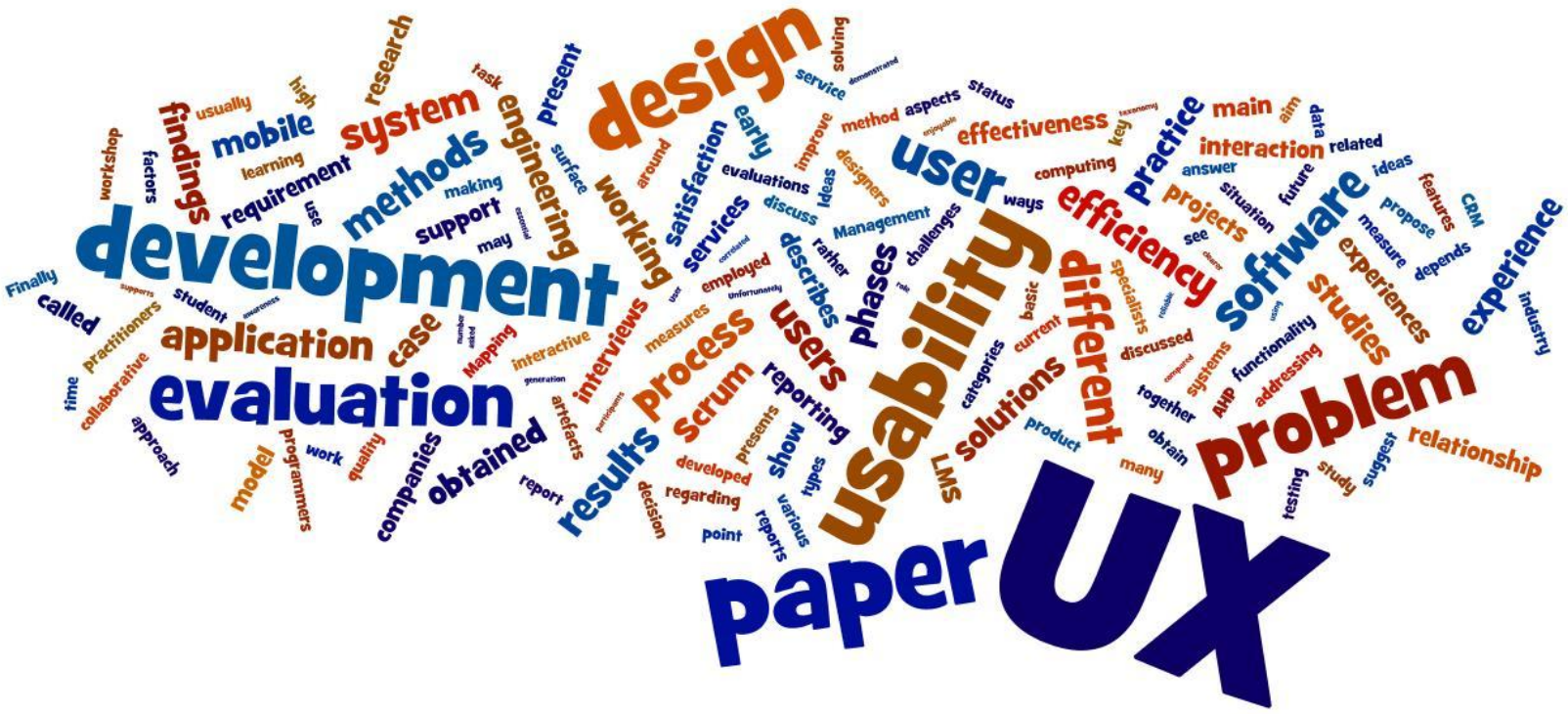
International Workshop on the Interplay between User Experience (UX) Evaluation and System Development (I-UxSED 2012)

NordiCHI 2012, Copenhagen, Denmark



Effie Lai-Chong Law

University of Leicester, UK/ ETH Zürich, Switzerland

Silvia Abrahão

Technical University of Valencia, Spain

Arnold P.O.S. Vermeeren

Delft University of Technology, the Netherlands

Ebba Thora Hvannberg

University of Iceland, Iceland

Title: Proceedings of International Workshop on the Interplay between User Experience Evaluation and Software Development (I-UxSED 2012)

Editors:

Effie L-C Law, Silvia Abrahão, Arnold, P.O.S. Vermeeren, and Ebba Thora Hvannberg

Printed by: Audio Visual Services (AVS), University of Leicester, UK

ACKNOWLEDGEMENTS

First of all, we are very grateful to the NordiCHI 2012 workshop chairs, Dr. Irina Shklovski and Dr. Joseph (Jofish) Kaye, for accepting our proposal. Thanks must also go to the authors of the workshop's papers, whose contributions serve as rich sources of stimulation and inspiration to explore the theme of the workshop from multiple perspectives. The quality of the contributions could further be ensured and improved with the generous help of the program committee members and additional reviewers. Their effective and efficient review works are highly appreciated.

List of Program Committee members:

- Asbjørn Følstad, SINTEF, Norway
- Gilbert Cockton, Northumbria University, UK
- Jan Stage, Aalborg University, Denmark
- Marianna Obrist, Newcastle University, UK
- Mark Springett, Middlesex University, UK
- Marta Kristín Lárusdóttir, Reykjavik University, Iceland
- Natalia Juristo, Universidad Politécnica de Valencia, Spain
- Rory O'Connor, Dublin City University, Ireland
- Virpi Roto, Aalto University, Finland

Last but not least, we express gratitude to our sponsor – COST (European Cooperation in the field of Scientific and Technical Research; http://cost.cordis.lu/src/home.cfm). The COST Office operated by the European Science Foundation (ESF) provides scientific, financial and administrative support to COST Actions. The COST Action IC0904 (http://www.twintide.org), which is commonly known as **TwinTide (Towards the Integration of Trans-sectorial IT Design and Evaluation)**, was officially launched in November 2009. The ultimate goal of the Action is to integrate and foster research and practice on design and evaluation methodologies for computing artefacts across sectors and disciplines. The current Workshop is the first open workshop implemented under the auspices of TwinTide. As with this and forthcoming events of TwinTide, we aim to provide the participants with enlightening environments to further deepen and broaden their expertise and experiences in the area of user experience and software development.

TABLE OF CONTENTS

Interplay between User Experience Evaluation and System Development: State of the Art

COST IC0904 TwinTide Open Workshop

Effie Lai-Chong Law
Department of Computer Science
University of Leicester
UK
elaw@mcs.le.ac.uk

Silvia Abrahão
Department of Computer Science
Technical University of Valencia
Spain
sabrahao@dsic.upv.es

Arnold P.O.S. Vermeeren
Faculty of Industrial Design Engineering
Delft University of Technology
The Netherlands
A.P.O.S.Vermeeren@tudelft.nl

Ebba Thora Hvannberg
School of Engineering and Natural Sciences
University of Iceland
Iceland
ebba@hi.is

ABSTRACT
User Experience (UX) is an emerging research area pertaining to as well as extending beyond the traditional usability. Issues in the realm of usability may be amplified in UX because of its larger scope. Four key non-orthogonal issues are: definition, modeling, method selection, and interplay between evaluation and development. Leveraging the legacy of a series of earlier workshops, I-UxSED 2012 aims to develop a deeper understanding of how evaluation feedback shapes software development, especially when experiential qualities such as fun, trust, aesthetic values are concerned. Is feedback on these fuzzy qualities less useful for problem prioritization or less persuasive for problem fixing? This and other challenging questions will be explored in I-UxSED 2012 that brings together researchers and practitioners from two communities - HCI and Software Engineering.

Author Keywords
User experience; Usability; Software development; Interaction design; Downstream utility; Interplay

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Evaluation; Measurement

BACKGROUND AND MOTIVATION
Leveraging the legacy of a series of successful workshops ([1] [2] [3]) that brought together people from Human-Computer Interaction (HCI) and Software Engineering (SE) communities to discuss the interplay between software evaluation and development, the proposed I-UxSED 2012 is further inspired by more recent insights into the issues pertaining to traditional usability (e.g. [4]) as well as the emerging User Experience (UX) (e.g. [5], [6]).

The shift of emphasis in the field of HCI from usability engineering to a much richer scope of user experience where users' emotions, affects, motivations, and values are given as much, if not more, attention than ease of use, ease of learning and basic subjective satisfaction [7]. Among others, four challenges engendered by the new focus of UX are particularly relevant to software development: (i) definition of UX; (ii) modelling of UX; (iii) selection of UX evaluation methods; (iv) interplay between UX evaluation feedback and software development.

The concept of UX is commonly understood as subjective, context-dependent and dynamic [7]. A "formal" definition of UX issued by ISO 9241-210: 2010 - A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service – is ambiguous and needs to be refined.

In contrast to usability, UX metrics are yet to be defined. The task is related to ongoing debates on the measurability of experiential qualities [8]. Both usability and UX measures should enable professionals to benchmark competitive design artefacts and to select right design options. The intriguing question is whether the respective measures have different persuasive power and impact on (re)design and development.

Modelling users' experiences is especially important for understanding, predicting and reasoning about processes of UX with consequences for software design. However, a number of issues pertaining to UX modelling remain to be resolved [9].

Recently, research efforts have been undertaken to collect, consolidate and categorize UX evaluation methods (e.g. [10]). It is envisaged that taxonomies of UX qualities, which can facilitate the selection of UX methods and

measures, will come to fruition from these ongoing endeavours.

The first three issues have significant impacts on their fourth one – the theme of I-UxSED 2012 - is only explored to a limited extent.

