Available online at www.sciencedirect.com
SCIENCE DIRECT®
Interacting with Computers 16 (2004) 1095–1132
Interacting with Computers
www.elsevier.com/locate/intcom

Evaluating the user-centredness of development organisations: conclusions and implications from empirical usability capability maturity assessments

Timo Jokela*

Department of Information Processing Science, P.O. Box 3000, 90014, University of Oulu, Finland

Received 23 March 2004; revised 11 June 2004; accepted 6 July 2004
Available online 10 August 2004

Abstract

Improving the position and effectiveness of user-centred design, UCD, in software and product development is a challenge in many companies. One step towards improvements is to carry out a usability capability maturity, UCM, assessment to evaluate the strengths and weaknesses of a development organisation in user-centred design. This article reports the lessons learnt from 11 empirical UCM assessments of R&D groups of Nokia, a software house, an SME, and a research institute in Finland. The first assessments were carried out using a standard process assessment model (a pre-version of ISO 18529); the last assessments were carried out using a new KESSU model that evolved during the research. It was found that the assessment model, its interpretation, and the viewpoints of the assessment team have a critical role in the success of assessments. In addition, it was found that the customers have different purposes for assessments and those purposes have an effect on how one should conduct the assessment—or whether to conduct it at all.

Keywords: User-centred design; Human-centred design; Process assessment; Usability maturity; Usability capability

1. Executive summary

The improvement of the position of *user-centred design, UCD*, in software and product development has been widely recognised as a challenge. A logical first step in the process

* Tel.: +358-40-5118250; fax: +358-8-5531890.
E-mail address: timo.jokela@oulu.fi.

0953-5438/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.intcom.2004.07.006

of making organisational improvements is to carry out a *current state analysis* of the development practices of a company. A current state analysis reveals the strengths and weaknesses in the organisational practices, and thereby forms a good basis for planning and implementing the action for organisational improvement. Organisational current state analyses (models such as CMM and ISO 15504) are widely used for providing the basis for software process improvement.

Several *usability capability maturity* (*UCM*) models have been presented since the early 1990s for the purpose of providing a basis for the improvement of UCD practices in development organisations. New UCM models have lately been introduced by ISO (International Standardisation Organisation) and in several countries such as Germany, Japan, and Finland. The Usability Roundtable (2004) in the US has produced a white paper on the topic. However, the literature on UCM assessments mainly focuses on presenting the different individual assessment models. Empirical research results on how to conduct UCM assessments are not widely reported.

The aim of this research is to fill this gap, by reporting the conclusions and lessons learnt from the analysis of 11 empirical assessments. The first assessment was carried out in August 1997. At that time the author—who was later involved in all the 11 assessments—worked as a usability practitioner in an industrial setting where he perceived the organisational position of UCD as difficult, because of unprofessional interventions in usability work, the use of the term 'usability' as a buzzword for political reasons, etc. The author initiated the assessment with the aim of improving that situation. The first assessment was soon followed by another assessment at a software house in November 1997. The third assessment took place two and half years later, in spring 2000, when a new national research project was set up in Finland. Further assessments followed, ending with the 11th assessment being conducted in autumn 2003.

The cases were conducted in different organisations: R&D groups of Nokia, a software house, an SME, and a research institute. We used two different main assessment approaches. In cases #1–#4 a standard process assessment was used (the assessment model was a pre-version of ISO 18529), and in cases #5–#11 we used the KESSU model developed ourselves, based on the experience gained from using the standard model. Although all of the cases had some pros and cons, cases #2, #6, #7, #10 and #11 can be regarded as successful and cases #1, #3, #4 as less successful. Case #5 was somewhere in between. If assessments #8 and #9 had been completed, they probably would not have been very successful.

The findings indicated many different—and even contradictory—phenomena in our cases. The expectations of the customers varied. For example, in cases #1 and #10 the customer was disappointed because the assessment did not produce 'bad' enough ratings. On the other hand, in case #4 the customer was disappointed because the results were not as 'good' as expected, and in cases #8 and #9, the results were not presented at all because they were so 'bad'. Moreover, it was discovered that a specific assessment model worked in one case, but did not work in another case.

To explain the contradictory phenomena, it was concluded that they were due to the fact that the customers had different *purposes* for the assessments. For example, if an assessment is used to 'awaken' the organisation, the assessment should focus on showing the organisational obstacles to UCD (and show the 'bad' results). For the *kick-off* of

a usability improvement program, an assessment should be a constructive and sense-making training occasion and the reporting of poor assessment results should be avoided. Carrying out an assessment for the purpose of *objective curiosity* is a more neutral setting; and in that scenario the results can be reported 'as they are'.

A number of lessons applicable to all assessments were also identified. An assessment is inherently an organisational intervention which can be an effective training opportunity for usability and UCD. The contents of the assessment model and its interpretation by the assessment team are critical. For a usability practitioner, being involved in assessments presents a very useful exercise which challenges his or her understanding and knowledge of UCD.

2. Introduction

2.1. What is user-centered design

Usability has been recognised as one of the important quality characteristics of software systems and products. Usability is recognised as one of the most important quality characteristics of software intensive systems and products. Usable systems are easy to learn, efficient to use, not error-prone, and satisfactory in use (Nielsen, 1993). Usability brings many benefits, which include 'increased productivity, enhanced quality of work, improved user satisfaction, reductions in support and training costs and improved user satisfaction' (ISO/IEC, 1999).

Usability is defined in ISO 9241-11 (ISO/IEC, 1998b) as follows: '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 definition emphasises how the usability of a product relates to its context of use. The definition is largely used in the HCI community. For example, it is used as a reference for usability in the Common Industry Format (CIF) for usability testing (ANSI, 2001).

To guide the development of usable products and software systems, *user-centred design*, *UCD* approaches and methodologies have been proposed. The standard ISO 13407 (ISO/IEC, 1999) is a widely used general reference of UCD, and is an important reference also in this study.

ISO 13407 identifies four principles of UCD: user involvement, iterative design, multi-disciplinary teamwork, and appropriate allocation of functions between users and the system. The standard further identifies four main activities of UCD, shown in Fig. 1. These activities represent a general overview of a user-centred development process: analysing users and the context of use, determining user-driven requirements, producing designs and evaluating the usability of the designs.

2.2. History of user-centered design

The book *User-Centered System Design* by Donald Norman and Stephen Draper (1986) is a pioneering work on UCD. In addition, there were also others who worked on the topic in the 1980s. John Gould and his colleagues worked with usability methodologies;

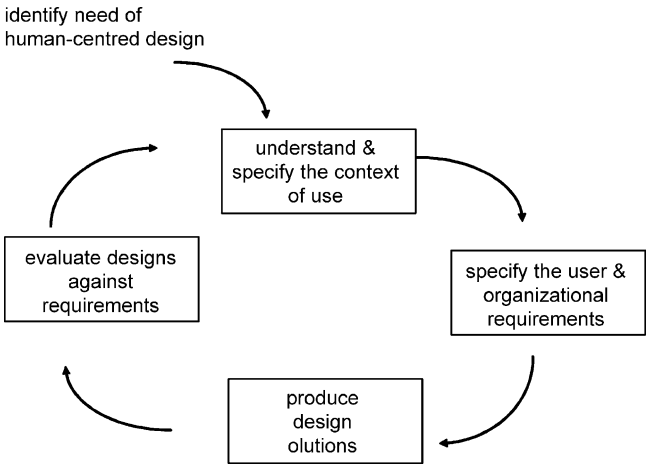

Fig. 1. The four usability engineering processes of ISO 13407.

their well-known work was the 1984 Olympic Message System (Gould et al., 1987). Dennis Wixon and Karen Holtzblatt at Digital Equipment developed Contextual Inquiry and later on Contextual Design. Also Carroll and Mack (1985) were early contributors.

Later, in the 1990s, Myers et al. (1996) summarised the past and visioned future directions on HCI. Different UCD methodologies were proposed, e.g. by Nielsen (1993), Hix and Hartson (1993), Beyer and Holtzblatt (1998), Mayhew (1999), Constantine and Lockwood (1999) and Rosson and Carroll (2002). Although there are some variations in the methodologies, they identify more or less the same basic activities and principles of UCD as does ISO 13407.

2.3. Assessment of user-centred design

Many products or systems with usability problems reveal that the position of UCD is not effective in many organisations. Thus, the improvement of the position of UCD has been widely recognised as a challenge in practice and in the literature: Axtell et al. (1997), Wilson et al. (1997), Bloomer and Wolf (1999), Rosenbaum et al. (1999, 2000), Anderson (2000), Bevan and Earthy (2001) and Vredenburg et al. (2002). The integration of user-centred design and software engineering has been a topic of a number of recent workshops, e.g. John et al. (2004) and Harning and Vanderdonckt (2003).

A logical first step in the process of making organisational improvements is to carry out a *current state analysis* of the development practices of a company. A current state analysis reveals the strengths and weaknesses in the organisational practices, and thereby forms a good basis for planning and implementing the action for organisational improvement. These kinds of organisational current state analyses are often called *capability maturity assessments*. Models for capability maturity assessments—such as the *Capability Maturity Model*, *CMM*, (Paulk et al., 1995) and *ISO 15504* (ISO/IEC, 1998a)—have been developed to provide a framework for guiding the assessment process in the field of software engineering.

Several *UCM models*[1] have been presented from the early 1990s. Following the capability maturity model trend in software engineering, the UCM models are aimed at assessing the status of user-centred design in a product or system development organisation. The first UCM models are *Trillium* (Coallier et al., 1994) by Bell Canada (a general assessment model including a specific part for usability engineering), *Usability Leadership Maturity Model*, *UMML*,[2] (Flanagan, 1995) by IBM, *HumanWare Process Assessment*, *HPA*, (Taylor et al., 1998) by Philips, and *User Centred Design Maturity*, *UCDM*, (Eason and Harker, 1997) by the Loughborough University. In the late 1990s, *Usability Maturity Model: Processes*, *UMM-P*, (Earthy, 1997)—which follows the format of software process assessment (ISO 15504)—and *Usability Maturity Model: Human-Centredness Scale; UMM-HCS* (Earthy, 1998b) were developed in a European research project INUSE. Later, a technical report *ISO TR 18529* (ISO/IEC, 2000) was produced from the basis of UMM-P. The latest developments are *ISO 18152* (ISO/IEC, 2000)—a more detailed version of ISO TR 18529-, *DATech* (2002) in Germany, *SDOS* (Kurosu et al., 2000) in Japan, and *KESSU* (Jokela, 2001) in Finland. A more detailed overview of the various UCM models is given in Jokela (2001).

Each UCM model defines—explicitly or implicitly—a *UCD reference model*[3] that can be regarded as an ideal model of UCD, i.e. 'this is what UCD should ideally be'. The principal idea in a UCM assessment is that the existing development practices of an organisation are analysed and mapped against the ideal UCD reference model (Fig. 2). The closer the organisation meets the requirements of the UCD reference model, the higher its ratings.

An assessment may serve for two main uses. One is to provide *a basis for improvement action*. The usefulness of a UCM assessment is that the results would help a practitioner to identify and prioritise areas of improvement in UCD. Typically, improvement actions are focused on those organisational areas with the lowest ratings. The assessment results may also identify the necessary strengths that need to be protected.

Another use of UCM assessments could be *third party certification*.[4] Through a UCM assessment, a purchaser organisation could gain an understanding of the capability of a supplier organisation in developing usable products. In other words, the main customer of the assessment is the customer of the company to be assessed, not the development organisation.

