – 1</td>
<td>Female , 31</td>
<td>2 years</td>
<td>Consultant organisation. 32 people employed.</td>
<td>Master of Computer Science with HCI emphasis</td>
</tr>
<tr>
<td>ID-2</td>
<td>Male, 34</td>
<td>5 years</td>
<td>Consultant. 8 employed.</td>
<td>Master of Computer Science, PHD studies in HCI</td>
</tr>
<tr>
<td>ID-3</td>
<td>Male, 35</td>
<td>11 years</td>
<td>Consultant organisation, product development. 11 employees</td>
<td>Bachelor in Cognitive Science</td>
</tr>
<tr>
<td>ID-4</td>
<td>Male, about 33</td>
<td>3 years</td>
<td>Product development for a focused user group. 2000 employees.</td>
<td>Master in Media Technology</td>
</tr>
<tr>
<td>ID-5</td>
<td>Male, 55</td>
<td>30 years</td>
<td>Different employers</td>
<td>University studies</td>
</tr>
<tr>
<td>ID-6</td>
<td>Male, 47</td>
<td>15 year</td>
<td>Company specialising in content management, 200 employees in Sweden</td>
<td>Private design degree</td>
</tr>
<tr>
<td>ID-7</td>
<td>Male, 50</td>
<td>11 year</td>
<td>Business systems and consultant services, 170 employees</td>
<td>Master degree in interaction design</td>
</tr>
</tbody>
</table>
Table 6. Business architects’ background information

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender, age</th>
<th>Experience from work in industry</th>
<th>Organisational context</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA-1</td>
<td>Female, 41</td>
<td>about 15 years</td>
<td>Consultant organisation. Management and development. 700 employed.</td>
<td>Bachelor in Computer Science and Human Computer Interaction</td>
</tr>
<tr>
<td>BA-2</td>
<td>Female about 40</td>
<td>12 years</td>
<td>Consultant organisation. Management and development. 700 employed.</td>
<td>Bachelor degree in Computer Science Psychology and pedagogic</td>
</tr>
<tr>
<td>BA-3</td>
<td>Female, 42</td>
<td>13 years</td>
<td>Consultant organisation. Management and development. 700 employed.</td>
<td>Master in Computer Science</td>
</tr>
<tr>
<td>BA-4</td>
<td>Male, about 40</td>
<td>About 20 years</td>
<td>Consultant organisation. 32 employed. System development, application security, integration, usability and project management.</td>
<td>Missing information on this</td>
</tr>
</tbody>
</table>

Table 7. Developers’ background information

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender, age</th>
<th>Experience from work in industry</th>
<th>Organisational context</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE-1</td>
<td>Male, 30</td>
<td>5 years</td>
<td>Consultant organisation. 800 employed</td>
<td>Master in Computer Science</td>
</tr>
<tr>
<td>DE-2</td>
<td>Female, 29</td>
<td>1 year</td>
<td>Business systems and consultant services. 170 employees</td>
<td>Master in Computer Science with courses in HCI</td>
</tr>
</tbody>
</table>

Table 8. Scrum Managers’ background information

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender, age</th>
<th>Experience from work in industry</th>
<th>Organisational context</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM-1</td>
<td>Male, 33</td>
<td>3 years</td>
<td>Product development for a focused user group. 2000 employees of which 400 work in Stockholm</td>
<td>Master in Media Technology</td>
</tr>
<tr>
<td>SM-2</td>
<td>Male, ca 30 years</td>
<td>3 years</td>
<td>Product development for focused user group. 12,500 employees, of which up to 200 work on the same project as the informant</td>
<td>PhD in Human Computer Interaction, 2008</td>
</tr>
<tr>
<td>SM-3</td>
<td>Female, 32 years</td>
<td>8 years</td>
<td>Worldwide international company with 120 employees</td>
<td>Master degree in CS</td>
</tr>
</tbody>
</table>
### B. Appendix

Table 9: Overview of type of evaluation used by each informant in the study

<table>
<thead>
<tr>
<th>Type of evaluation</th>
<th>Evaluation method</th>
<th>Empirical Quantitative Evaluation</th>
<th>Empirical Qualitative Evaluation</th>
<th>Analytical Qualitative Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Measuring users' performance and survey</td>
<td>Observing Users</td>
<td>Asking user their opinions</td>
</tr>
<tr>
<td>UE-1</td>
<td></td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>UE-2</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>UE-3</td>
<td></td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>UE-4</td>
<td></td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>UE-5</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Usability Experts</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>ID-1</td>
<td></td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ID-2</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ID-3</td>
<td></td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>ID-4</td>
<td></td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ID-5</td>
<td></td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ID-6</td>
<td></td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ID-7</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Interaction Designers</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>BA-1</td>
<td></td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>BA-2</td>
<td></td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BA-3</td>
<td></td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>BA-4</td>
<td></td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Business Analysts</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>DE-1</td>
<td></td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>DE-2</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Developers</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SM-1</td>
<td></td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>SM-2</td>
<td></td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>SM-3</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Scrum Managers</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>4</td>
<td>13</td>
<td>18</td>
</tr>
</tbody>
</table>