WORKSHOP GOALS AND THEMES

We understand the relationship between UX and usability as the latter is subsumed by the former. Usability evaluation methods (UEMs) and metrics are relatively more mature [11]. In contrast, UX evaluation methods (UXEMs) which draw largely on UEMs [12] are still taking shape. It is conceivable that feeding outcomes of UX evaluation back to the software development cycle to instigate the required changes can even be more challenging than doing so for usability evaluation (UE). It leads to several key issues:

- UX attributes are (much) more fuzzy and malleable, what kinds of diagnostic information and improvement suggestion can be drawn from evaluation data. For instance, a game can be perceived by the same person as a great fun on one day and a terrible boredom the following day, depending on the player's prevailing mood. The waning of novelty effect (cf. learnability differs over time in case of usability) can account for the difference as well. How does the evaluation feedback enable designers/developers to fix this experiential problem (cf. usability problem) and how can they know that their fix works (i.e. downstream utility)?
- Emphasis is put on conducting UE in the early phases of a development lifecycle with the use of low fidelity prototypes, thereby enabling feedback to be incorporated before it becomes too late or costly to make changes [13]. However, is this principle applicable to UX evaluation? Is it feasible to capture authentic experiential responses with a low-fidelity prototype? If yes, how can we draw insights from these responses?
- The persuasiveness of empirical feedback determines its worth. Earlier research (e.g. [14]) indicates that the development team needs to be convinced about the urgency and necessity of fixing usability problems. Is UX evaluation feedback less persuasive than usability feedback? If yes, will the impact of UX evaluation be weaker than UE?
- The Software Engineering (SE) community has recognized the importance of usability. Efforts are focused on explaining the implications of usability for requirements gathering, software architecture design, and the selection of software components [15]. Can such recognition and implications be taken for granted for UX, as UX evaluation methodologies and measures could be very different (e.g. artistic performance)?
- How to translate observational or inspectional data into prioritised usability problems or redesign proposals is thinly documented in the literature [4]. Analysis approaches developed by researchers are applied to a limited extent by practitioners [4]. Such divorce between research and practice could be bitterer in UX analysis approaches, which are essentially lacking.

While the gap between HCI and SE with regard to usability has somewhat been narrowed (e.g. [1]. [2]), it may be widened again due to the emergence of UX.

The main goal of I-UxSED 2012 is to bring together people from HCI and SE to identify challenges and plausible resolutions to optimize the impact of UX evaluation feedback on software development.

RELEVANCE TO THE FIELD

The main contribution of I-UxSED 2012 to the field of HCI and SE is the understanding of state-of-the-art about the interplay between UX evaluation feedback and system development. Specifically, there are limited studies investigating how different UX evaluation feedback formats such as textual (e.g. diary), audio (e.g. interview), visual (e.g. pictorial scale) and physiological (e.g. eye-tracking) determine their usefulness as well as persuasiveness. Besides, visual and physiological data are more commonly used in UX than in usability, based on the observations that experiences are more difficult to verbalize and more subjective. The role of such evaluation data in system redesign entails further exploration. Besides, there are very few methodological and practical guidelines on integrating UX evaluation and system design in a software development process. The workshop will heighten the awareness of the need for more research studies on the above-mentioned issues.

CONTRIBUTIONS

Eleven quality contributions have been accepted. They are categorized into four groups:

- Domain-specific design and evaluation case study (Winckler et al. on e-citizen, Panayiotis et al on e-learning, Nilsson & Følstad on emergency services)
- Models on usability and UX evaluation (Oliveria et al on customer satisfaction, Sikorski on customer relationship, and Srđević et al on decision-making)
- Agile and UX practice (Lárusdóttir et al on UX role in scrum, Lindell on design-driven organization, and Jokela on the role of evaluation in UX)
- Attitudes towards and awareness of UX (Ardito on UX practice in companies; Law and Schaik on attitudes towards UX measurement)

In-depth discussions in the workshop can shed light on these aspects with regard to the interplay between UX evaluation and software development. Future research challenges along this inquiry will be identified.

REFERENCES

[1] Abrahão, S., Law, E. L-C., Stage, S., Hornbæk, K., Juristo, N. (2008). *Proceedings of the First Workshop on the Interplay between Usability Evaluation and Software Development*, Pisa, Italy, September 24, 2008, CEUR-WS.org 2008

[2] Abrahão, S., Hornbæk, K., Law, E. L-C., Stage, J. (2009). Interplay between Usability Evaluation and Software Development (I-USED 2009). *In Proceedings of INTERACT (2) 2009:* 969-970

[3] Law, E. L-C., Abrahão, S., Stage, J. (2010). *Proceedings of the First International Workshop on the Interplay between User Experience and Software Development (I-UxSED 2010) (*CEUR-WS.org 2010), NordiCHI 2010, Iceland.

[4] Følstad, A., Law, L-C., & Hornbæk, K. (2012): Analysis in practical usability evaluation: a survey study. In *Proc. CHI 2012* (pp. 2127-2136).

[5] Bargas-Avila, J. A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proc. CHI 2011* (pp. 2689-2698).

[6] Hassenzahl, M. (2011): User Experience and Experience Design. In: M. Soegaard, & R. F. Dam, (eds.), Encyclopedia of Human-Computer Interaction". http://www.interaction-design.org/encyclopedia/user_experience_and_experience_design.html

[7] Law, E. L-C., Roto, V., Hassenzahl, M., Vermeeren, A., & Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach. In *Proc. CHI 2009* (pp. 719-728).

[8] Law, E. L-C. (2011). The measurability and predictability of user experience. In *Proc. EICS 2011* (pp. 1-10).

[9] Law, E. L-C., & van Schaik. P. (2010). Modelling user experience – an agenda for research and practice. *Interacting with Computers, 22*(5):313–322.

[10] Vermeeren, A. P.O.S. Law, E. L-C., Roto, V., Obrist, M., Hoonhout, J., Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In *Proc NordiCHI 2010* (pp. 521-530).

[11] Law, E. L-C., Hvannberg, E., & Cockton, G. (eds.) (2008). *Maturing usability: Quality in software, interaction and value*. Springer.

[12] Tullis, T., & Albert, W. (2008). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Interactive Technologies)*. Morgan Kaufman.

[13] Hertzum, M. (2006). Problem Prioritization in Usability Evaluation: From Severity Assessments Toward Impact on Design. *International Journal Human Computer Interaction 21*(2): 125-146.

[14] Nørgaard, M., & Hornbæk, K (2009). Exploring the Value of Usability Feedback Formats. *International Journal Human Computer Interaction 25*(1), 49-74.

[15] Juristo, N., Moreno, A. M., Sanchez-Segura, M. (2007). Guidelines for Eliciting Usability Functionalities. *IEEE Transactions on Software Engineering, 33* (11), 744-758.

Ideas Mapping, Surface Computing and User Experience

Panayiotis Zaphiris, Andri Ioannou, Fernando Loizides, Christina Vasiliou
Cyprus University of Technology
Limasol, Cyprus
panayiotis.zaphiris@cut.ac.cy, andri.ioannou@gmail.com, fernando.loizides@gmail.com, vasi.christina@gmail.com

ABSTRACT

This paper reports work regarding the design, development and evaluation of a surface computing application to support collaborative decision making.. The domain-independent application, so called *Ideas Mapping*, builds on the principle of Affinity Diagramming to allow participants to analyze a problem and brainstorm around possible solutions while they actively construct a consensus artifact -- a taxonomy of their ideas. During idea generation, *Ideas Mapping* replicates physical post-it notes on a multi-touch tabletop. Additional functionality supports student collaboration and interaction around the organization of ideas into thematic categories associated with the problem at hand. We report on the functionality and user experience while interacting with the application which was designed and developed using a user-centered approach. We also report initial findings regarding the affordances of surface computing for collaborative decision making.

Author Keywords
surface computing, user experience, collaborative decision making, CSCL

ACM Classification Keywords
K.3.1 Computer Uses in Education: Collaborative learning

General Terms
Design, Human Factors

INTRODUCTION
A multi-touch interactive tabletop can support collaboration, allowing different patterns of turn taking, negotiation and interaction [5, 2]. In this paper we report the design, development and evaluation of a surface computing application that supports idea generation, collaborative decision making and group artifact construction. The paper starts by covering related research literature and continues with the description of the design and development of *Ideas Mapping* and its use in two studies aiming to understand the affordances of surface computing for collaborative decision making. The paper concludes with a discussion of the key findings and makes suggestions to researchers and practitioners.

EXISTING KNOWLEDGE
The work reported in this paper draws from literature in the areas of Human Computer Interaction (HCI) methods (Affinity Diagramming) and Multi-touch interactive surfaces. The current state of the art in these areas is briefly summarized in this section.

Affinity Diagramming
HCI techniques exist to facilitate discussion in groups and to extract ideas from users' initial conceptual models. For example, the Kawakita Jiro diagrammatic method [8], also known as Affinity Diagramming, is a team-based knowledge elicitation technique. It is used for grouping information into categorical domains [10] and bears similarities to open card sorting. Users write down items of knowledge or descriptions on sticky notes and then organize the notes into groups before creating group headings. These methods are useful to HCI specialists as techniques for creating and analyzing categorizations of knowledge and are considered among the foremost usability methods for investigating a user's (and groups of users') mental model of an information space [9]. In affinity diagramming, the method is enforced in teams usually working on a shared whiteboard or large piece of paper. They are encouraged to communicate their reasoning verbally; thus, collaborative team decisions upon consensus lead to category cluster formation [1].

Multi-touch Interactive Tabletops

Multi-touch interactive tabletops have recently attracted the attention of the HCI and Computer Supported Collaborative Learning (CSCL) communities. Based on preliminary evidence from the education and computer-science literature, Higgins et al. [7] provide a review of the technological characteristics of multi-touch interactive tabletops and their pedagogical affordances. Overall, as pointed out by Higgins et al. [7], most of what we know in this area concerns technical issues related to interaction of users with the technology, but we know little about the use and value of multi-touch tabletops on collaborative learning situations within formal educational settings. Below we summarize some recent empirical evidence related to multi-touch tabletops and learning.