The area of our research was solely the first one, i.e. assessments as a basis for improvement action. To the author's knowledge, UCM assessments for third party certification are not widely carried out, if at all.

2.4. Motivation for the research

The literature on UCM assessments mainly focuses on presenting the different individual assessment models (Jokela, 2001). Earthy et al. (2001) provide an important

[1] Other terms, such as usability maturity model, UMM, are also used in literature.
[2] The abbreviations of the models used in this article are by the author.
[3] 'UCD reference model' is the author's term, and not used systematically by others.
[4] The original motivation for the development of CMM in software engineering.

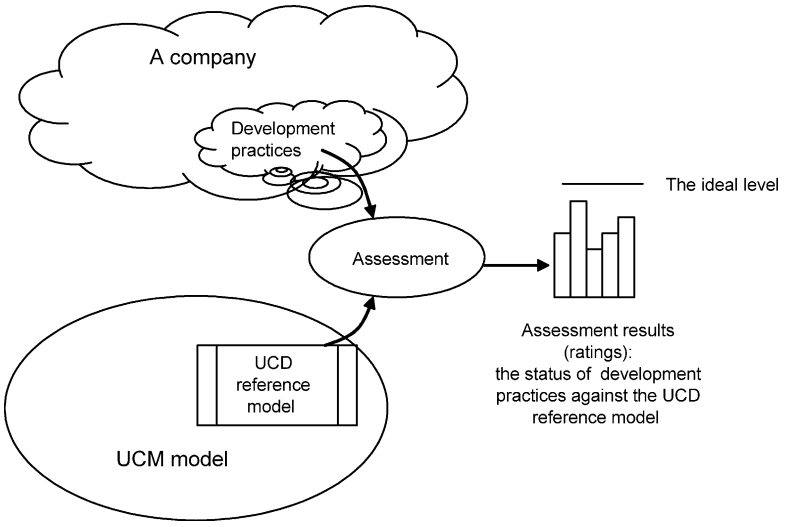

Fig. 2. Development practices are analyzed against a UCD reference model.

overview of assurance of usability. They identify six approaches: assessment of (1) product attributes, (2) user performance and satisfaction, (3) process certification, (4) organisational human-centredness, (5) technical competence, and (6) process capability. They specifically emphasise the importance of the last category, i.e. process capability assessments[5] (6), which is also the main focus of the research reported in this article.

Empirical research results on UCM assessments are not widely reported. IBM reports that they conducted a large number of assessments (Flanaghan, 1996) but they do not report any experiences or other research findings from the assessments. Bevan and Earthy (2001) report the experiences from a European TRUMP project where they used the ISO TR 18529 model in two organisations. They carried out an assessment in the beginning and in the end of the project, and found that improvements in UCD had taken place in both organisations. They conclude that ISO TR 18529-based assessments are generally useful ('the results were judged to be highly beneficial') but they also identified some challenges (the customer had an 'initial difficulty in understanding the assessment model').

In this research, 11 usability capability assessments in different development organisations were carried out. The initial motivation for the research was practical. In 1997, when the first assessment was conducted, the author—who was later involved in all the 11 assessments—worked as a usability practitioner in an industrial setting where he perceived the organisational position of UCD as difficult because of unprofessional interventions in usability work, the use of the term 'usability' as a buzzword for political reasons, etc. He initiated the assessment with the aim of improving the situation. The idea for carrying out assessments originated from the author's earlier knowledge of

[5] A more general term 'UCM assessment' is used in this article.

assessments. The author had formal training in software process assessment and had been involved in software process assessments. Further, the author had acquired basic knowledge about UCM assessments. He had attended two special interest group sessions on usability maturity models run by representatives of IBM at the CHI'95 conference (Flanagan and Rauch, 1995), and learnt about the assessment models developed in the European INUSE project.

The first assessment was soon followed by another assessment. The third assessment took place two and a half years later, in spring 2000, when a new national research project was set up in Finland. Further assessments followed so that the 11th one was conducted in autumn 2003. In the first assessments (#1–#4), we used very much the same UCM model as Bevan & Earthy, i.e. the different versions of the UMM-P model (UMM-P was the pre-version of ISO TR 18529). Later (assessments #5–#11), we used a different UCM model KESSU[6] developed by ourselves, in order to overcome interpretation problems that we met with UMM-P and to fit better our specific assessment contexts.

The purpose of this paper is to report the conclusions and lessons learnt from our assessments. Our trials complement the work of Bevan and Earthy by reporting experiences from several cases where the assessments were carried out by different people, in different contexts, and using partly different UCM models. Furthermore, the focus of our research was limited to just learning about conducting assessments. In other words, we tried to learn what is important in assessments, and how they are perceived by the different stakeholders involved in the assessment process. We did not focus on following whether subsequent improvements took place in the organisations or not, although we followed the improvements to some extent.

This study has merit for several reasons. There is a recent and growing interest in UCM models. The recent developments in Germany (DATech, 2002) and Japan (Kurosu et al., 2000) and the initiative of the Usability Roundtable (2004) in the USA indicate and support this interest and popularity. In this article, an analysis of several trials and empirical experience of conducting assessments are presented, about which there is very little literature to date. This is of relevance for practitioners and researchers alike. The lessons learnt and the implications from the assessment should be useful for practitioners who either wish to conduct an assessment themselves, or wish to have an assessment conducted. From the research perspective, developments in UCM models can take place only when one has a good understanding of the relevant issues around practical assessments. Thus, this research provides a platform for further research into UCM models.

In Section 3 of the article, the basics of the UCM models and the UCM assessments are introduced. In Section 4, the research method is described. Descriptions of the 11 case studies are presented in Section 5. The conclusions on the different phenomena around the assessments, the main results of the study, are summarised in Section 6. In Section 7, the limitations of the research as well as the implications of this research for practitioners and researchers are discussed.

[6] The name of the research project.

3. On UCD and UCM assessments

To understand the results of this research, an introduction is given on some basic issues related to UCD assessments. A UCM assessment is about analysing the status of UCD in a development organisation. In Section 3.1, I first describe what kind of organisational areas potentially could be examined as part of 'the status of UCD'. Then the main concepts of assessments and UCM models are presented.

3.1. Areas of user-centred design capability

A natural basic focus of a UCM assessment is to examine the UCD practices that are carried out in a development project: if there is any effective UCD in an organisation, it should be visible in development projects. In practice, this means an examination of the extent, quality and effectiveness to which *usability engineering*[7] activities (user analysis, task analysis, usability requirements determination, usability evaluations, etc.) are carried out in a specific individual development project (shown as 'a' in Fig. 3). High performance means that usability engineering activities are extensively carried out to a high quality, and the results are truly taken into account in the development project.

There is, however, a set of other organisational areas related to UCD which could be examined in an assessment, too:

- ☐ The *management* of user-centred design in a specific development project ('b' in Fig. 3) means the examination of issues such as the inclusion of UCD activities in a project plan, follow-up of the implementation of the plan during the project, configuration management of the documents produced, etc. It should be noted that high performance of usability engineering does not necessarily require a good level of management. Effective UCD work can be carried out without being very well planned or controlled.
- ☐ The *UCD infrastructure* ('c' in Fig. 3) means those UCD resources that any development project can utilise when planning and carrying out UCD activities. Typically these include the procedures, templates, staff training programs, etc. which are documented as a part of the quality process system in the company.
- ☐ There are a number of other foci ('d' in Fig. 3) that also could be examined in the assessments and which are identified by some UCM models: UCD skills, management support, awareness, attitude, organisational culture, etc.

3.2. On UCM models

The term UCD model is used as a high-level concept, denoting a specific assessment approach such as Trillium, ULMM, ISO TR 18529, etc. A UCM model includes two sub models. In addition to the UCD reference model (see discussion in Section 2.3), UCM

[7] I use *UCD* as a high level concept and *usability engineering* to denote the specific UCD engineering activities (user analysis, usability requirements determination, etc.) in a development project.

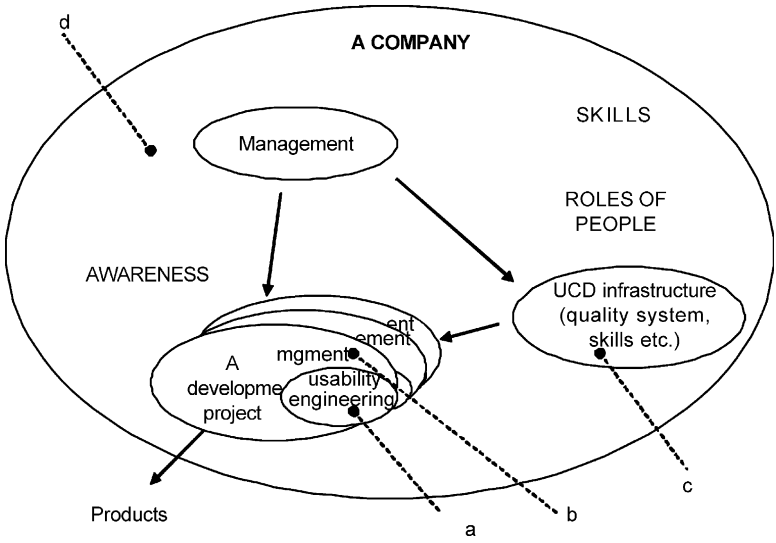

Fig. 3. Different organisational areas that have impact on the effectiveness of UCD.

models include an *assessment method* for guiding the process of conducting an assessment.

Most UCD reference models identify several different organisational areas, each of which is investigated in an assessment and each of which is rated separately (see the illustration of five ratings in Fig. 2). For example, a model may identify a set of UCD processes (such as context of use analysis, usability requirements determination, and usability evaluation). There are, however, differences between the different UCD reference models. The contents as well as the number of organisational areas vary from one model to another.

The UCD reference model defines the boundaries of an assessment: i.e. what kinds of organisational areas can be the focus of an assessment. The focus of a specific UCM reference model may be wide, and include the examination of areas such as organisational awareness and skills in UCD, the inclusion of UCD in the quality systems of a company, and the performance of UCD in individual development projects. For example, the UMML model (Flanagan, 1995) represents a wide focus and includes the examination of the organisational awareness about usability UCD, management support towards UCD and the skills of UCD in the organisation ('a–d' in Fig. 3). Process assessment (such as UMM-P) examines the performance and management of UCD in development projects as well as the incorporation of UCD in the quality system of the company ('a–c' in Fig. 3). An example of a narrow focus is KESSU which examines the performance of UCD in individual development projects only, excluding the management aspects, ('a' in Fig. 3). The DATech model (DATech, 2002) has a specific focus, among others: it examines the position of the usability persons in the organisational structure of a development project.

One should note that the focus of an assessment may be—and often is—more limited than the focus of the UCD reference model. In other words, all of those organisational areas defined by the UCD model are not necessarily examined in a specific assessment.

3.3. Standard process assessment

The scope of this paper is not to describe the various assessment (UCM) models in detail. However, some basics of the assessment models are introduced—especially of the standard process assessment ISO 15504 that was used in the first cases—so that a reader can understand our different assessment cases and the research results. It must be stated that process assessment is a rather complex area, and only some of the basics are covered. For complementary discussion of process assessment, see e.g. Earthy et al. (2001).

3.3.1. Process definitions

ISO 15504 as such does not define the software processes; it defines the format by which processes should be defined. There is actually another standard (ISO 12207) that provides the definitions of software processes. The format of ISO 15504 makes it possible to define processes also for other disciplines—such as user-centred design.

The UMM-P model defines UCD processes within the format of ISO 15504. In our first assessments we used different versions of UMM-P (Earthy, 1996, 1997, 1998a) as the UCD reference model. Later, the process definitions of the latest version of UMM-P were published as a technical report ISO TR 18529 (ISO/IEC, 2000). A more comprehensive set of UCD processes—more than 20—was later published in the technical report ISO TR 18151 (ISO/IEC, 2003).

The ISO 15504 format defines processes through *purpose statements*. As an example of the definition of a UCD process, ISO TR 18529 identifies the process 'Specify the user and organisational requirements' and defines the process as follows: 'The purpose of the process is to establish the requirements of the organisation and other interested parties for the system. This process takes full account of the needs, competencies and working environment of each relevant stakeholder in the system'.

3.3.2. Capability levels

An ISO 15504 style process assessment identifies six capability levels, Table 1. In an assessment, one first examines whether a process reaches level 1 of the capability levels. This means examining the extent to which a process—carried out in a selected

Table 1
Capability levels of the ISO 15504 model

Level of capability	Description
Level 5: optimising	The organisation can reliably tailor the process to particular requirements
Level 4: predictable	The performance of the process is within predicted resource and quality limits
Level 3: established	The process is carried out in a manner specified by the organisation and the resources are defined
Level 2: managed	The quality, time and resource requirements for the process are known and controlled
Level 1: performed	The process achieves its purpose. Individuals carry out processes
Level 0: incomplete	Organisation is not able to carry out the process

Each process gets an individual rating in the scale Level 0 (lowest) to Level 5 (highest) in an assessment.

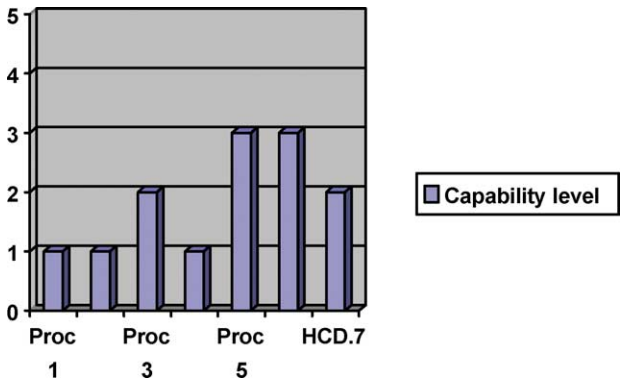

Fig. 4. The capability profile of seven processes. The higher the rating, the better capability (see Table 1) of a process.

development project—achieved its purpose (as defined in the process definitions, see above). If the process is found to have achieved its purpose, then capability level 1 ('performed') is reached. For example, capability level 1 of the process *Specify the user and organisational requirements* should be based on the mapping of the UCD practices in a development project against the purpose statement mentioned above.[8]

If level 1 capability is reached, the next level ('2') is to examine to which extent the process was managed in that specific development project: to which extent the process was planned and its progress monitored, the extent to which the documents were under configuration management, etc. If an adequate level of management of a process within a development project is achieved, level 2 capability is reached. At level 3 capability, UCD is carried out in individual projects in a manner specified by the organisation, e.g. in a quality system. Levels 4 and 5 mean more advanced organisational capability to predict and optimise the processes, see Table 1.

One relevant aspect to remember is that the 'basic focus', the performance of (UCD) processes, is examined at capability level 1. All the examinations from levels 2 to 5 are about process management issues, not the substance of the usability engineering processes. Level 1 is about examining the process definitions, levels 2–5 are assessed against the general criteria—defined in ISO 15504—that are not process specific.

The result of process assessment is a *capability profile*, shown in Fig. 4. Each process that is assessed is rated by a capability rating: the higher the rate, the higher the capability of the specific process.

3.3.3. Implementing assessments

An assessment is carried out by an *assessment team* that consists of a *lead assessor* and *assistant assessors*. The lead assessor manages the assessment process, takes charge of conflicting situations, e.g. when there are disagreements among the assessment team on the results—and makes the key decisions.

[8] The full definition include further details.

The process of assessment starts by planning the assessment together with the key stakeholders of the customer organisation. The assessment process typically starts with an opening briefing to the organisation. Then the assessors gather data by examining available documentation and interviewing the different stakeholders in the company. The assessment team interprets the findings, generates the results and presents the results to the customer.

4. Research method

The main context of the assessments was two national research projects[9] in Finland between 1997 and 2003. The main objective of the assessments was to provide a basis for improvement actions in UCD in the companies—i.e. the assessments were not performed primarily for the reason of conducting research into UCM assessments. All the 11 assessments—apart from the last one—were conducted in companies on real development projects; and even the last one—where an experimental project of a research organisation was examined—had aims for improvement action in the long run.

The research approach that we used was partly constructive research, partly interpretive research. The constructive part of the research was to develop a new assessment model, which is discussed in other papers, e.g. Jokela (2001). The interpretive part of the research—the topic of this article—was about understanding the phenomena around assessments: how the assessment models worked and how the different stakeholders perceived the assessments.

The main source of data arose from the observations of the author and other assessors gathered in the different assessment cases, and in different situations within the cases. We also gathered data on how other stakeholders—i.e. the development staff, management, and usability specialists of the organisations—perceived the assessments during and after the different types of interventions (briefings, interviews, workshops and result reporting sessions). All assessments were carried out to provide a basis for improvements (vs. third party certification). Therefore, we found it important to gather feedback from the organisation assessed, which probably had not been so important in assessments for third party certification.

The details of how the data was gathered varied from case to case. In all cases, data was gathered by observations and notes from the researchers, i.e. by those who conducted the assessments. The number of researchers varied from seven (some sessions in cases #3 and #4) to one (cases #10 and #11). After various assessment sessions, the assessment team gathered together to explore and discuss their findings. We drew conclusions based on our observations, and decided what changes should be made to the assessment approach for the next assessment.

Feedback from the companies was gathered by questionnaires, interviews and email. A large questionnaire with detailed questions was used in case #3 but in most cases a simple feedback form (i.e. ‘list the pros and cons’) was used. The number of answers to

[9] Called *KÄYPRO*, *KESSU*.

the questionnaires varied depending on the number of participants, from a one or two people to approximately 30 persons in case #3. In addition, post-assessment interviews were carried out in assessments #3, #4, #5 and #7.

We took additional feedback from our contact persons (usability specialists) in the companies by email and in project meetings. In most of the cases we were able to follow to some extent what happened in the organisations after the assessments. However, we could do the follow-up only in quite a limited way.

The primary criterion for assessing the validity of our interpretations was the hermeneutic one of seeing if they 'made sense' (Myers, 1997) and were believable both to ourselves and to the key people involved with the assessments. Interpretive case study research allows the researcher to generalise from the case to the theory, and to obtain deep insights about phenomena (Walsham, 1995). Conversely, one of its weaknesses is that it does not generally have much breadth. Our research was not so deep, e.g. data gathering was not equally systematic in all cases—as in studies of one case, such as (Larsen and Myers, 1999). On the other hand, we had a large number of case studies.

5. Descriptions of the case studies

The companies involved in the assessments were different sizes: they included some R&D groups or project groups in Nokia, TeamWare—a software company of 200 employees—Buscom—which is an embedded systems development company of 50 employees—and VTT—a national research institute. Although the companies were very different in size (Nokia vs. SMEs), the sizes of the organisational units assessed were not so different. The size of the groups or projects assessed varied between 10 and approximately 100 people. All the companies operate in international markets. The applications were different, and included mobile consumer products, transportation management devices and software, web based software, customer documentation, electro-mechanics and software of telecommunication network elements and mobile wireless services. In some organisations, the assessment was carried out twice.

The background in UCD in the organisations varied. One organisation had had a usability lab from the late 1980s, while in another organisation the first usability person had joined the company not long before the assessment was carried out. There was some variation in the UCD methodologies the companies used. Contextual Design (Beyer and Holtzblatt, 1998) was used in two organisations, while the UCD book by Hackos and Redish (1998) was used as the main reference in another organisation. Some of the organisations did not refer to any specific main UCD methodology. The timeline of the cases is shown in Fig. 5.

The overall purpose of the assessments was to provide a basis for improvements in the organisational position or performance of UCD. In most cases, the initiators of the assessment were the usability specialists of the organisations. In other words, the assessments were not required by management, but the management gave their permission and support to the assessments.

The author—with usability experience from the early 1990s and with formal training in process assessment—was the lead assessor or member in the assessment team in all

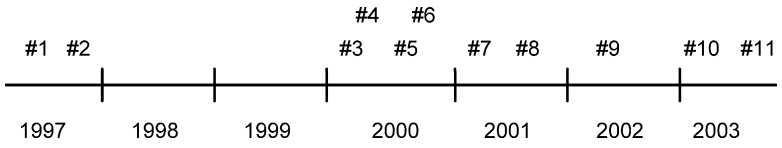

Fig. 5. Time frame of the cases.

assessments. There were several researchers as assistant assessors. Some assessors participated in most of these assessments, while some people attended only one assessment. All of the researchers had at least the basic knowledge of usability and UCD, and received training on process assessment. The members of the assessment teams came from the Oulu University (at least one representative in all cases), Helsinki University of Technology (cases #2 and #4), Lloyds Register (case #1), TeamWare (case #1), and Nokia Research Center (case #2).

In Sections 5.1–5.10, the cases are described one by one in chronological order. In the limited space of this article, the aim is to give an overall picture of the main features of each case. More detailed reports on the assessments can be found from other publications: Kuutti et al. (1998), Jokela (2001), Jokela et al. (2001) and Jokela and Abrahamsson (2004).

5.1. Case #1: starting with a pre-version of UMM-P process model

The author was the initiator of the first assessment, being a usability practitioner in the company. The motivation for the assessment was to find objective evidence of those organisational issues that the author perceived were problematic (unprofessional interventions in usability work, the use of the term 'usability' as a buzzword for political reasons, etc.). It was expected that the results of the assessment would provide clear evidence about those organisational problems.

There were some specific reasons why we took standard process assessment as our assessment approach in the first trial. First, standard process assessment is used very widely in software engineering. Second, it was the latest development in UCM models—i.e. the first version of the UMM-P had been recently introduced. Third, process assessment is well documented and there is training available (the author and some other assessors had attended such training). Fourth, we were able to have the author of the UMM-P model—a person with a solid background in software process assessment and usability—in our assessment team as the lead assessor.

We used the early version of UMM-P (Earthy, 1996). This version of the model did not identify specific UCD processes, but it extended the definitions of existing software engineering processes (such as Requirements determination) from ISO 12207 with UCD practices. Basically, we tried to follow the standard process assessment procedure rigorously up to level 3 capability, Fig. 4 (assessing up to level 3 is typical in software process assessments). In other words, the focus included the performance of UCD activities (as a part of the software development processes), their management and the UCD infrastructure. The assessment procedure lasted almost one week, including a briefing session, interviews, interpretation meetings, and a results reporting session.

The main finding in case #1 was the 'too complicated' structure of the UCD reference model. We were able to generate the results but the assistant assessors were not able to totally follow the assessment. The complexity of the assessment model also became evident even in the session where the results were reported: the assessors had difficulties in remembering their interpretations of the model while explaining the results to the organisation.

Another main finding was that the author—who was also in the role of the customer—was disappointed with the results because the assessment did not focus on those areas that he had experienced as problematic in the organisation. In other words, the assessment did not respond to the original goal of the assessment.