Multi-touch tabletops have been used with disabled user groups to promote development of social skills. SIDES, for example, is a four-player cooperative computer game designed to support adolescents with Asperger's syndrome to practice social skills and effective group work during their group therapy sessions [11]. SIDES provided an engaging experience for this audience who remained engaged in the activity the entire time and learned from the activity (unlike typical behavior of this population) [11]. Similarly, StoryTable has been used to facilitate collaboration and social interaction for children with autistic spectrum disorder with positive effects [4]. StoryTable was initially designed to support children's storytelling activity in groups [3]; Evaluation of StoryTable showed that it enforced cooperation between children during the storytelling activity, by allowing simultaneous work on different tasks, while forcing them to perform crucial operations together in order to progress [3]. In some other work, multi-touch tabletops have been studied for their added benefits compared to single-touch tabletops. Harris et al. [5] contrasted groups of children in multi-touch and single-touch conditions and found that children talked more about the task in the multi-touch condition while in the single-touch condition; they talked more about turn taking. Furthermore, the technology is considered engaging For example, the overall (perceived) usefulness and benefit of using interactive tabletops in collaboration contexts was assessed in a recent experiment by [2] with 80 participants. That study showed that groups in the tabletop condition had improved subjective experience and increased motivation to engage in the task.

With regards to using tabletops in formal learning settings, a series of studies are currently being conducted as part of the SynergyNet project [7]. SynergyNet goes beyond using single tables to studying a network of tabletops that can communicate with each other. SynergyNet focuses on how this technology can best support collaboration within small groups, while undertaking the development of curricula and tabletop applications for classroom integration [7]. A recent SynergyNet study contrasted groups of children in multi-touch and paper-based conditions to examine the differences in their collaborative learning strategies [7]. The authors found that student groups in the multi-touch condition maintained better joint attention on the task than groups in the paper-based condition. Another recent SynergyNet study examined NumberNet, a tool designed to promote within and between group collaboration in a mathematic classroom using a network of tabletops [6]. In this study, pilot results from 32 students showed significant knowledge gains from pre to post testing.

DESIGN METHODOLOGY

We adopted a strongly user-centered approach, emphasizing the engagement of students and instructors in all phases of the design process. Four university students and three instructors were involved, contributing to design elements of the application.

First, through low-fidelity paper-based prototypes, we simulated a collaborative activity with four students around a (turned-off) tabletop using paper and pencil. The scenario involved "the creation of a computer games industry in Cyprus and the factors involved." First, students generated ideas individually for 10 minutes. They wrote a (physical) post-it note for each new idea. Next, the ideas appeared one-by-one on the table and became subject to discussion, after a brief explanation from their originator, in an effort to categorize them in thematic units. Students revisited and changed ideas, rejected less promising ones, and generated new ideas during a collaborative decision making process leading to their thematic categorization. Finally, the activity concluded with a consensus of the main factors (i.e., resulting thematic categories) involved in the creation of a computer games industry in Cyprus. After the completion of the activity, instructors (who observed and kept records of all interactions during the activity) and students discussed the potential surface computing application and contributed to elements of the design from their own viewpoints.

Following the low-fidelity design discussions and analysis of user needs, a prototype Beta version application was developed in Action Script 3.0, for a multi-touch tabletop, the MagixTable. The application, so called *Ideas Mapping,* was designed to be domain-independent with a mild learnability curve. Our participants were called back to collaborate on different scenarios using *Ideas Mapping* and provide feedback on its user experience and further suggestions for improvement. Evaluation sessions took place in a fully equipped usability lab and all sessions were video recorded and analyzed. *Ideas Mapping* was optimized and finalized in three major iterative cycles of design, development and evaluation.

OVERVIEW OF THE APPLICATION

Overall, *Ideas Mapping* is designed to support idea generation, collaborative decision making and group artifact construction. The application builds on the principle of Affinity Diagramming to allow participants to analyze a problem and brainstorm around possible solutions while they actively construct a consensus artifact; namely, a taxonomy of their ideas. This is done in three stages:

Stage 1: With a scenario at hand, each collaborator generates new ideas. Ideas are typed into a web application (producing an XML file associated with *Ideas Mapping*) through the use of a mobile device (laptop, tablet, smartphone connected to the Internet). The need for the integration of mobile devices and a web application emerged from a constraint imposed by the MagixTable (also true for other platforms such as the MS Surface) -- that text entry can be done from one pre-existing keyboard at a time. For the kind of activity we sought, this constraint

would be significant. To resolve this problem, we developed four virtual keyboards on the tabletop (one for each user). However, users experienced difficulties typing extended ideas on the virtual keyboard during stage 1; the keyboard interaction suffered from input latency and mistyping issues. Thus, the use of mobile devices for input via a web application was considered as a practical solution to this problem for stage 1. This problem demonstrates both the still existing technical limitations of tabletops but also the importance of user input in developing applications for such technologies.

Stage 2: Next, the ideas are presented one-by-one, as digital post-it notes in the middle of the tabletop surface and become subject to discussion amongst the collaborators. For each idea, collaborators make an effort to categorize it in a thematic unit. Functionalities include:

- Each post-it note must be categorized before the next one appears. If controversy exists, an idea can be placed in the "Decide Later" depository to be revisited upon the categorization of other ideas. Post-it notes are automatically oriented to face their contributor, which encourages them to elaborate on the idea. This functionality was implemented as a result of users' feedback and is consistent with previous work by [12] showing that orientation can play an important role in collaborative interactions around tabletops by signifying ownership and directing attention.
- Thematic units can be created by any participant using the virtual keyboard. Once a participant begins the categorization of an idea (e.g., either begins to type a thematic unit or simply touches the post-it note), others must wait as only one keyboard is presented at any given time. Thematic units can be renamed if needed.
- Participants can drag and drop a post-it note over a thematic unit to categorize it. Post-it notes can be manipulated in order to move them across the surface, rotate and resize them.
- In this stage participants cannot edit ideas, or generate new ideas notes, and thematic units cannot be deleted. These design decisions aimed to scaffold the collaborative activity by allowing time for learners to consider all contributed ideas before making significant decisions.

Stage 3: In this last stage, more flexibility is given to the participants to finalize their taxonomy. In addition to the above, users can now edit ideas or generate new ones, delete ideas or thematic units that are less promising, and reallocate ideas into thematic units for a better fit. Overall, students engage in a collaborative decision making process, leading to the construction of a group artifact -- a taxonomy of their ideas.

STUDIES WITH IDEAS MAPPING

To examine the affordances of surface computing for collaborative decision making two studies were conducted with groups of university students: a small pilot study and a larger scale investigation.

THE PILOT STUDY

Participants and Setting:

Four university students, aged between 22-27 years old, were recruited to participate in a short activity around the tabletop. The scenario involved the "creation of an action plan that can improve university students' experiences at the Cyprus University of Technology, including social and educational aspects." The session was video recorded and analyzed.

Video Analysis and Preliminary Findings

An exploratory approach was used to trace the kinds of interactions amongst the collaborators and the technology and to better understand the role of tabletops in supporting learning. General research questions guided our video analysis such as: what kinds of interactions take place around the tabletop? and what evidence is present regarding the value of multitouch interactive tabletops for collaborative decision making?

One of the researchers considered the video corpus in its entirety – a total of 57 minutes. Most interaction occurred during the 2nd and 3rd stages of Ideas Mapping, which became the focus of the analysis. The researcher repeatedly watched the video, marked segments of interest, and created transcripts, in an effort to categorize the types of discourse and gestures used by the group members around the tabletop. A preliminary coding scheme is presented in Table 1. This coding scheme will be further refined as more studies are conducted in this context. Understanding collaborative decision making around tabletops is currently limited. It is thus important to establish a coding scheme of the interactions evident around this technology (particularly, the synergetic dialog and physical gestures) to be able to examine the phenomenon further. Ultimately, the coding scheme should help us examine interesting patterns of collaborative decision making around multi-touch interactive tabletops.

Overall, the pilot study provided initial evidence that the CSCL setting encouraged and stimulated discussion and physical interaction around shared artifacts.

Spoken Contributions

- Information Sharing – Defining/describing/identifying the problem
- Proposing – Proposing a thematic unit/new idea
- Elaborating – Building on previous statements, Clarifying
- Negotiating meaning – Evaluation of proposal, Questioning/ answering, Expressing agreement/disagreement, Providing arguments for/against
- Stating consensus – Summarizing ideas, Metacognitive reflections
- Other talk – Tool-related talk, Social talk, Laughter

Gesture Contributions

• Communicative Gestures – Show on the table without touching, Dominating/blocking gestures • Touch Gestures – Resize, Rotate, Type, Move something across, Random touching or touching to explore

Table 1: Preliminary Coding Scheme

LARGER SCALE INVESTIGATION

Participants

To further examine the value of multitouch interactive tables for collaborative decision making, we recruited postgraduate students in Cyprus to discuss a scenario related to peace. The sample was composed of 17 postgraduate students enrolled in a CSCL/CSCW course at a public university in Cyprus, aged between 22-45 years old (*M*=30).

The participants were divided into five groups: 3 groups of 3 students and 2 groups of 4 students, suitable for the four-sided tabletop. Group members were familiar with working together through other course learning activities. All, but one student, had no prior experience with using a multi-touch tabletop.