For the development staff the assessment was mainly considered more as an interesting research action than a truly useful activity. It did not have any major subsequent impact in the organisation.

5.2. Case #2: a success story of traditional process assessment

Trial #2 was initiated by a usability manager of another organisation, who was a member of the assessment team in case #1. We used a refined version of UMM-P (Earthy, 1997) where the structure of the UCD reference model was considerably changed (specific UCD processes were defined). Otherwise, trial #2 was conducted in a similar way as trial #1: the standard assessment procedure was followed and the processes were examined up to level 3 where applicable.

The new version of the UMM-P model identified seven UCD processes altogether, out of which four were actual usability engineering processes (adopted from ISO 13407, Fig. 1), and three related UCD processes (such as planning UCD activities). All the seven processes defined by the model were included in the assessment.

The assessment was perceived generally as much more successful than case #1. Specifically, the structure of the UCD reference model—explicit UCD processes—was perceived as much simpler now. Results were generated, and acceptance of the assessment by the customer was good. For example—although no systematic customer feedback data was gathered—the overall impression from the results reporting session was that the staff found the results useful and interesting. The assessment results also led to an improvement effort on a specific process that was found to be at a rather low level of capability. The improvement effort was also implemented.

One distinctive feature of the assessment was that it turned into a kind of 'show' by the lead assessor who was an accredited lead assessor of software process assessments but with only limited background in usability. The findings and ratings were discussed within the assessment team but the role of the assistant assessors was relatively low. On occasions the author could really not understand and agree with all the judgments, especially when examining the performance of the processes stage (i.e. when the achievement of level 1 of process capability was determined). Overall, the assessment was an (enjoyable!) show for the lead assessor.

More details of cases #1 and #2 are reported in Kuutti et al. (1998).

5.3. Case #3: troubles with the process definitions

This assessment was the first one in the second research project (KESSU), and was initiated by the researchers. It was thought to be an effective step to start research work in the new project.

The same assessment model (Earthy, 1997) was used in case #3 as in case #2. However, we decided to limit the assessment to examining the performance of processes up to level 1 of process capability only (Table 1). In other words, we excluded the examination of project management and quality systems, Fig. 3. The lead assessor's (i.e. the author) opinion was that it was better to focus just on the basics of the UCD processes since the organisation to be assessed had a much shorter history in UCD than in cases #1 and #2. As a further limitation, we included only the four 'core' usability engineering processes in the assessment—i.e. those described in Fig. 1—but excluding the other UCD processes of UMM-P (such as planning UCD).

There was, however, an additional focus in assessment #3: we tried to assess the UCD skills and awareness of UCD in the organisation. One of the lessons learnt from case #1 was to cover the organisational obstacles of UCD more widely. We also planned—and tried to some extent—to expand our focus on these wider organisational issues.

Although the focus of the assessment of the usability engineering processes was simpler than in case #2, the assessment was perceived as much more difficult than it was in case #2. This time we examined the UCD reference model—i.e. the process definitions of UMM-P—in detail and tried to understand what the model really requires for the achievement of level 1 capability—and had many interpretation problems. When parts of the model were perceived as unclear, it was very difficult for the assessors to map the findings of the organisation against the definition and give ratings. This led us to giving the results in our own style rather than following the standard way of presenting results.

Our communication with the organisation was not totally successful. In particular, we were not able to communicate the UCD reference model—i.e. the process definitions and capability scale of UMM-P successfully. It was perceived by the staff as difficult to understand. We later heard that the staff had described the assessment with comments such as 'totally mysterious' and 'academic stuff driven by the interests of the university'.

We did not complete the assessment of other organisational issues (awareness, skills). Firstly we just perceived reporting such things as too sensitive. Secondly, we found the assessment of the 'basic things'—usability engineering processes—challenging enough and did not want to assign our resources to further challenges.

Our main lesson from the assessment was that, for the next assessment, we needed to have a clear interpretation of the UMM-P model. We also concluded that the main focus of the assessment—limitation of the focus on the performance of UCD processes—was meaningful. Moreover, we concluded that making an assessment sense-making to the staff is extremely important.

5.4. Case #4: repeat of case #2, but unsuccessfully

The assessment in case #4 was carried out in the same organisation as case #2, also using the same UMM-P model used in case #2. Despite the lesson learnt from case #3,

we did not make a 'clear interpretation' of the UMM-P model. There were two specific reasons for this. First, the customer was the same as in case #2 and was interested in a reassessment after a period of two and half years. It was logical to repeat the UMM-P model, as it had been used in the earlier assessment. Second, the lead assessor in the assessment—an experienced usability professional—was an assistant assessor in case #2 but was not involved in cases #1 and #3.

This time the assessment was much less fluid than in case #2. Rather, the experiences resembled those of case #3. There were many disagreements about the right interpretations of the process definitions of the UMM-P model within the assessment team. The negotiations about interpreting the findings and generation of the results were prolonged and time-consuming. Determining the ratings was perceived as difficult, and disagreements ensued.

In the end, the key customer (the usability manager) was not happy with the results; the customer had expected better results than they got. As a consequence, some processes had to be reassessed to check whether the ratings could be made higher.

The feedback from the staff was not very good. Many reported that the results of the assessment were vague and they felt that many important areas related to usability engineering were not discussed. The terminology, the process definitions and the maturity scales were perceived as difficult to understand.

5.5. Case #5: a new UCD reference model

In this assessment, there was a considerable change in the UCD reference model. Based on the lessons learnt from cases #3 and #4, we started to make a 'concrete interpretation' of the UMM-P model. After making this interpretation, however, we realised that our interpretation was so different from the original model that we concluded we actually had a new UCD reference model, the *KESSU process model*.

Both the UMM-P and KESSU models define the UCD processes. A key difference between the models is in the format of how the processes are defined. In the UMM-P model, the process definitions typically identify many desirable features. The process definitions in the KESSU model were made simpler through defining the processes through outcomes, i.e. defining what a process should produce. A four-grade scale[10] (none, partially, largely, fully) is used to rate the extent to which an outcome is produced, as well as the quality and integration of the outcome (for details, see Jokela (2001)). We found the new style of definition work throughout the assessments that followed, and actually did not even consider changing it.

The process structure of the UCD reference model (how many and which processes) evolved during the assessment. We started assessment #5 with the same set of processes, i.e. four, as in UMM-P (or ISO 13407). During the assessment, however, we decided to change the process structure so that it had five processes instead of four. We found the activities of designing users' work and designing user interface such different kinds of outcomes (task descriptions vs. user interface design) that it was logical to make them

[10] The scale is adopted from process assessment (used in rating the achievement of base practices).

separate processes. Therefore, we split the process 'Product design solutions' into two processes: 'User task design' and 'Interaction design', see Fig. 7.

The focus of the assessment in case #5 was the same as in case #3: the performance of usability engineering processes. This limited focus was kept because we found that the performance of UCD is the basic issue that needs to be understood. The author was the lead assessor (as in all the later cases, too) and all the members of the assessment team had been assistant assessors also in cases #3 and #4.

The overall experience of the assessment was more positive than in cases #3 and #4. The struggle to interpret 'someone else's model' was not an issue any more. Rather, the challenge was to create a model that would make clear the strengths and weaknesses of the usability engineering in this specific case. This thinking process fell mainly to the lead assessor (the author). The experience was very motivating and interesting.

The assessment was completed, the results reported and the organisation agreed on an improvement plan. The fact that an assessment of another team within the organisation was initiated later (case #9) indicates the perceived usefulness of the assessment by the customer.

5.6. Case #6: short but successful

The preparation for the cased #6 came with short notice and the assessment had to be carried out with few resources and little time. We used the same UCD reference model as in case #5 and the focus was also the same: the performance of usability engineering processes. The implementation of the assessment, however, was very different. Instead of the usual timeframe of several days and hours (for a planning session, a briefing session, individual interviews, a results reporting session, etc.), this assessment was implemented as a workshop in four hours. No formal results (ratings) were given. We just presented our findings qualitatively (using statements such as: 'you did not pay very much attention to this process').

The overall experience and customer feedback was very good. The customer gave only positive comments: 'The assessment pointed out many issues to be considered in later projects'; 'Now we know that task analysis is important. We also need to work on usability requirements', and 'We found that your model worked well also in our domain'. The assessment team felt that it succeeded in forming a credible picture of the development process in a short time.

Further indication of the usefulness of the assessment as perceived by the customer usability specialist was when she later moved to another organisation, she ordered another assessment there as well. In fact, that assessment is case #8.

5.7. Case #7: challenging contextual design

The positive experience of case #6 led us to conduct case #7 basically in a similar style. But now we had more resources and time. We had three separate workshop sessions: one for group interviews, one for reporting the results and a third one for planning improvement actions.

We generally perceived the assessment as fluid. We, however, found it reasonable to make refinements to the process structure of the original KESSU process model. Our findings revealed that the project team had not really identified and thought of the different users of the product under development. To be able to communicate this finding clearly to the project team, we made the activity more explicit in the KESSU process model through defining a new explicit process 'Identify user groups', Fig. 7.

When the results were reported, discussions were perhaps livelier than in any of the earlier cases (e.g. in the feedback: 'There was a lot of discussion'). The reactions varied somewhat among the people. Most of the staff found the results constructive ('The model opened my eyes'; 'A good model') but some were not so impressed ('Nothing new'). The project manager expected better results and apparently felt insulted after seeing some low ratings (although the ratings were relatively high!) and started to defend the work done in the project. The usability specialist, on the other hand, accepted a critical and objective analysis. Some of the staff challenged the contents of the process model—which we interpreted as a good sign: if the contents were challenged, the model was also understood.

The specific setting in this organisation was that Contextual Design, CD, was used as the design methodology. One finding was that some 'CD practitioners' found the assessment useful in the sense that the assessment challenged some parts of the methodology.

The major follow-up action of this assessment was that the KESSU process model was taken as the basis for planning UCD activities in the next project in the organisation. Moreover, the success of this new project was tied to its adherence to the KESSU process model. In other words, the better the project follows the KESSU process model, the better its success would be. This setting led to assessment #10 where the success of the new project was assessed.

5.8. Cases #8 and #9: assessment of no-existence

These two assessments were initiated by usability specialists who had been in the role of the customer in earlier assessments. Case #8 was initiated by the usability specialist in case #6, and assessment #9 was initiated based on the experiences in case #5.

The approach used in case #7 was planned to be used: i.e. the KESSU process model and three workshop sessions for interviews, results presentation, and improvement actions. The interview session of case #8, however, revealed that there were practically no UCD activities carried out in the project. As a consequence, we decided to change the planned results reporting session into a general UCD training session. It was not explicitly explained that the assessment was to be interrupted—we just organised the session at the time planned, but instead of reporting the results we provided UCD training.

Case #9 was similar to case #8. Having obtained very few findings on UCD in the interview session, the next session was changed 'on the hop' into a UCD training session.

In both cases, there were quite a large number of participants (around ten in each) in the workshops. They had booked the whole day for the results reporting session. The reason for deciding not to complete the assessment was that we just did not find it constructive to use the whole day to report 'bad news'.

5.9. Case #10: judging money

The specific feature of case #10 was that the purpose of the assessment was to evaluate the project success: we examined to what extent the project truly followed the KESSU process model used in assessment #7. The financial bonuses for the project staff were tied to the extent to which the process model was followed during the project.

The assessor (the author) was not very critical and gave relatively good ratings. However, the usability specialist was not totally satisfied with the results because they were 'too good' and did not pinpoint the organisational problems she had perceived. In other words, the usability specialist hoped for a highlighting of the organisational problems of UCD, rather than receiving higher financial bonuses for herself and the project staff!

5.10. Case #11: UCD and extreme programming

This assessment was initiated by a project manager without specific usability background. The background for this assessment was the personal acquaintance of the project manager and the assessor (the author). The project manager had recently finished a successful extreme programming (Beck, 1999; Cockburn, 2002) project, where communication between the project team and the customer was emphasised. The project manager was aware of the author's work on user-centred design, and was curious to see how his project would rate in terms of UCD.

There were only very few UCD findings on the project. As a consequence, the project got very poor ratings but the results were reported 'in the raw' to the customer.

Basically the assessment approach—and the focus of the assessment—was the same as it had been in the earlier assessments (from case #4 onwards). However, to make the findings clearer to communicate, we found it appropriate to refine the KESSU process model again. We separated the two different types of usability evaluation activities into two processes: 'Usability feedback' (could be also called formative evaluation) and 'Usability verification' (could also be called summative evaluation). This case study is discussed in detail in Jokela and Abrahamsson (2004).

5.11. Summary

The main characteristics of the different cases are summarised in Fig. 6. Two main UCM models were used. Different versions of the UMM-P model—i.e. a standard process assessment model—were used in trials #1–#4. A new assessment model ('KESSU') was created for trial #5 and further developed in the subsequent assessments, up to assessment #11.

Although all of the cases had some positives and negatives, one can probably summarise that cases #2, #6, #7, #10 and #11 were successful while cases #1, #3, #4 were less successful. Case #5 was somewhere in between. If the assessments #8 and #9 had been completed, they probably would have been some kind of disaster.

All the assessments covered one common focus. The performance of usability engineering activities was examined in all assessments. Moreover, the examination

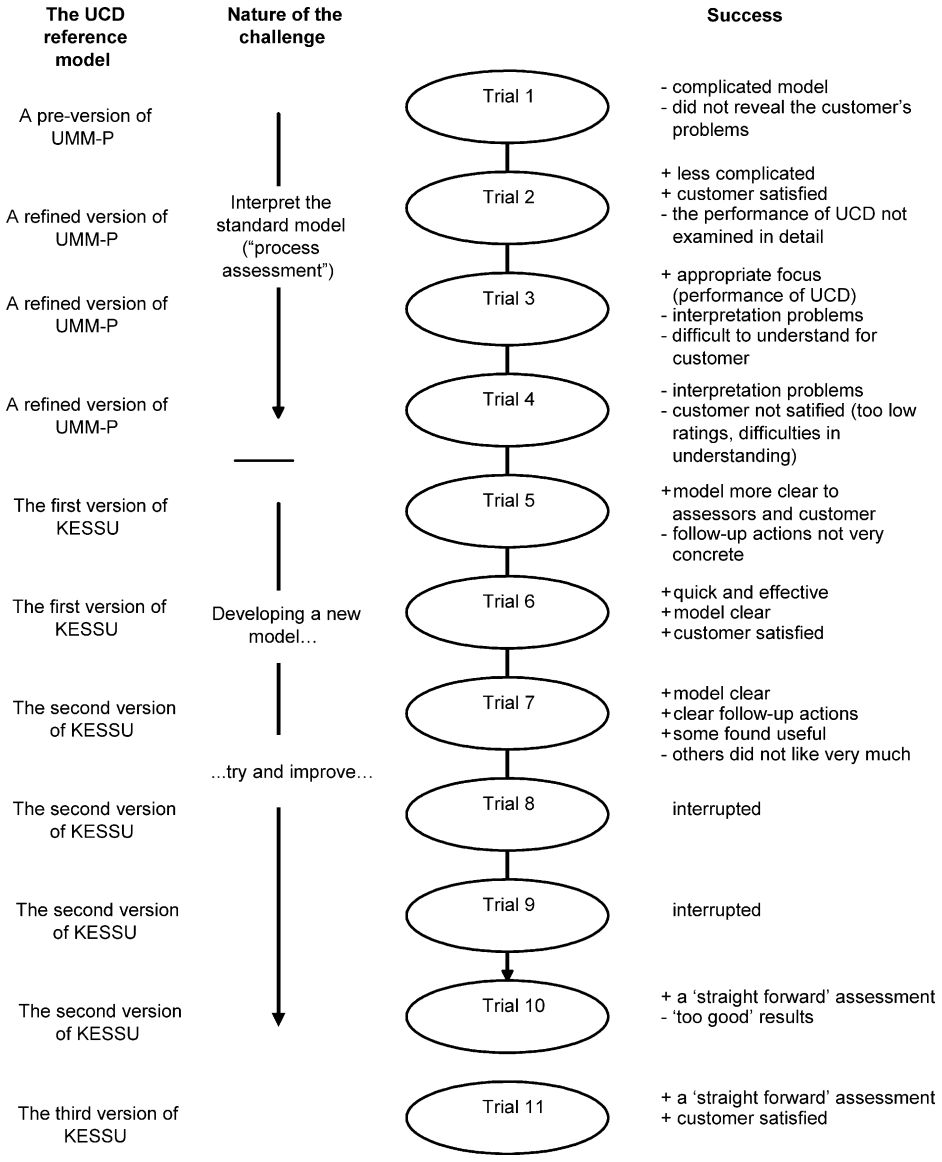

Fig. 6. Overview of the trials.

covered the whole usability engineering lifecycle (i.e. from user analysis to usability #5–#11evaluations) in all cases. Cases #5–#11 included only this focus ('a' in Fig. 3).

In assessments #1, #2, and #4 the focus was one of traditional process assessment. In other words, the assessment also covered the examination of how UCD activities were managed in the projects, and the extent to which UCD was included in the quality systems

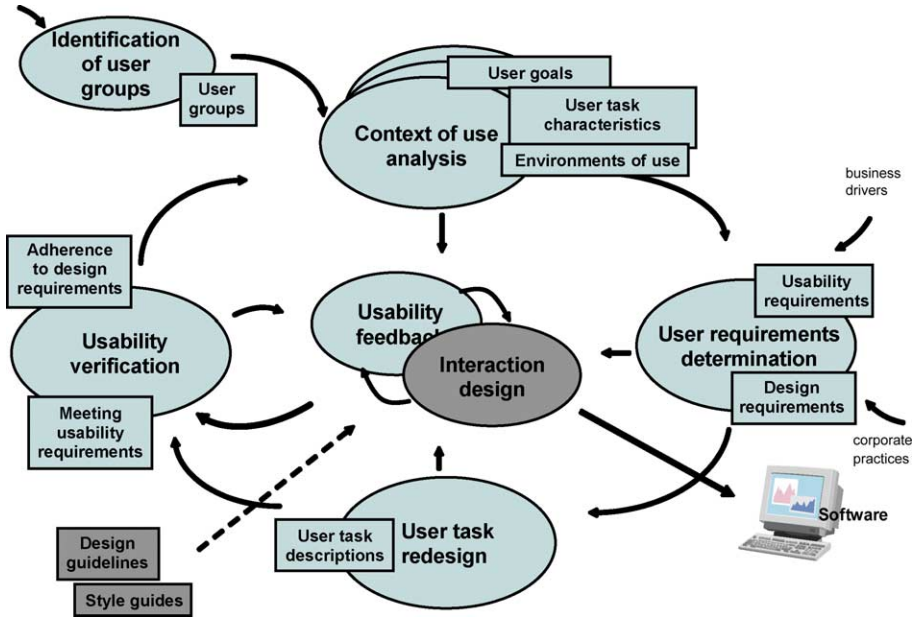

Fig. 7. The latest version of the KESSU process model (contrast the model with the one of Fig. 1).

of the organisations ('a–c' in Fig. 3). In case #3, other organisational aspects were included in the assessment (but the assessment was not completed).

The latest version of the KESSU process model is illustrated in Fig. 7. In summary, the reason for the evolvement of the KESSU process model was to make it clearer to communicate the findings and results of the assessments to the organisations. A detailed description of the KESSU process model is not in the scope of this article. It is described, e.g. in Jokela (2001a,b, 2002) and Jokela and Abrahamsson (2004).

6. Interpreting the case studies—conclusions on UCM assessments

One can see that there were also many different—and even contradictory—phenomena in our cases. For example, in cases #1 and #10 the customer was disappointed because the assessment did not clearly point out the problems in the way they expected. On the other hand, in case #4 the customer was disappointed because the results were not as good as expected, and in cases #8 and #9 the results were not presented at all because they had been too 'bad'. In case #11 the 'bad' results were presented, and the customer was satisfied.

Moreover, the experiences were not always in line with the UCM model that was used. For example, the same UCD reference model was used in cases #2–#4—still the experiences from these assessments were quite different.

It is concluded that this kind of phenomena in the assessments can be explained by the different *purposes* that the customer had for the assessments. In other words, although the overall purpose of all our assessments was to provide a basis for improvement action

(vs. the purpose of third party certification), there were specific purposes that altered the phenomena around the assessments. Altogether, four different main purposes that can be identified in our cases:

(I) to *awaken* the organisation
(II) *kick-off* for a usability improvement program
(III) *monitor* progress
(IV) curiosity

Some of the cases had one main purpose. In some other cases, two or more main purposes were present.

(I) Case #1 and also partly cases #7 and #10 represent category I, where the purpose of the usability specialists for conducting the assessment was to 'awaken' the management of the organisation. The usability specialists were not happy with the organisational position of UCD, and hoped to get concrete evidence about the problems through an assessment. The usability specialists expected to get 'low' ratings that would provide evidence about the problems in the organisational position of UCD. Case #10 shows this purpose very clearly. The purpose of assessment #10 was to monitor the project success. However, the usability specialist perceived the results as 'too good', which indicates that she wanted to awaken the organisation to the need for further improvements to the position of UCD.
(II) Category II represents a situation where an assessment was used as a kick-off for a usability program. This category of motivation was present in most of our cases (#2–#9). The assessments were partly initiated as a 'me too' action: a company wanted to have an assessment because the other companies (in the research project) had had one. The overall idea for conducting an assessment was that it would make sense to understand the past in order to know how to do it better in the future. We learnt most lessons from this category.
(III) Cases #4 and #10 represent category III: they were conducted to monitor progress in the performance of UCD in the companies. In other words, the companies wanted to find out whether they had progressed in UCD compared with the time when an earlier assessment had been carried out. Case #4 was a follow-up of case #2, and case #10 was a follow-up of case #7.
(IV) Case #11, especially, represents category IV. The initiation of the assessment was based on 'objective curiosity'—the customer was just interested in knowing how his project rated in terms of UCD. This category of purpose represents personal and individual interest in UCD, and is less connected to organisational improvement efforts or other impacts than the other categories. This purpose was also partly present in cases #5 and #7 where the usability specialists of the customers were rather experienced and had an interest in benchmarking their understanding and knowledge of UCD through an external reference.

In summary, the different categories of assessments based on the different purposes are shown in Table 2.

Table 2
Purposes for assessments in different cases

Purpose	Case										
	1	2	3	4	5	6	7	8	9	10	11
(I) Awakening	×						×			×	
(II) Kick-off		×	×	×	×	×	×	×	×		
(III) Monitoring				×						×	
(IV) Curiosity					×		×				×

On the other hand, one can find many phenomena that were common to all the assessments. In Section 6.1, the conclusions that are generally applicable to the assessments are presented. Thereafter the conclusions that are specific to the different purposes of assessments are discussed.