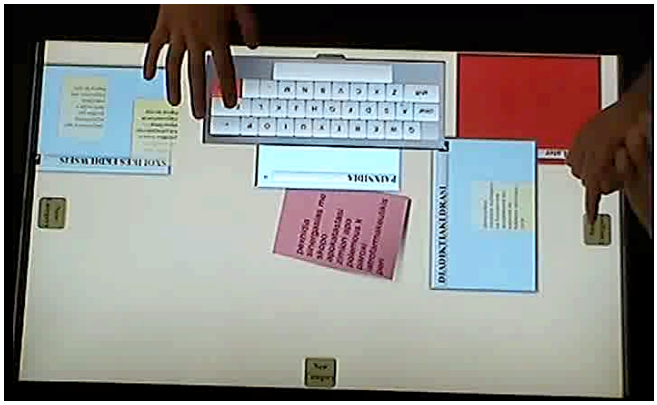


Figure 2: Categorization of ideas it in thematic units

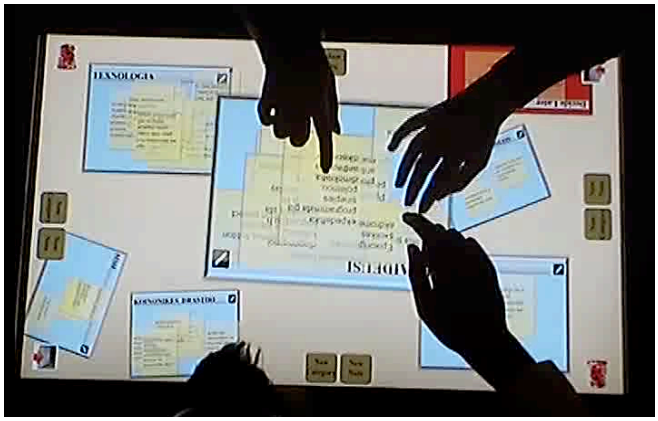


Figure 3: Consensus on a group artifact

Procedures

In this study there was a preparatory phase before students engaged in group work around the tabletop. That is, Stage 1 of *Ideas Mapping* was completed in distance, during the week before the tabletop investigation. The preparatory week aimed to allow students to research the scenario and think at their own pace. During the preparation week, students were tasked to investigate the topic, think creatively and record at least 10 ideas into the *Ideas Mapping* web application.

The following scenario was presented to the students: "Your team works at a non-governmental organization dealing with global peace. Your project is to create a roadmap of actions to promote global peace using technology." The specific scenario was chosen for it to be thought-provoking and without obvious answers to it. The goal was to stimulate critical thinking, dialog, and creative problem solving. Cyprus is a country in a long lasting political conflict. Thus, the topic was both personally important for the student participants, but also required their emotional and mental engagement.

The next phase involved collocated collaboration around the tabletop. Following the prep week, each group met face-to-face and engaged in collaborative work as described in Stages 2 and 3 of *Ideas Mapping*. Briefly, the ideas of each group were presented on the tabletop one-by-one. Students engaged in discussion and physical interaction with the tabletop in an effort to categorize the ideas in thematic units (i.e., taxonomy of ideas).

Data collection

The sessions of all five groups were video recorded for subsequent utterance coding and analysis. To complement the video data, a questionnaire was administered to all participants soon after the completion of the activity. The questionnaire aimed to assess students' perceptions of the collaborative learning experience and the usability of the surface computing application.

Video Analysis

An extensive video analysis of the data was carried out but due to the scope and space limitations of this paper the results will be presented elsewhere.

Quantitative Data Analysis

The questionnaire included 30 Likert-type items with a 7-point agreement response scale (from 1: completely disagree to 7: completely agree). These items measured three constructs of interest: (1) *Collaboration Support*, assessing the extent to which students thought the technology supported their collaboration such as, "The technology helped me work effectively in my group", "The technology met my needs as a collaborator"; (2) *Learning Experience*, assessing the extent to which students were satisfied with their learning experience overall, such as "Overall, my collaborative learning experience was positive", "I am satisfied with my experience through this activity", and (3) *Usability Satisfaction* (adapted from Lewis, 1995), assessing the extent to which students were satisfied with the usability of the system such as, "It was simple to use this system", "I can effectively complete my work using this system", "I like using the interface of this system".

A total of 17 students completed the questionnaire. First, the internal consistency for each subscale was assessed using Cronbach's alpha; all 3 subscales had acceptable internal consistency (Cronbach's alphas > .80). Then, subscale mean scores were calculated for every participant (i.e., an un-weighted composite score for each participant on each subscale) followed by computation of descriptive statistics. As shown in Table 2, means were well above the midpoint of the 7-point response scale for all three measures, suggesting that the technology was positively endorsed by the participants overall. Specifically, the participants thought the technology supported their collaboration (*M*=5.53, *SD*= .22), and were satisfied with their learning experience (*M*=5.77, *SD*= .51). With regards to the third measure, participants found the system usable overall (*M*=4.93, SD= .77), but individual item means pointed to some aspects which may need improvement. The rating average was lower for three particular items in this scale, suggesting that we should improve the way participants recover from mistakes ("The system gives error messages that clearly tell me how to fix problems" *M*=3:00 and "Whenever I make a mistake using the system, I recover easily and quickly" *M*=3.36), as well as extend the application to include more functionality ("This system has all the functions and capabilities I expect it to have" *M*=3.88).

The questionnaire also included an open-ended question concerning the pros and cons of using tabletops for collaborative learning activities. We reviewed students' responses to identify themes. Several students commented on how the tabletop promoted collaboration, helped them maintain attention to the task and was enjoyable to use. For example, one of the participants commented: *"The tabletop helped us collaborate and the resulting product was a group effort. It helps you pay attention. I also found it very enjoyable"*. Often, students pointed out the capabilities of the system that enabled effective collaboration, such as "*It was nice all of us could use the tools at the same time, to rotate a note, to make it larger to read, or to put it in the box to revisit later."* On the negative side, a few participants found the virtual keyboard difficult to use and that the system needed improvement in handling mistakes, which was consistent with the findings from the quantitative data. These results confirmed our views regarding the affordances of multi-touch tabletops to support collaboration activities and also contributed to further refinement of *Ideas Mapping*.

Subscale	# Items	Cronbach's Alpha	M (SD)
1. Collaboration Support	6	.94	5.53 (.22)
2. Learning Experience	5	.96	5.77 (.51)
3. Usability Satisfaction	19	.97	4.93 (.77)

Table 2: Subscales statistics and descriptive statistics (N=17)

DISCUSSION - CONCLUSION

This study reports on the functionality and user experience while interacting with a multitouch application which was designed and developed using a user-centered approach. We also report initial findings regarding the affordances of surface computing for collaborative decision making.

Ideas Mapping builds on the principle of Affinity Diagramming to allow participants to analyze a problem and brainstorm around possible solutions while they actively construct a consensus artifact -- a taxonomy of their ideas. We feel *Ideas Mapping* makes the Affinity Diagramming technique more collaborative. By allowing for an extension sorting activity, it provides a way for participants to negotiate around an emerging group artifact and make sense of challenging problems, such as how to promote world peace using technology.

We further have evidence that the CSCL setting of the study, and surface computing more generally, encouraged and stimulated dialog and collaborative work around an authentic problem . Following the individual generation of ideas, *Ideas Mapping* supported a 2-stage collaborative activity that promoted ideas sharing, negotiating, sorting and constructing a group artifact while coming to a consensus.

Moreover, we believe that traditional user experience evaluation methods (e.g. questionnaires) were useful for evaluating Ideas Mapping. However qualitative evaluation (e.g. video analysis and the establishment of a coding scheme) is also important; such methods can reveal

interesting patterns of interactions amongst the participants and with the technology beyond what is self-reported.

Below, we identify some implications of this work for future research and practice in the fields of HCI and CSCL.

Suggestions to Practitioners:

1. Designers should focus on engaging students and instructors in the design process of educational surfaces computing applications.
2. Current interactive tabletop technologies come with a lot of user interface limitations. These should be taken into account when designing applications for such surfaces.
3. The CSCL setting of the study encouraged and stimulated active dialogue with a problem at hand and a multitouch interactive tabletop application to support them.
4. Self-reported measures showed that students positively endorsed the use of multitouch interactive tabletops for small group work.

Suggestions to Researchers:

1. The proposed coding scheme can be applied and extended to more studies in the area.
2. New qualitative analysis methodologies for evaluating user experience are needed.
3. The role of surface computing in promoting dialogue around sensitive topics (like peace) is an interesting area for further research.
4. A framework for using surface computing for collaborative decision making in general (especially related to sensitive issues) can be developed and tested.

REFERENCES

1. Beyer, H., & Holtzblatt, K. (1998). Contextual design: Defining customer-centered systems. California: Morgan Kaufmann Publishers CA, USA.

2. Buisine, S., Besacier, G., Aoussat, A., Vernier, F. (in press). How do interactive tabletop systems influence collaboration? *Computers in Human Behavior*.

3. Cappelletti, A., Gelmini, G., Pianesi, F., Rossi, F., & Zancanaro, M. (2004). Enforcing cooperative storytelling: First studies. In *Proceedings of the 4th IEEE Intentional Conference in Advanced Learning Technologies*, Joensuu, Finland.

4. Gal, E., ,N. Goren-Bar, D., Pianesi, F., Stock, O., Zancanaro, M. & Weiss, P. (2009). Enhancing social communication of children with high-functioning autism through a co-located interface. AI & SOCIETY, 24 (1), 75-84, DOI: 10.1007/s00146-009-0199-0

5. Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., et al. (2009). Around the table: Are multipletouch surfaces better than single-touch for children's collaborative interactions? In *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning* (pp. 335–344). ISLS.

6. Hatch, A., Higgins, S., Joyce-Gibbons, A. & Mercier, E. (2011) NumberNet: Using Multi-touch Technology to Support Within and Between Group Mathematics Learning. In Proceedings of the 9th International Conference on Computer Supported Collaborative Learning (pp. 176-183). International Society of the Learning Sciences.

7. Higgins, S.E., Mercier, E.M., Burd, E., & Hatch, A., (2011) Multi-touch Tables and the Relationship with Collaborative Classroom Pedagogies: A Synthetic Review. International Journal of Computer-Supported Collaborative Learning. 6 (4), 515-538. DOI: 10.1007/s11412-011-9131-y

8. Kawakita, J. (1975). KJ method: A scientific approach to problem solving. Tokyo, Japan: Kawakita Research Institute.

9. Martin, S. (1999). Cluster analysis for Web site organization. Internetworking 2, 3. Retrieved February 29, 2012, from http://www.internettg.org/newsletter/dec99/cluster_analysis.html

10. Nielsen, J. (1994). Usability engineering. CA, USA: Morgan Kaufman.

11. Piper, A. M., O'Brien, E., Morris, M. R., & Winograd, T. (2006), SIDES: A cooperative tabletop computer game for social skills development. In *Proceedings of the 6th International Conference on Computer Supported Cooperative Work*, Alberta, Canada.