6.1. Findings applicable to assessments generally

6.1.1. An assessment is an organisational intervention

Ideally, assessments could just be a neutral way of determining the status of UCD in a development organisation. However, during an assessment, the assessors are in touch with the organisation in several ways. Assessors discuss with staff, interview them and present briefings and the results. Our trials show that assessments are inherently organisational interventions that are not only neutral status finding actions. Assessments have an impact on the people, on how they perceive and understand usability and UCD.

The more people are involved in the assessment, the more the knowledge of usability and UCD can be spread throughout the organisation and the bigger the impact the assessment has. On the other hand, the more a person assigns his or her resources to an assessment, the bigger the expectations she or he has. Therefore, an assessment means risk. The impact may be negative if the people do not perceive the time they spend on assessment as meaningful and worthwhile. This risk was partly realised in practice, at least in cases #3 and #4.

6.1.2. The UCD reference model has a critical role

The UCD reference model determines what kind of foci an assessment may cover. The model is the basis for data gathering—examination of documentation and interviewing people. The model guides and has an impact on the kind of data which should be sought. The model is used for determining the assessment results. Then the results are reported to the staff using the model as a reference. Finally, the model may be used as a reference in future development projects, as in case #7.

6.1.3. An assessment is about training in usability and UCD as defined in the UCD reference model

The practices of an organisation are mapped against ‘what UCD should be’ in an assessment. In practice, when the assessment results are reported to the staff, an explanation should be given about the nature and purpose of UCD (i.e. the contents of

the UCD reference model). For example, an assessor first explains what does 'user groups should be determined' mean. Then the assessor reports the findings (or non-findings) from the development project related to the 'user group determination' as the justification for the capability rating. Similarly, each activity of UCD is gone through. In practice, in the end, the assessor has trained the participants in the basics of UCD—as defined in the UCD reference model.

6.1.4. A specific interpretation of the UCD reference model is important

It is important for the assessors to have a profound understanding of the contents of the UCD reference model. If the reference model is perceived as unclear, the consequences are very laborious interpretation sessions and disagreements between the assessors. The difficulties in cases #1, #3 and #4 reveal the importance of this.

If an existing UCD reference model is used (i.e. a new model is not developed), the model should have very good documentation and be unambiguous, clear and understandable. The assessment team made big efforts to examine the documentation of the UMM-P model—and the sources of the model—in cases #3 and #4. The UMM-P model was found to be the best documented model of the existing ones. In addition, most of the assessors had training in the underlying assessment paradigm (i.e. process assessment). Still, we faced interpretation problems and did not have a deep and common understanding of the contents of the model.

6.1.5. Interpretation of the status of UCD in the organisation requires clear mapping between the findings and the UCD reference model

In the course of cases #5–#11, the UCD reference model (the KESSU process model) evolved step by step. The reason for each refinement of the model was the specific findings found in the state of UCD in an organisation. We made refinements to the KESSU process model because we judged it useful to make the results as clear as possible to the customer.

6.1.6. The role of the lead assessor and the assessment team is critical

The role of the assessment team is central—especially the role of the lead assessor. The ways of conducting the assessments, how the UCD reference model and the findings were interpreted and the results reported were very much dependent on the viewpoints of the lead assessor and the assessment team.

The lead assessors were different persons and also the role of the assistant assessors was different in cases #2–#4. This is probably the main reason for the different experiences in these assessments. This is not to claim that the assessment team in cases #3 and #4 (a less successful assessment) were in any way 'worse' than in case #2 (a more successful assessment). The key difference was that the UCD reference model was interpreted differently by the lead assessors, which led to different assessment experiences. In cases #3 and #4, the assessment teams really tried to understand the process definitions thoroughly. The author was in the position of an assistant assessor (among many others) in case #4. On reflection, I can say that I probably was a key reason for the disagreements in case #3—I simply disagreed or questioned the viewpoints of the lead assessor about the interpretation of the UMM-P model.

6.1.7. There are personal differences about how assessments are perceived

This is probably not a surprising finding but important to bear in mind. Some people just do not like assessments or they are just not interested in usability and UCD. For example, in case #7 most of the feedback was positive, but some participants did not find the assessment useful ('nothing new').

There also are personal differences about how people perceive the results. For example, in case #7 there was a clear difference between how people reacted to the results (which were generally relatively good). Some of the staff was disappointed when the assessment identified some clear problems. On the other hand, some other staff found the results interesting and useful ('opened my eyes').

6.1.8. An assessment is a useful exercise for a usability practitioner

An assessment is a personal challenge for a usability practitioner. He or she needs to map his or her understanding and knowledge of the essence of usability and UCD practice against the UCD reference model.

An assessment also clearly helped the practitioners to understand UCD beyond its specific methodology. For example, we found in case #7 that part of the project team perceived a specific UCD methodology (Contextual Design) as 'the way' to carry out UCD work.

Being an assessor is even a more challenging exercise for a practitioner. One needs to form a clear understanding of what exactly UCD is (i.e. a model of UCD with which to evaluate a company). The model should clearly state what the elementary things are for user-centredness. Thinking of the contents of the reference model is a useful thinking exercise for an assessor. One should be able to objectively assess different organisations where different UCD methods and approaches are used.

6.2. Specific findings in category I: awakening

In this category, where an assessment is used as a means for a usability specialist to awaken the organisation, one specific factor can be identified:

6.2.1. It is important that the results clearly show the problems in the organisational position of UCD

In cases #1 and #10, the assessment did not meet the expectations of the customer in the sense that it did not produce those 'bad' ratings that the usability specialists hoped to have. In case #1, the problematic areas were not examined at all; in case #10, the ratings were just too good.

The conclusion is that in this kind of situation it is important that the focus of the assessment covers those organisational issues that the usability specialist finds problematic. However, one should make one remark here. What if the assessment had produced the 'bad' ratings that the usability specialists hoped to have? We did not get any evidence of the usefulness of such results (because the assessments did not produce those 'bad results'). It may be that such 'bad' results might not help at all. For example, in case #1 the organisational setting was so entrenched, that even pinpointing the core problems might have brought no new changes.

6.3. Specific findings in category II: kick-off

Most of our assessments fall into this category where an assessment is used as an asset for a usability specialist as a kick-off for usability actions. As a result, there are more conclusions about this category.

6.3.1. A constructive training opportunity is more important than assessment results

Probably the main conclusion from these cases is that it is not very important to find out the exact status of UCD in the organisation. Rather, an assessment is a success when it can be a constructive training occasion on usability and UCD. The exact assessment results—the ratings of the current status of UCD—are not so important. In case #6, the results were given only qualitatively and the customer was happy. In cases #8 and #9 no ratings were given at all, and a probable disaster was avoided.

It could be said that in this category the purpose of an assessment is to 'sell' usability to the company. Therefore, an assessment should be a positive experience, and reporting 'bad' results probably does not support this objective.

6.3.2. An appropriate focus is the performance of usability engineering processes

Another conclusion in this kind of setting is that it is sensible to focus on the examination of the 'basics focus', i.e. the performance of usability-engineering activities. This focus means that all the discussions during an assessment are truly about the essence of UCD: what UCD activities are carried out, how they are done, how the results are utilised in a development project. This is important because, as stated above, an assessment is essentially an opportunity for training in usability and UCD. In standard process assessment the focus would be mainly on the management aspects (how a UCD activity is planned; is the activity written in the project plan; is the performance of the activity reviewed; is UCD documented as part of the quality system, etc.) Although these are important issues, discussions of those issues are far removed from the resources that could be assigned to discussing and training in the essence of UCD.

6.3.3. Examination of all usability engineering processes is useful

Another conclusion is that all the usability engineering processes should be examined (in contrast with the standard process assessment where it is typical to predefine a set of processes and assess only those). We started the trials by examining all the UCD engineering processes—from user analysis to usability evaluation—and kept this focus all the time. The examination of the whole UCD lifecycle—i.e. not only focusing on some selected UCD activities—is important because UCD activities form a logical continuum. If one activity is not carried out, then there is a hole in the user-centred design lifecycle. Further, because assessments are also training opportunities, it is essential to give an overall picture of UCD. Third, the process structure may evolve. For example, even the last case (#11) raised the need to restructure the KESSU process model. Predefining the processes to be assessed would potentially mean that some relevant and interesting phenomena would not be revealed.

6.3.4. Concreteness and sense-making of the UCD reference model are important

The understandability and sense-making of the UCD reference model are specifically important in this category. The reference model determines whether the results make sense or not to the people. Improvement actions will be built on the model—as in case #7. The UCD reference model represents the 'truth' of user-centred design to the staff.

If the models and the results do not make sense, as e.g. in case #3 ('it was not understood here'; 'totally mysterious'; 'academic stuff driven by the interests of the university') or were found uninteresting—as in case #4 ('did not address the relevant aspects')—the impact was more or less negative. It was reported later that 'usability' became almost a swear word in the ears of some staff in case #3. On the other hand, in those cases where the staff perceived the model as making sense, the assessment was a positive experience. For example, in case #6 the assessment was perceived as very useful and educational (e.g. in case #6: 'the assessment pointed out a lot of issues to be considered in later projects').

6.3.5. A logical structure and visual presentation of the UCD reference model are of relevance

We found it important to be able to report the findings of the status of UCD as clearly as possible. As a consequence, the process structure of the KESSU UCD process model evolved step by step—although the essential UCD substance of the model did not. We wanted to make the model structure clear enough so that the findings could be mapped in an understandable and elegant way against the model. There were altogether three main versions of the KESSU process model. The assessments #5 and #6 were carried out using the first version. The 2nd version was developed during assessments #7–#9. This version of the model was also used rather extensively in publications, presentations and training. Still, assessment #11 brought up a third version of the model.

When discussing the findings and results, it just happened that the visual picture of the UCD reference model (such as shown in Fig. 7) was used as the basic tool. Textual descriptions were referred to when required, but the visual model remained central.

6.4. Findings in category III: Monitoring

Two assessments (#4 and #10) fall into this category where assessments were carried out for monitoring the progress. The main conclusions are discussed in the following sections.

6.4.1. The selection of the assessment approach depends on the organisation's earlier experiences

The selection of the assessment approach in this category is very much dependent on the earlier assessment experience. This is logical—not really a finding—but worth mentioning because this does not apply to other categories. In case #4, a reassessment was needed; therefore, the assessment approach was the same as the earlier one. In that case, at that stage there really was no true choice.

In the setting of case #10, the selection of the assessment approach was even more connected to the organisation's history. One target of the project was to follow the UCD

reference model used in case #7—thereby the only possible selection was to use exactly that model.

6.4.2. Other purposes and personal issues have an impact on how the customer perceives the results

The customers perceived the results differently in these two cases. In case #4, when the initial assessment results (ratings) proved not to be as good as expected, the customer requested further interviews to check whether the ratings could be better. In case #10, the customer hoped for 'worse' results than those produced by the assessment.

These kinds of different phenomena depend on personal issues on the one hand, and on the potential purposes of the assessment on the other hand. As discussed above, there was also another purpose—to awaken the organisation—in assessment #10. The expectation of 'worse' results is explained by that purpose.

6.5. Findings in category IV: curiosity

In this case, the assessment was carried out for objective curiosity.

6.5.1. Reporting results 'as they are' is important

It is naturally important to give ratings in this setting. Quantitative ratings provide a clear picture of the status of UCD. In this category, the objectivity and preciseness of the results are probably more important than in the other categories. This is true especially if the customer is in a 'neutral' position in terms of UCD and thereby does not have any specific expectations about the results.

From other viewpoints, the many conclusions of category II are applicable. One should focus on the performance of usability engineering and cover all usability engineering activities. The clarity and sense-making of the UCD reference model is important.

6.6. Summary

This section presents the interpretations of the 11 assessments. Some phenomena that are common to assessments generally are identified, such as the critical roles of the UCD reference model and the assessment team. In addition, there are phenomena that depend on the specific purposes that the customer has about assessments. The conclusions are summarised in Table 3.

When the research of Bevan and Earthy (2001) is examined in the light of our results, one can conclude that their assessments seem to represent two different purposes: first the purpose of 'kick-off' and later the purpose of 'monitoring' (in each of the two organisations, a 'start-up' assessment was conducted, and the progress was monitored later). They used practically the same UCD reference model as we did in our cases #2–#4 but their experience was better. The different interpretations of the UMM-P model by the assessors are probably the main reason for the differences between their and our experiences.

Table 3
Summary of findings

	Purpose of assessment			
	Awakening	Kick-off	Monitor	Curiosity
General findings				
Assessment is an organisational intervention	×	×	×	×
The UCD reference model has a central role	×	×	×	×
The interpretation of the UCD reference model is decisive	×	×	×	×
The role of the lead assessor and the assessment team is critical	×	×	×	×
There are personal differences about how people perceive assessments	×	×	×	×
An assessment is about training of usability and UCD to those involved	×	×	×	×
Method-independence of the UCD reference model is important	×	×	×	×
Conducting an assessment is a useful exercise for a usability practitioner	×	×	×	×
Specific findings				
It is important that the results clearly show the problems in the organisational position of UCD	×			
Constructive and sense-making training is more important than assessment results		×		
An appropriate focus is the performance of usability engineering processes		×		×
Examination of all usability engineering processes is useful		×		×
Clarity and sense-making of the UCD reference model are important		×		×
Visual presentation and a logical structure of the UCD reference model are of relevance		×		×
The selection of the assessment approach depends on the organisations' earlier experiences			×	
The other purposes have an effect on how the customer perceives the results			×	

7. Discussion

The purpose of usability maturity assessment is to evaluate user-centred development organisations or projects. The idea is that this information could then be used as a basis for improving the performance of UCD. Several usability maturity models have been developed from the early 1990s. New developments exist for example in Germany

and Japan, and new interest is emerging in the US. There are, however, very few empirical studies on carrying out assessments.

In this article, experiences and conclusions from 11 usability maturity assessment cases are presented. The cases are described in Section 5. The research covered 11 assessments of projects where different kinds of products and applications were developed—from software systems to customer documentation. The lessons learnt and conclusions on conducting the assessments are discussed in Section 6.

The results further develop the knowledge in the field about UCM models and assessments. The existing knowledge of UCM assessments is expanded by identifying several lessons learnt on UCD assessments and different purposes that the customers have for assessments. The cases show that it is important to understand the customers' purpose for an assessment, the roles of the assessment team, and the selection of the assessment approach. The findings confirm the claim by Earthy et al. (2001) that an assessment may be 'a powerful tool to introduce and train user-centred design in organisations'. Our findings, however, show that the impact could be the opposite if the assessment is not carried out in an appropriate way.

7.1. Limitations of the research

One obvious limitation of the research is that all the assessments were conducted for the purpose of providing a basis for improvement action in UCD. Our cases did not include a single case where the other main purpose—third party certification—had been the objective (i.e. the assessment had been carried out to understand the capability of a supplier organisation to develop usable software). A third party certification setting would represent partly the category 'objective curiosity'—a customer would be interested objectively on the usability capability of a supplier company. However, a third party certification would probably also lead to some further conclusions. For example, it might be important that the UCM model used is formally standardised or otherwise generally accepted.

Four different purposes for conducting an assessment (under the broad purpose of 'providing basis for improvement action') were identified. However, one can identify other potential purposes for an assessment that were not represented in our cases. One such purpose would be to compare the performance of UCD of one company with that of other companies. This kind of assessment is typical in the world of software engineering where there exist statistics about the capability levels that companies have reached in software process maturity. This kind of benchmarking possibility is missing from UCM assessments. Therefore, none of the case studies were driven by the interests of benchmarking the performance of one company to others.

Another specific purpose that was not represented in our cases would be to integrate the assessment of UCD with other assessments. For example, the purpose for the development of the UMM-P model was to include UCD processes in large, organisation wide software process improvement programs (Earthy et al., 2001). Software process assessments are frequently performed especially in large organisations, and having a compatible model makes it possible to integrate the assessment of UCD processes into the software process assessment.

The purpose of our assessments was to provide a basis for improvement actions. One limitation of this research is that we were not able to systematically follow up the improvement actions. It is, however, not easy to isolate the role of assessment in the success or failure of the improvement actions that follow. The success of improvement depends on many factors: the resources that the organisation can assign to improvement efforts; the ability and interest of the usability specialists to change, etc. how the staff generally feels towards usability; the support of management, etc.

7.2. Practical implications

Two main categories of implications are discussed: the implications for practitioners who wish to be in the role of an assessor (or lead assessor), and the implications for conducting assessments in different situations.

7.2.1. Being an assessor

Overall, the author would like to encourage practitioners to participate in assessments. The assessments are very useful professional exercises. The specific usefulness of assessment is the fact that one needs to challenge his or her understanding of the essence of usability and UCD.

There is, however, one remark that should be taken into account. A UCD reference model that one really understands and believes in should be selected. It is problematic to use a model without really understanding it and accepting its contents. Presenting assessment results often means discussions between the customer and the assessment team. It is essential that an assessor is able to explain and make sense of the UCD reference model and justify their findings.

7.2.2. General considerations for carrying out an assessment

If an assessment of the UCD practice of a company is considered, there are a number of issues to be considered.

Firstly, here are some general things to be remembered:

- ☐ One should understand that carrying out an assessment may be a useful action but is also a risk. An assessment is always an organisational intervention. It necessarily has some impact on the organisation—positive or negative (or both). An assessment may be perceived as negative by some persons in the organisation, no matter how well the assessment is planned, how ‘good’ the UCD reference model is, how right the focus of the assessment is. People perceive assessments differently—so it is not easy to anticipate the success of an assessment;
- ☐ Attention should be paid to the selection of the UCD reference model. The role of the model is critical; it has a major impact on many different aspects in assessments;
- ☐ Attention should be paid to the selection of the assessment team—and especially to the selection of the lead assessor. The way an assessment is conducted and how the findings are interpreted depends on the understanding and viewpoints of the assessment team. The assessment team should also be ‘compatible’ with the selected

UCD reference model. In other words, the team should have a profound understanding of the UCD reference model and believe in it;

- An assessment is a research action. The assessment procedure and the UCD reference model should be challenged. It should be understood that no assessment model represents an absolute truth. In our cases, continuous development of the assessment model took place from assessment to assessment. Tailoring was not done only before an assessment, but also during the assessment: some things were not explored if found irrelevant (decided by the assessor). However, in many cases, tailoring is not purposeful and not even possible. Although an assessment model is followed strictly in an assessment, one should be critical of it and learn from the experience.

Finally, one should try to understand what the specific purpose behind the assessment is: i.e. awakening, kick-off, monitoring, or curiosity.

7.2.3. Using an assessment as a means to awaken the organisation

This kind of situation emerges when usability specialists find problems in the organisational position of UCD. Could an assessment—through providing 'evidence' about the organisational problems—be an effective asset for a usability practitioner to improve the position of UCD in this kind setting?

Our research does not provide a solid answer to this question: we do not have any empirical evidence of a 'successful' assessment in this category. The effectiveness of an assessment is probably dependent on the specific situation in the organisation. The organisational problems may be so entrenched that even pinpointing the organisational problems might lead to no significant changes.

An assessment in this situation, however, could be useful for a practitioner if it is difficult for him or her to understand what exactly the organisational problems that negatively affect the position of UCD are. An assessment using an appropriate UCD reference model could be useful in helping a usability practitioner to identify the root reasons of the problems in the organisation.

7.2.4. Using an assessment in the kick-off of a usability improvement program

When an assessment of this kind of situation is considered, the goal should be that the overall perception of the assessment is positive in the organisation.

When it is obvious that there are no, or very few, usability activities in a company it might be wise to consider not carrying out the assessment at all. When looking back over our case studies, in the beginning we did large-scale assessments in many situations where we skipped over the actual assessments at the end of our research project. An indication of this kind of situation is, for example, if a usability specialist has only recently started in the company. An assessment in such a case would only give 'bad' results. There is a risk that reporting such results probably would just lead to negative attitudes towards usability and UCD among the staff. In companies where one can anticipate that there is no, or very little, UCD, a better way of starting improvement action might be to simply start doing UCD in some specific development project.

If an assessment is in anyway planned, the main benefit of an assessment should be to make people understand the essence and application of UCD—not necessarily to provide

exact and objective ratings on the level of usability capability. Mapping the existing practices against a reference model is an effective way of doing this but showing and emphasising poor results should be avoided.

Attention should be paid to the selection of the UCD reference model. As discussed earlier, the model has a critical role in many aspects of an assessment. Probably the main point to be remembered is that training in UCD can be provided through the model. The model should thus make sense and be understandable. Visual representation and an appropriate process structure are important. The whole UCD engineering lifecycle of a development project should be examined. In addition, as discussed above, the assessors should believe in the 'truth' of the UCD reference model.

7.2.5. Monitoring the progress of UCD in an organisation

This is a limited setting in the sense that the earlier experiences of the organisation in assessments affect the selection of the assessment model and the focus of the assessment.

If the goal is to compare the current status with an earlier one, one should naturally select the same UCD reference model, and one should aim for objective ratings. On the other hand, one should try to understand the expectations and other purposes that the customer sets for the assessment. As cases (#4 and #10) indicate, the customer also had other purposes.

This kind of setting is easier than the other purposes in the sense that the organisation has earlier experience of assessments.

7.2.6. Using an assessment 'just for curiosity'

This is a situation of which we gained only very little experience. This kind of setting, however, could be more general. In case #11, the initiator of the assessment was a project manager. However, a similar situation may happen when senior management takes an interest in usability and is interested to know where the company is in terms of usability.

Providing objective results would be a key in such a situation. Probably it would be wise to keep the organisational intervention as small as possible, and report the results to the management only at first.

7.2.7. The KESSU process model

The KESSU process model, Fig. 7, is a generic, method-independent model of UCD. The model might also be useful for other purposes than assessments. The author has used it in education and training. The model could be further useful for planning and implementing UCD activities in projects, as was done at the end of assessment #7 in that specific organisation.

7.3. Further research topics

As discussed above, one can identify the potential purposes for assessments that our cases did not cover: assessment for benchmarking against other companies, and an assessment that is integrated with a larger assessment program in an organisation.

Neither did our cases include any third party assessment. Empirical studies in these kinds of different settings would provide complementary findings to the results presented in this article. Other trials by others using other UCM models in different settings would be an important follow-up to this research.

The UCD reference model has a very critical role in assessments. It represents more or less the truth of UCD to the staff of the organisation to be assessed. This raises a question: what exactly is the ideal UCD reference model? The existing UCM models have differences in how they define 'ideal UCD'. The assessment cases of our research led to the development of another model (KESSU). The recent activities in Germany and Japan have led to models that, again, are different. Research into understanding the fundamental contents of UCD would be most relevant. The specific challenge to defining the 'ideal model of UCD' is that, on the one hand, it should include all the critical aspects of UCD but, on the other hand, it should not be tied to any specific UCD methodology or practice and thereby set limitations on the innovations of new UCD methods.