12. Wigdor, D., & Balakrishnan, R. (2005). Empirical investigation into the effect of orientation on text readability in tabletop displays. In Proceedings of Computer-Supported Cooperative Work (Vol. 18, p. 22). Springer.Retrieved 15.06.2010 from http://www.springerlink.com/index/p62165406wv725r4.pdf.

Are software companies aware of UX?

C. Ardito, P. Buono, M. F. Costabile, R. Lanzilotti
IVU Lab, Dipartimento di Informatica, Università di Bari Aldo Moro,
Via Orabona 4, 70125 Bari, Italy
{carmelo.ardito, paolo.buono, maria.costabile, rosa.lanzilotti}@uniba.it

ABSTRACT
The efforts of addressing user experience (UX) in product development keeps growing, as demonstrated by the proliferation of workshops and conferences bringing together academics and practitioners, who aim at creating interactive software able to satisfy their users. Unfortunately, human-centred design and methods addressing usability and UX are always mentioned in research papers but yet very seldom applied in the current practice of software development in industry. In this paper, some findings of studies we have recently performed with software companies are reported. They show that either companies still neglect usability and UX, or they do not properly address them. Thus, in this workshop that seems to consider UX evaluation as a usual practice and aims to optimize the impact of UX evaluation feedback on software development, our provocative statement is: Are software companies (at least) aware of UX? The studies summarized in this paper show that, in many cases, the answer is NO. We are working to overcome the current situation and the paper concludes by providing some suggestions to fill the gap between research and practice of UX.

Keywords
Software life cycle, human-centered design, ethnographic studies.

Categories and Subject Descriptors
H5.m. [**Information interfaces and presentation (e.g., HCI)**]: *Miscellaneous*; D2.10 [**Software**]: *Methodology*.

General Terms
Design, Human Factors.

INTRODUCTION
Designing for UX requires understanding user requirements from both a pragmatic (system functionalities and interaction) and a hedonic point view [16]. It is necessary to iteratively design and evaluate prototypes, according to the human-centered design (HCD) process [7]. Unfortunately, HCD and methods addressing usability and UX are always mentioned in research papers but yet very seldom applied in the current practice of software development. Our position is that, in order to successfully address interplay between UX evaluation and system development in current work practices, we still have to do a lot in order to make software companies aware of the importance of UX.

In this paper, we briefly report some findings of recent studies involving software companies. On the basis of such studies, we provide some indications for making UX an explicit goal of software developers, as well as suggestions on how to fill the gap between what Human-Computer Interaction (HCI) researchers propose about design and evaluation of UX and the activities performed by software engineers in their daily practices of software development.

ABOUT USABILITY ENGINEERING IN SOFTWARE COMPANIES
Our research group has been working for defining HCD techniques and methodologies that could be pragmatically integrated in the work activities of software developers. For example, in [4] it was proposed how to augment the standard waterfall life cycle to explicitly address usability issues; the Pattern-Based (PB) inspection reported in [9] has been defined in order to provide a cost-effective method that could satisfy the companies' need of effective and easy to use evaluation methods.

Despite the efforts of HCI researchers, HCD approaches are applied only to a limited extent by practitioners, as shown in [3], [6], [8], [10], [11], [14], [18], [19]. Such studies indicate that the main reasons why companies are reluctant to adopt HCD practices include: 1) time and costs of the HCD methods; 2) cultural prejudices; 3) lack of frameworks guiding the software development team in applying HCD methods. Some studies actually involved designers with a strong HCI background, and even HCD practitioners (e.g. [11], [18]). Thus, the situation is even worst when software engineers are addressed.

The above results have been confirmed in our recent survey reported in [1]. Specifically, we collaborated with colleagues of the Aalborg University to investigate the

NordiCHI '12, October 21-25, 2012, Copenhagen, Denmark.

practical impact of usability engineering in software development organizations in two different geographical areas in Europe, namely Northern Denmark and Southern Italy. The survey was conducted in order to identify possible obstacles that prevent organizations to take into account usability issues. It showed that the number of organizations conducting some form of usability activities is rather low. Even if software developers are becoming more aware of what usability is and of its importance in order to improve their products, one of the main problems still remains what we call "Developer mindset", i.e. many developers have their minds set more on programming aspects, technical challenges and functionality of the product than its usability. Still too many of them do not know well what usability is. Another main obstacle they report is the lack of suitable methods that could be integrated in their work practices without demanding a lot of resources. Software development companies do not consider involving final users during the requirement analysis and the evaluations activities. This pushes usability researchers and practitioners to deeply consider devoting more attention on how to transfer academic work into practical value for industry. As we said in [1], we believe "it is responsibility of academics to translate scientific articles, which formally describe evaluation methods, into something that makes sense for companies and it is ready to be applied".

HOW ETHNOGRAPHIC STUDIES MIGHT HELP?

As follow-up of the study in [1], we wanted to know more about the advantages and problems of usability engineering as perceived by individual organizations. We focused on companies whose software developers appeared to be motivated to improve the usability of the products they develop. The key question to be addressed is why such developers do not push for the adoption of usability engineering methods in their development processes. We also decided to consider ethnographically based research in order to get an in-depth understanding of the socio-technological realities surrounding everyday software development practice [5], [15] this should provide other indications on how to overcome obstacles to a wider account for usability engineering.

In this paper, we briefly report on a study we have performed in order to know more about the software development life cycle of a company of medium-high size. The study had two main objectives: 1) to view, capture and understand the work practice by employing observational methods and in-situ interviews; 2) to integrate HCD activities in key points of the software development life cycle, such as interviews and usage scenarios during the requirement analysis, as well as prototyping and evaluation during system design.

The study was conducted at a medium software company located in Southern Italy, which develops products in different domains, primarily public administration and bank. The company accounts three different Business Units (BUs): Public Administration, Finance, and Research. The latter is mainly involved in research projects. Each BU could be considered as a separate small company, with its own personnel for carrying out all the activities in the software life cycle: project leaders, analysts, designers, developers, BU managers, etc. All BUs adopt a traditional waterfall life-cycle model for several reasons, primarily management background and project constraints, which completely neglect usability and UX issues. The study has been carried out in the Public Administration and Research BUs. Two master students participated in the study, each one involved in the activities of a BU. Their work was part of their master thesis in HCI. They were in the company for a total of 120 working days. Specifically, Rossana, the student in the Public Administration BU, was assigned to a project for creating an application for tourists visiting a certain town, running on a mobile device; it was committed by the town municipality. Diego, the student in the Research BU, was assigned to a research project on "Technologies for Situational Sea Awereness", whose aim is to develop hardware and software to provide services to various people, from oceanography researchers to skippers, and others.

The details of the study and the analysis of the collected data will be described in another paper we are currently writing, and can be discussed at the workshop. We summarize here some findings, which were confirmed by the interviews to the BU managers, performed about a month after the end of Rossana's and Diego's work. As most important effect, they were surprised to see how effective and efficient the HCD methods that Rossana and Diego used were. Thanks to this experience, they finally understood that the minimal resources spent in the iterative prototyping were widely fulfilled by the obtained benefits.

The Research BU manager appreciated a lot the fact that Diego, in the requirement analysis, insisted a lot for including a detailed specification of user requirements. He did it and also performed semi-structured interviews to validate such requirements with other stakeholders. The manager actually understood how fruitful these activities were and how meeting other stakeholders helped resolving several concerns. Diego insisted for involving more real users, pointing out that how different final users are from other stakeholders in terms of needs and expectations, but this was not possible.

Both Rossana and Diego used paper prototypes a lot, discussing them in participatory meetings with other stakeholders, i.e. the other project partners in the case of Diego research project, while Rossana organized short

meetings with all designers. Because she was involved in the design of an application devoted to people visiting a certain town, she was able to involve a few other persons in the company (secretaries and staff members), who acted like tourists interacting with the prototypes. Even if the approach might appear a bit naïf, HCI researchers know how useful these "quick and dirty" methods might be. To test a running prototype with real users, Diego contacted two friend of him, who are professional skippers, and performed a thinking aloud test. They pointed out a feature that was not as useful as designers considered, and indicated some other problems.

After an analysis of various tools for rapid prototyping, Diego selected Justinmind Prototyper (http://www.justinmind.com/) and used it for creating several successive prototypes. The BU managers are now enthusiastic of this tool and are getting it to use in the early design phase. Rossana and Diego also performed several heuristic evaluations of the prototypes. Thus, they used methods that are very cost effective in order to demonstrate that methods that require limited resources and little training of company employees, who could perform them, actually exist.

What performed in the above study is in line with other works. For example, Jim Hudson states that a variety of methods have to be used at all phases of the product life cycle [5]. For example, in order to understand customer needs, the design team can choose from casual conversations to more formal focus groups. He also found very important discussing with small groups of customers on the paper prototypes once or twice each week. During these meetings, customers have to be observed during the interaction with a product prototype.

SUGGESTIONS

The ethnographic study confirmed how it is important to develop paper prototypes and to discuss them with other stakeholders, including end users. This is a first important suggestion for companies. It might appear that it is not a novel finding, but it is worth emphasizing that it is obvious within the research community, whereas the actual problem is to transfer the use of iterative prototyping in the practice of companies. With our ethnographic study, we provided evidence for the company of the advantages of informal meetings in which several stakeholders, including end users, analyse prototypes, starting from those on papers. This study and other previous experiences of ours on HCD in practice (e.g., see [2]), as well as other relevant work in literature [20], provide another important suggestion: running prototypes have to be evaluated with samples of their end users in a real context of use, since "end users can raise significant issues about system usability only when they get down to using the system, or even a running prototype, in their real activity settings". Only then, they are able to provide the right indications about what is working well and what is not. If this is true for usability, it is further true for UX, both because usability is part of UX and because the subjective aspects that UX impacts can be really assessed only by end users in real contexts of use.