Carrying out research into developing a well-founded model of UCD would not only serve assessments but also provide a foundation for identifying needs for enhancing existing, and developing new, usability methods. For example, there are not very many effective methods for the usability requirements phase.

Our research focused mainly on the evaluation of the performance of usability engineering processes. The first assessments (#1, #2, and #4) also covered project and process management. The focus of the assessment could well be wider. Other organisational factors, such as skills, technology and organisational structure could be included in assessments. Some UCD models, such as UMML and DaTech, include these aspects. However, research findings from the assessment of these aspects have not been reported, and would provide an interesting window on the field.

Our experience reveals the importance of good documentation of the UCD models—and especially good documentation of the UCD reference model. Any model, meant to be used by others, should be very well documented. This is a true challenge for anyone developing such a model. Many interesting UCM models, such as UMML, suffer from inadequate documentation. The author is aware of this research challenge. The documentation of our own model, the KESSU model, is not as good as it could be. Producing high quality documentation of a model is not an easy task.

8. Conclusion

Usability capability maturity, UCM, assessments are useful as a basis for strategic initiatives to improve the performance of UCD in a company. Our trials show that such assessments are not a straightforward matter. We draw some general conclusions on conducting assessments but found that the specific purposes that the customer has for an assessment affect how it should be conducted. Therefore, it is imperative to understand the detailed purposes of the assessment, the importance of the assessment models used, and the critical role of the assessment team.

Acknowledgements

I would like to thank many persons who were involved in the assessments, including Pekka Abrahamsson, Jonathan Earthy, Petri Hyyppä, Netta Iivari, Mikael Johnson, Pirkko Jokela, Erkki Jylhä-Ollila, Mikko Jämsä, Tonja Molin-Juustila, Marko Nieminen, Tero Posio, Mikko Rajanen, Risto Rauhala, Juha Rikkilä, Timo Salmu, Tuula Strömberg, Pekka Suhonen, Juha Vaihoja, and Juhani Warsta. Special thanks to those who further gave comments on this article: Marjo Favorin, Harri Kiljander, Kari Kuutti, Katriina Nevakivi, Mikko Siponen, Eija Suikola and Johanna Tapio. A major part of this study was sponsored by TEKES.

References

Anderson, R., 2000. Organisational limits to HCI. Conversations with Don Norman and Janice Rohn. Interactions 7 (2), 36–60.

ANSI, 2001. Common Industry Format for Usability Test Reports. NCITS 354-2001 2001.

Axtell, C.M., Waterson, P.E., Clegg, C.W., 1997. Problems integrating user participation into software development. International Journal of Human–Computer Studies 1997, 47.

Beck, K., 1999. Embracing change with extreme programming. IEEE Computer 1999;, 70–77.

Bevan, N., Earthy, J., 2001. Usability Process Improvement and Maturity Assessment, IHM-HCI 2001, Lille, France. Cépaduès-Editions, Toulouse.

Beyer, H., Holtzblatt, K., 1998. Contextual Design: Defining Customer-centered Systems. Morgan Kaufmann, San Francisco.

Bloomer, S., Wolf, S., 1999. Successful Strategies for Selling Usability into Organizations. CHI 99, Pittsburgh.

Carroll, J.M., Mack, R.L., 1985. Metaphor, computing systems and active learning. International Journal of Man–Machine Studies 22 (11), 39–57.

Coallier, F., McKenzie, R., Wilson, J., Hatz, J., 1994. Trillium Model for Telecom Product Development and Support Process Capability, Release 3.0. Bell Canada. Internet edition, 1999.

Cockburn, A., 2002. Agile Software Development. Addison Wesley, Boston, MA.

Constantine, L.L., Lockwood, L.A.D., 1999. Software for Use. Addison-Wesley, New York.

DATech, 2002. DATech-Prüfbaustein, Usability-Engineering-Prozess, Version 1.2. Deutsche Akkreditierungs-stelle Technik e.V., Frankfurt/Main p. 70.

Earthy, J., 1996. Development of the Usability Maturity Model. IE2016 INUSE Deliverable D5.1.1(t). Lloyd's Register, London.

Earthy, J., 1997. Usability Maturity Model: Processes. INUSE/D5.1.4(p), EC INUSE (IE 2016) Final Deliverable (Version 0.2). Lloyd's Register, London.

Earthy, J., 1998a. TRUMP Version Usability Maturity Model: Processes. Version 2.1. Lloyd's Register of Shipping, London, p. 85.

Earthy, J., 1998b. Usability Maturity Model: Human Centredness Scale. INUSE Project deliverable D5.1.4(s). Version 1.2. Lloyd's Register of Shipping, London.

Earthy, J.V., Sherwood-Jones, B., Bevan, N., 2001. The Improvement of human-centred processes—facing the challenge and reaping the benefit of ISO 13407. International Journal of Human Computer Studies 2001;.

Ease, of Use Roundtable, 2004. Why the PC Industry Must Improve New Technology's Quality and Ease of Use 2004, http://www.eouroundtable.com/.

Eason, K., Harker, S.D., 1997. User Centred Design Maturity. Internal working Document. Department of Human Sciences, Loughborough University, Loughborough.

Flanagan, G.A., 1995. in: Katz, I., Mack, R., Marks, L., Rosson, M.B., Nielsen, J. (Eds.), Usability Leadership Maturity Model (Self-assessment Version). Delivered in a Special Interest Group session, CHI '95: Human Factors in Computing Systems.

Flanagan, G.A., Rauch, T.L., 1995. Usability management maturity. Part 1. Self assessment—how do you stack up? CHI Companion 95, Conference on Human Factors in Computing Systems. ACM, Denver.

Flanaghan, G.A., 1996. Usability management maturity. Part 1. Self assessment—how do you stack up?. SIGCHI Bulletin 28 (4).

Gould, J.D., Boies, S.J., Levy, S., Richards, J.T., Schoonard, J., 1987. The 1984 Olympic message system: a test of behavioral principles of system design. Communications of ACM 30 (9), 758–769.

Hackos, J.T., Redish, J.C., 1998. User and Task Analysis for Interface Design. Wiley, New York.

Harning, M.B., Vanderdonckt, J., 2003. Closing the gaps: software engineering and human–computer interaction. Interactions 2003.

Hix, D., Hartson, H.R., 1993. Developing User Interfaces: Ensuring Usability Through Product and Process. Wiley, New York.

ISO/IEC, 1998a. 15504-2 Software Process Assessment. Part 2. A Reference Model for Processes and Process Capability, ISO/IEC TR 15504-2: 1998 (E) 1998.

ISO/IEC, 1998b. 9241-11 Ergonomic Requirements for Office Work with Visual Display Terminals (VDT)s. Part 11. Guidance on Usability, ISO/IEC 9241-11: 1998 (E) 1998.

ISO/IEC, 1999. 13407 Human-Centred Design Processes for Interactive Systems, ISO/IEC 13407: 1999 (E) 1999.

ISO/IEC, 2000. 18529 Human-centred Lifecycle Process Descriptions, ISO/IEC TR 18529: 2000 (E) 2000.

ISO/IEC, 2003. 18152 A Specification for the Process Assessment of Human–System Issues, ISO/PAS 18152: 2003 2003.

John, B.E., Bass, L., Kazman, R., Chen, E., 2004. Identifying Gaps between HCI, Software Engineering and Design, and Boundary Objects to Bridge Them, CHI 2004.

Jokela, T., 2001a. An assessment approach for user-centred design processes Proceedings of EuroSPI 2001. Limerick Institute of Technology Press, Limerick.

Jokela, T., 2001b. Assessment of User-Centred Design Processes as a Basis for Improvement Action. An Experimental Study in Industrial Settings. Oulu University Press, Oulu.

Jokela, T., 2002. Making User-Centred Design Common Sense: Striving for an Unambiguous and Communicative UCD Process Model NordiCHI 2002. ACM, Aarhus, Denmark.

Jokela, T., Abrahamsson, P., 2004. Usability assessment of an extreme programming project: close co-operation with the customer does not equal to good usability. Profes 2004; 2004.

Jokela, T., Iivari, N., Nieminen, M., Nevakivi, K., 2001. Developing a usability capability assessment approach through experiments in industrial settings Joint Proceedings of HCI 2001 and IHM 2001. Springer, London.

Kurosu, M., Ito, M., Horibe, Y., Hirasawa, N., 2000. Diagnosis of human-centeredness of the design process by the SDOS Proceedings of UPA 2000. Usability Professionals Association, Asheville, NC.

Kuutti, K., Jokela, T., Nieminen, M., Jokela, P., 1998. Assessing human-centred design processes in product development by using the INUSE maturity model. Proceedings of the Seventh IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man–Machine Systems—MMS'. IFAC, Kyoto, Japan.

Larsen, M.A., Myers, M.D., 1999. When success turns into failure: a package-driven business process re-engineering project in the financial services industry. Journal of Strategic Information Systems 8, 395–417.

Mayhew, D.J., 1999. The Usability Engineering Lifecycle. Morgan Kaufman, San Fancisco.

Myers, M.D., 1997. in: Stowell, F. (Ed.), Interpretive Research Methods in Information Systems. Information Systems: An Emerging Discipline. McGraw Hill, London, pp. 239–266.

Myers, B., Ioannidis, Y., Hollan, J., Cruz, I., Bryson, S., Bulterman, D., Catarci, T., Citrin, W., Glinert, E., Grudin, J., 1996. Strategic directions in human–computer interaction. ACM Computing Surveys (CSUR) 28 (4), 794–809.

Nielsen, J., 1993. Usability Engineering. Academic Press, San Diego.

Norman, D.A., Draper, S., 1986. User Centered System Design. Erlbaum, Hillsdale, NY.

Paulk, M., Weber, C., Carcia, S., Chrissis, M., Bush, M., 1995. The Capability Maturity Model: Guidelines for Improving the Software Process. Addison-Wesley, New York.

Rosenbaum, S., Rohn, J., Humburg, J., Bloomer, S., Dye, K., Nielsen, J., Rinehart, D., Wixon, D., 1999. What Makes Strategic Usability Fail? Lessons Learned from the Field. A Panel CHI 99 Extended Abstracts, Pittsburgh, USA. ACM, New York.

Rosenbaum, S., Rohn, J.A., Humburg, J., 2000. A toolkit for strategic usability: results from workshops, panels, and surveys CHI 2000 Conference Proceedings, The Hague. ACM, New York.

Rosson, M.B., Carroll, J.M., 2002. Usability Engineering. Scenario-based Development of Human–Computer Interaction. Morgan Kaufmann, San Fancisco.

Taylor, B., Gupta, A., Hefley, W., McLelland, I., Van Gelderen, T., 1998. HumanWare process improvement—institutionalising the principles of user-centred design Tutorial PM14 H Human-centred Processes and their Impact, Human–Computer Interaction Conference on People and Computers XIII, Sheffiel Hallam University 1998.

Walsham, G., 1995. Interpretive case studies in IS research: nature and method. European Journal of Information Systems 4, 74–81.

Wilson, S., Bekker, M., Johnson, P., Johnson, H., 1997. Helping and hindering user involvement. Involvement: a tale of everyday design Proceedings of CHI '97. ACM Press.

Vredenburg, K., Isensee, S., Righi, C., 2002. User-Centred Design. An Integrated Approach. Prentice Hall, Upper Saddle River.