In several interviews conducted with company managers as follow up of the study in [1], it emerged that another reason why companies neglect usability and UX is that such requirements are not considered in public tenders. In most of their work, company develop software systems committed by public organizations, which specify the system requirements in Call for Tenders. It is evident that the companies' interest is to satisfy all and only the requirements specified in the Call. Thus, another suggestion for changing the current situation is to convince such public organisations of the need of explicitly mentioning UX requirements in their Calls for Tenders. According to this, we are already in touch with people working at the office of the Apulia region (the region where our University is located), which is publishing in the last years several Call for Tenders about ICT systems, and we are discussing such issues. In trying to convince them to address UX, we are actually facing the lack of usability and UX requirements that are objectively verifiable; consequently, it is not easy to specify them in the Calls. HCI researchers are urged to find proper solutions to this problem.

Our last suggestion is that, once we succeed in getting companies aware of usability and UX, we try to satisfy their request of suitable methods requiring limited resources and help integrating them in their work practices. Current situation shows that this is still very challenging. Only a few scattered experiences of designing and evaluating UX in practice are reported in literature. For example, at Nokia, which has a long history in designing for experience, the product development process includes continuous evaluation of usability and UX in different phases of the life cycle. After the release on the market of the product, feedback is gathered from the field through controlled and uncontrolled studies in order to collect information for improving successive products [13].

Despite the effort spent by Nokia and some other companies in designing for and evaluating UX, there is yet no consensus on approaches and methods to be widely adopted in order to develop software systems able to provide users with pleasurable and satisfying experiences [12], [17]. We look forward to the discussions at the workshop, hoping that they might provide more insights.

ACKNOWLEDGMENTS

The research reported in this paper has been partially supported by: 1) EU COST Action IC 0904 TWINTIDE (Towards the Integration of Transectorial IT Design and Evaluation); 2) DIPIS (DIstributed Production as

Innovative System) Project, sponsored by Apulia Region. We are particular grateful to the companies which were involved in our studies, and to Rossana Pennetta and Diego Paladini who participated in the ethnographic study.

REFERENCES

[1] Ardito, C., Buono, P., Caivano, D., Costabile, M. F. Lanzilotti, R., Bruun, A., Stage, J. 2011. Usability evaluation: a survey of software development organizations. In *Proc. of SEKE 2011*. Knowledge Systems Institute, Skokie, Illinois, USA, 282-287.

[2] Ardito, C., Buono, P., Costabile, M.F., Lanzilotti, R., Piccinno, A., and Simeone, A.L. 2010. Analysis of the UCD process of a web-based system. In *Proc. of DMS 2010*. Knowledge Systems Institute, Skokie, Illinois, USA, 180-185.

[3] Bias, R.G., and Mayhew, D.J. 2005. *Cost-Justifying Usability*, 2nd Edition, Academic Press.

[4] Costabile M. 2001. Usability in the Software Life Cycle. In S.K. Chang. In *Handbook of Software Engineering and Knowledge Engineering*, 1, 179-192.

[5] Dittrich, Y., John, M., Singer, J. and Tessem, B. 2007. Editorial: For the Special issue on Qualitative Software Engineering Research. *Information and Software Technology*, 49, 6 (2007), 531-539.

[6] Hudson, J. 2008. Beyond Usability to User Experience. *Workshop UXEM at CHI 2008*, Available at: http://www.cs.tut.fi/ihte/ CHI08_workshop/papers.shtml. Last access on August 3rd, 2012.

[7] ISO/IEC 2010. 9241-210 *Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems*.

[8] Lallemand, C. 2011. Toward a closer integration of usability in software development: a study of usability inputs in a model-driven engineering process. In *Proc. of EICS '11*. ACM, New York, NY, USA, 299-302.

[9] Lanzilotti, R., Ardito, C., Costabile, M. F., De Angeli, A. 2011. Do patterns help novice evaluators? A comparative study. *International Journal of Human Computer Studies*, 69, 1-2, 52-69.

[10] Nielsen, J. 1993. *Usability Engineering*. San Diego, CA: Academic Press.

[11] Rosenbaum, S., Rohn, J.A., and Humburg, J. 2000. A toolkit for strategic usability: Results from workshops, panels, and surveys. In *Proc. of CHI'2000*. ACM, New York, NY, USA, 337-344.

[12] Roto, V., Law, E. L-C., Vermeeren, A., and Hoonh-out, J. User Experience White Paper. Available at: http://www.allaboutux.org/. Last access on August 3rd, 2012.

[13] Roto, V., Ketola, P., and Huotari, S. 2008. User Experience Evaluation in Nokia. *Workshop UXEM at CHI 2008.* Available at: http://www.cs.tut.fi/ihte/ CHI08_workshop/papers.shtml. Last access at on August 3rd, 2012.

[14] Seffah, A., Donyaee, M., Kline, R. B., and Padda, H. K. 2006. Usability measurement and metrics: A consolidated model. *Software Quality Journal*, 14, 2, 159-178.

[15] Sharp, H., deSouza C., and Y. Dittrich, 2010. Using ethnographic methods in software engineering research. In *Proc. of ICSE 2010*. ACM, New York, NY, USA, 491-492.

[16] Väänänen-Vainio-Mattila, K., Roto, V. and Hassenzahl, M. 2008. Now let's do it in practice: user experience evaluation methods in product development. In *Proc. of CHI* '08. ACM, New York, NY, USA, 3961-3964.

[17] Vermeeren, A. P.O.S. Law, E. L-C., Roto, V., Obrist, M., Hoonhout, J., Väänänen-Vainio-Mattila, K. 2010. User experience evaluation methods: current state and development needs. In *Proc. of NordiCHI 2010*. ACM, New York, NY, USA, 521-530.

[18] Venturi, G., and Troost. J. 2004. Survey on the UCD integration in the industry. In *Proc. of NordiCHI '04*. ACM, New York, NY, USA, 449-452.

[19] Vredenburg, K., Mao, J.Y., Smith, P.W., and Carey, T. 2002. A survey of user-centered design practice. In *Proc. of CHI '02*. ACM, New York, NY, USA, 471-478.

[20] Wagner, E.L., and Piccoli, G. 2007. Moving beyond user participation to achieve successful IS design. *Communication of ACM*, 50, 12, 51-55.

Influence of Usability on Customer Satisfaction: A Case Study on Mobile Phone Services

Rodrigo de Oliveira
Telefonica Research
oliveira@tid.es

Mauro Cherubini*
Google
mauroc@google.com

Nuria Oliver
Telefonica Research
nuriao@tid.es

ABSTRACT
Designing for better user experiences (*e.g.,* interactions more satisfying, enjoyable) is usually more difficult than aiming for clearer usability goals (*e.g.,* improve systems' efficiency, easy of use). In this paper, we present a conceptual model validated with data from 603 mobile phone users that clarifies the relationship between usability of basic mobile services and the users' satisfaction with them. Our findings indicate that satisfaction is mostly influenced by how users perceive the usability of these services, more specifically their efficiency. We discuss the model and propose three implications that shall increase satisfaction with basic mobile services: a few solutions to minimize routine disruption, personality-based service personalization, and persuasive strategies to raise awareness of one's technology consumption saturation point.

Author Keywords
Big Five; mobile phone services; personality; structural equation modeling; usability.

ACM Classification Keywords
H.1.2 Models and Principles: User/Machine Systems: Human Factors

INTRODUCTION
The Human-Computer Interaction community was once concerned primarily with usability, but has since become more interested in understanding, designing for and evaluating a wider range of user experience aspects. According to Rogers *et al.* [29], interactive systems should now be designed in terms of their objectives classified in terms of usability and user experience goals. Traditionally, usability goals are related to specific usability engineering criteria (*e.g.*, systems designed to be more efficient, effective, easy to use), whereas user experience goals aim to explain the nature of the user experience (*e.g.*, interactions more satisfying, enjoyable, engaging) [29]. Although usability goals are nowadays better established and integrated into Software Engineering, UX goals are still considered somewhat fuzzy, being their connection with usability goals even less clear. In this paper we focus on clarifying this connection in the context of mobile phone services, particularly between two key usability goals (*i.e.* efficiency and ease of use) and an important UX goal: user satisfaction. More specifically, we present findings of a conceptual model validated with data from 603 customers of a telecommunication operator that provides insights on the relationship between perceived usability of basic mobile phone services and their satisfaction with them. The model also captures the influence of other variables, such as the users' personality profile and their usage of mobile services. In the following sections we explain how the proposed model was empirically validated and discuss how designers and software engineers could leverage the model towards improving customers' satisfaction with basic mobile services.

*Research conducted while working for Telefonica Research.

CONCEPTUAL MODEL
The way people appropriate technology has been previously studied. Several theoretical models have been introduced and tested to explain user acceptance behavior, such as the *Theory of Reasoned Action* [15], the *Theory of Planned Behavior* [2] and the *Technology Acceptance Model* [11]. While these models have contributed a great deal to our understanding of users' preferences and acceptance behavior of technological artifacts, they fall short in explaining the users' experience with technology.

User experience encompasses the experiential, affective, and cognitive aspects of a person interacting with a product, system or service[1]. Therefore it is not limited to the user's intention to use a certain technology. However, user experience models do not typically capture the role of the user's personality when interacting with a certain piece of technology. Ryckman [30] defined personality as a "dynamic and organized set of characteristics possessed by a person that uniquely *influences* his or her cognitions, motivations, and behaviors in various situations". Recent studies have demonstrated that personality influences directly how people experience the world [28]. Hence, we believe that there is an opportunity to better understand the user's interaction with technology by taking into account his/her personality profile.

Personality profiles are typically assessed by means of surveys. Goldberg [17]'s Big Five model is one of today's most well-known, accessible—and of public domain—and empirically validated personality assessment models. It structures a personality profile into five factors (or traits): Extroversion,

[1]Adapted from `en.wikipedia.org/wiki/User_experience`, last retrieved September 2012.

Agreeableness, Conscientiousness, Emotional Stability, and Intellect (also known as Openness). The five factor model is not only well known in Personality Psychology, but also extensively used by the HCI community [25, 14, 6].

Our proposed model aims at explaining the customer's satisfaction with basic mobile phone services by means of his/her: (1) personality traits, (2) perceived usability of the services, and (3) actual usage of these services. Figure 1 depicts the model with references to prior work related to each of the five hypothesized relationships among the different concepts. Detailed explanations on relationships 4 and 5 from Figure 1 are out of the scope of this paper. Next we therefore concentrate on presenting prior art that sheds light on the first three hypothesized relationships.

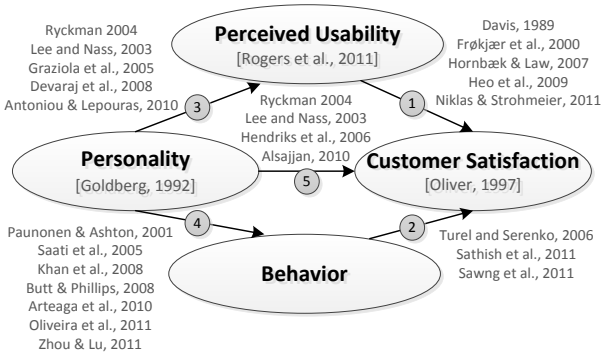


Figure 1. Proposed conceptual model. References that address each relationship are indicated onto the corresponding arrow or ellipse.

Relationship 1: Perceived Usability of mobile phone services influences the customers' satisfaction with them. Usability goals (*e.g.,* effectiveness, efficiency, learnability) have been said to be positively correlated with how people evaluate their user experience with technology (*e.g.,* satisfying, enjoyable) [29]. However, these correlations depend in a complex way on the application domain, the user's experience and the context of use [16]. Additionally, effectiveness, efficiency and satisfaction should be considered to be different goals [16, 22]. These recent findings motivate the study of our hypothesis in the case of mobile services. In this regard, Heo *et al.* [21] created a framework to evaluate the usability of mobile services, and showed that there were correlations between usability and user experience constructs, such as satisfaction. Another support for this hypothesis comes from the Technology Acceptance Model [11] that has been adapted to the specific case of mobile services [27]. In both cases significant correlations between usability goals and user satisfaction were found. In this paper we investigate the impact of perceived usability on customer satisfaction with mobile phone services.

Relationship 2: Mobile phone usage influences customer satisfaction with mobile phone services. The way customers use mobile technology influences their experience of the mobile services they use. Turel & Serenko [34] worked on a model that incorporated self-reported behavioral accounts of mobile service usage. They found that it was possible to use these measures to benchmark service operators in terms of customer satisfaction and loyalty. Similarly, Sawng *et al.* [33] worked on a model that included social benefits, satisfaction and service risks and that could be used to predict customer behavior when using mobile phone services. In market research, behavioral patterns are typically used to predict switching to a different operator (*i.e.*, churn). For instance, Sathish *et al.* [32] studied the factors that affected churn decisions in India. They found that self-reported call frequency was among the most important factors in determining whether customers were satisfied with their carriers. In this paper, we investigate the impact that actual—as recorded by the operator—mobile phone usage has on customer satisfaction with mobile services.

Relationship 3: Personality influences the perception of usability of mobile phone services. Many researchers have worked on the relation between personality and the measures that are usually taken into account to define the usability of a system. Ease of use and usefulness were studied by Devaraj *et al.* [13], who conducted a study with 180 new users of a collaborative technology and found correlations between the personality dimensions and the perceived usefulness and ease of use. Other related measures of usability have been studied for mobile services. Antoniou & Lepouras [5] worked on an adaptive mobile museum guide and showed that personality traits are related to the acceptance of the adaptivity dimensions of the service. A similar study was conducted by Graziola *et al.* [19], who found a relation between personality traits and the user's preferences of interface modality. Our work builds on these previous findings and investigates whether and how they hold in the context of the proposed model.

METHODOLOGY

According to Rogers *et al.* [29], usability testing has been increasingly performed remotely, thus allowing services to be evaluated with larger samples and improving ecological validity by keeping participants in their own environment. Furthermore, Nielsen & Levy [26]'s work on the relationship between self-reported measures and objective measures of usability have encouraged the community to also consider measuring usability in a subjective manner. We therefore opted for measuring both usability and user satisfaction using an online survey approach. Participants were recruited via email from an online panel with members living in Mexico and who satisfied two filtering criteria: they all owned a Telefonica[2] pre-paid mobile phone number, and were using basic mobile phone services for at least the past six months (*i.e.*, calls, SMS, MMS, and basic GPRS/3G related services). The online survey had two main sections. The first section included 50 questions [1] to assess their personality traits according to the Big Five model (*i.e.* extroversion, agreeableness, conscientiousness, emotional stability and intellect) [17], whereas the second section collected the participants' opinions about the basic mobile phone services that they were using.

[2]Telefonica S.A. is currently the 3rd largest telecommunication company worldwide with over 300 Million customers (21 Million in Mexico). See **www.telefonica.com** for further details.

Measures. Items were measured either subjectively or objectively. A total of seven constructs were created from survey items and hence *subjectively* measured: extroversion, agreeableness, conscientiousness, emotional stability, intellect, perceived usability, and satisfaction with mobile phone services. Each of the five personality traits were captured by 10 survey items that were later grouped into personality facets using Goldberg's [18] classification (shown in Table 1). This was performed by computing summated scales for each facet, *i.e.*, summating all positive survey items and reversed negative items related to the same facet. For instance, if one participant gave the ratings 2, 8, and 7 to the survey items $q8r$, $q33$, and $q43$ respectively (see Table 1), then the summated scale for his/her Orderliness personality facet would be: $(10 - 2) + 8 + 7 = 23$. The remaining two subjectively measured factors—customer satisfaction and perceived usability—were assessed in relation to the mobile services contracted by the participants (phone calls, messages, *i.e.* SMS and MMS, Internet access and operator's mobile Web portal). Finally, mobile phone usage was the only factor composed of items that were measured *objectively*: the total number of mobile phone calls made/received between January and June 2010, the total duration of phone calls, and the total number of messages sent/received during the same period. Table 1 summarizes data and constructs used in the study.

Participants. A total of 603 valid responses (male: 50.2%, controlled for a balanced distribution) were obtained in the final study. Participants' age ranged between 18 and 35 years old ($\bar{x} = 25.87$, $s = 5.25$)—as per our invitation filtering criteria—and they predominantly belonged to the middle socioeconomic class. The majority reported using computers (93.4%) and the Internet (92.4%) at least once a week. In terms of mobile phone use, 81.6% reported using their mobile phone everyday and 14.8% several times a week. Based on their mobile phone call data, participants made or received an average of 101 calls per month and sent or received an average of 171 messages per month.

Data analysis. The conceptual model depicted in Figure 2—note that we expanded the personality variable from Figure 1 into the Big Five traits—was evaluated using Structural Equation Modeling (SEM) [7]. We highlight at least three reasons for using this approach: (1) SEM models relationships between concepts given that its objective function maximizes the probability of predicting the covariance matrix instead of predicting values of a certain variable; (2) SEM takes measurement unreliability into account by modeling equation errors and non-measurable concepts—*e.g.*, extroversion, satisfaction—as latent variables, thus avoiding unrealistic assumptions of error-free measurements; and (3) SEM allows researchers to leverage previous knowledge given that it uses confirmatory rather than exploratory factor analysis.

The conceptual model was evaluated using Maximum Likelihood (ML) estimation and the data was bootstrapped (1000 samples) to meet the estimation's assumption of joint multivariate normality of observed variables [7]. The SEM estimation process was split in two steps as recommended by Anderson and Gerbing [4]. First we developed a measurement model, *i.e.* relationships between each factor construct—*e.g.* usability—and its corresponding items—*e.g.* efficiency and easy of use. Then we estimated the structural paths—*e.g.* between factors usability and satisfaction. The measurement model was evaluated for uni-dimensionality, reliability, convergent and discriminant validity. Finally, the hypothesized structural paths between constructs were included in the model for the final estimation.

RESULTS AND DISCUSSION

Figure 2 depicts the validated conceptual model with the most relevant statistics. Fit measures like SRMR (.05), RMSEA (.05), CFI (.94), and PRATIO (.80) reveal that our model has a good fit according to widely accepted cutoff criteria [23, 7]. Next we discuss only those results related to the influence of perceived usability on customer satisfaction, and how one can leverage the findings of the model in order to propose new design solutions for basic mobile phone services that encompasses both usability and UX goals.

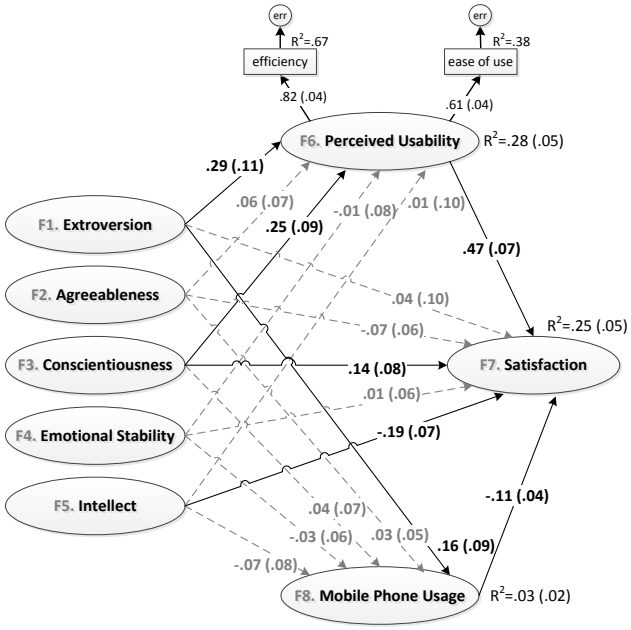


Figure 2. Validated conceptual model. Standardized loadings next to the corresponding arrows with standard errors in parenthesis (bootstrapping to 1000 samples). Significant paths ($p < .05$) indicated by solid black arrows and non-significant paths indicated by grey dashed arrows. Error variables and covariance paths omitted for clarity.

Perceived usability positively influences customer satisfaction with mobile phone services. The validated conceptual model corroborated that the usability of mobile phone services is positively correlated with the customers' satisfaction with these services ($\beta_{76} = .47$; $p = .002$). The standardized direct effect of perceived usability on satisfaction was .47, which means that when usability goes up by 1 standard deviation, satisfaction goes up by .47 standard deviation, and hence has a very strong influence on it. In fact, this is the strongest direct influence present in the model. With

Table 1. Construct factors and associated items captured subjectively by the survey and objectively by the mobile phone operator.

Construct Factor *Item code*	**Summated item** Item name	Survey code [a]	**Item description in English** / Item description in Spanish (used in the survey)	Removed from the model [g]
Extroversion [b]				
x1	**Gregariousness**	*q1*	**Am the life of the party** / Soy el alma de la fiesta	
		q6r	**Don't talk a lot** / No hablo mucho	
		q16r	**Keep in the background** / Prefiero mantenerme al margen [h]	•
		q21	**Start conversations** / Comienzo las conversaciones	
		q31	**Talk to a lot of different people at parties** / En las fiestas hablo con muchas personas diferentes	
		q36r	**Don't like to draw attention to myself** / No me gusta llamar la atención	•
		q46r	**Am quiet around strangers** / Cuando estoy entre desconocidos me mantengo callado	
x2	Poise	*q11*	**Feel comfortable around people** / Me siento cómodo con la gente	
x3	Leadership	*q26r*	**Have little to say** / No tengo mucho que decir	
x4	Provocativeness	*q41*	**Don't mind being the center of attention** / No me importa ser el centro de atención	•
Agreeableness [b]				
x5	**Understanding**	*q2r*	**Feel little concern for others** / Me preocupo poco por los demás	
		q17	**Sympathize with others' feelings** / Soy sensible hacia las emociones de otros	
		q22r	**Am not interested in other people's problems** / No me interesan los problemas de otras personas	
x6	**Warmth**	*q7*	**Am interested in people** / Me intereso por la gente	
		q32r	**Am not really interested in others** / En realidad no me intereso por los demás	
		q37	**Take time out for others** / Dedico tiempo a los demás	
		q42	**Feel others' emotions** / Siento las emociones de los otros	
		q47	**Make people feel at ease** / Hago sentir cómoda a la gente	
x7	Pleasantness	*q12r*	**Insult people / Ofendo a la gente**	•
x8	Nurturance	*q27*	**Have a soft heart** / Tengo un corazón sensible	
Conscientiousness [b]				
x9	Conscientiousness	*q28r*	**Often forget to put things back in their proper place** / A menudo olvido poner las cosas en su lugar	•
x10	**Orderliness**	*q8r*	**Leave my belongings around** / Dejo mis pertenencias en cualquier lado	
		q33	**Like order** / Me gusta el orden	
		q43	**Follow a schedule** / Hago un programa y lo sigo	
x11	Organization	*q13*	**Pay attention to details** / Pongo atención en los detalles	•
x12	**Efficiency**	*q23*	**Get chores done right away** / Realizo mis tareas inmediatamente	
		q48	**Am exacting in my work** / Soy perfeccionista en mi trabajo	
x13	**Purposefulness**	*q3*	**Am always prepared** / Siempre estoy preparado	
		q18r	**Make a mess of things** / Soy desordenado	
		q38r	**Shirk my duties** / Evado mis obligaciones	
Emotional Stability [b]				
x14	**Stability**	*q4r*	**Get stressed out easily** / Me estreso con facilidad	
		q24r	**Am easily disturbed** / Me molesto fácilmente	
		q29r	**Get upset easily** / Me disgusto con facilidad	
		q34r	**Change my mood a lot** / Cambio mucho de humor	
x15	**Tranquility**	*q9*	**Am relaxed most of the time** / Estoy relajado la mayor parte del tiempo	
		q39r	**Have frequent mood swings** / Tengo cambios frecuentes de estado de ánimo	
x16	**Happiness**	*q14r*	**Worry about things** / Me preocupo por todo	
		q19	**Seldom feel blue** / Rara vez me siento triste	
		q49r	**Often feel blue** / Me siento triste frecuentemente	
x17	Calmness	*q44r*	**Get irritated easily** / Me irrito fácilmente	•
Intellect [b]				
x18	**Intellect**	*q5*	**Have a rich vocabulary** / Tengo un vocabulario amplio	
		q20r	**Am not interested in abstract ideas** / No me interesan las ideas abstractas	
		q40	**Use difficult words** / Utilizo palabras difíciles	
x19	Creativity	*q10r*	**Have difficulty understanding abstract ideas** / Me cuesta entender ideas abstractas	•
x20	Imagination	*q15*	**Have a vivid imagination** / Tengo mucha imaginación	
x21	**Ingenuity**	*q25*	**Have excellent ideas** / Tengo excelentes ideas	
		q30r	**Do not have a good imagination** / No tengo una buena imaginación	
		q50	**Am full of ideas** / Estoy lleno de ideas	
x22	Quickness	*q35*	**Am quick to understand things** / Soy rápido para entender las cosas	•
x23	Introspection	*q45*	**Spend time reflecting on things** / Dedico tiempo a reflexionar	•
Usability [c]				
x24	Ease of Use	*q51*	**I find it easy to make mobile phone services do what I need** / Me resulta fácil conseguir que los servicios de telefonía celular hagan lo que necesito	
x25	Efficiency	*q52*	**Using mobile phone services saves my time** / Utilizar los servicios de telefonía celular me hace ahorrar tiempo	
Satisfaction				
x26	General Satisfaction [d]	*q53*	**What is your general satisfaction level with the mobile phone services that you are paying for?** ¿Cuál es tu nivel de satisfacción general con los servicios de telefonía celular que estás pagando?	
x27	Expectations Met [e]	*q54*	**How do you think the mobile phone services that you are paying for meet your expectations?** / ¿Cómo consideras que los servicios de telefonía celular que estás pagando cumplen con tus expectativas?	
x28	Ideal Mobile Services [f]	*q55*	**How close are the mobile phone services that you are paying for to your ideal mobile services?** ¿Dónde consideras que se encuentran los servicios de telefonía celular que tienes contratados con respecto a tu ideal de servicios de telefonía celular?	
Mobile Phone Usage				
X29	Calls	N/A	[not survey]: Number of mobile phone calls made/received between January and June 2010	
x30	Duration of calls	N/A	[not survey]: Total duration of mobile phone calls made/received between January and June 2010	
x31	Messages	N/A	[not survey]: Number of phone messages (SMS, MMS) sent/received between January and June 2010	

[a] Numbers in item code indicate the order of appearance in the survey while the letter "*r*" indicate the item is reversed.

[b] Associated survey items measured in a 9-point scale ranging from 1: "almost never" and 9: "almost always" as suggested by Goldberg (1992).

[c] Associated survey items measured in a 9-point scale ranging from 1: "strongly disagree" and 9: "strongly agree".

[d] Measured in a 9-point scale ranging from 1: "completely not satisfied" and 9: "completely satisfied".

[e] Measured in a 9-point scale ranging from 1: "don't meet my expectations at all" and 9: "meet all of my expectations".

[f] Measured in a 9-point scale ranging from 1: "very far" and 9: "very close".

[g] Item-analysis suggested that personality facets measured by one survey item were violating unidimensionality of their corresponding factors and should therefore be removed. Furthermore, convergent validity analysis and subjective inspection of questions pointed out that the extroversion factor should be improved by removing items *q16r* and *q36r*.

[h] When reusing the Spanish translation, change this item for: *"Intento no llamar la atención"* as suggested by Cupani (2009).

respect to the key usability goals that defined customer satisfaction, service efficiency came in first place ($R^2 = .67$), followed by ease of use ($R^2 = .38$). The model changed significantly when usability loadings for these variables were constrained to be equal ($\chi^2/df = 8.813$, $p = .003$). These results indicate that the efficiency of basic mobile phone services might be the most important usability goal determining user satisfaction—in the context considered herein.

Mobile phone usage influences customer satisfaction with mobile phone services. According to our model, this influence is rather negative ($\beta_{78} = -.11$; $p = .005$), meaning that the more one uses basic mobile phone services, the less satisfied s/he is with them. One possible explanation of this finding is that technology consumption might have a saturation point. Satisfaction could be maintained up to a point where the given technology addresses people's needs without compromising their daily routines and personal values. If by overusing mobile services one jeopardizes these routines and values, then dissatisfaction might be a natural outcome due to several reasons, *e.g.*, realizing that too much time is being wasted using them, creating anxiety to keep up with the flow of messages and calls, *etc.* Note that the construct factor for Mobile Phone Usage comprised more information about synchronous disruptive activities like phone calls ($R^2 = .94$) and their durations ($R^2 = .83$), than about sent/received asynchronous text messages ($R^2 = .45$). Therefore, the mobile phone usage patterns as captured by our model include mostly activities that can break daily routines and hence be more susceptible to the argument of technology consumption saturation point. While previous work demonstrated the existence of a link between usage behavior and satisfaction with mobile services [34, 33], our work goes one step further by finding that these are negatively linked (and quantifying the relationship), suggesting a possible explanation, and considering actual mobile phone usage as captured by the mobile operator.

Personality influences the perception of usability of mobile phone services. More specifically, extroversion ($\beta_{61} = .29$; $p = .004$) and conscientiousness ($\beta_{63} = .25$; $p = .006$) had significant effects on perceived usability of mobile phone services. The interpretation of this finding is grounded on behavior theories associated to personality traits. If today's mobile phone services are useful to shorten distances between people and allow them to efficiently interact more often, it is expected that extroverts—who interact with peers more frequently—will recognize such qualities and hence highly evaluate these services' usability. Likewise, if these services indeed help people save time, one would expect that those who care about efficiency when following daily schedules—*i.e.*, people with high scores on the conscientiousness trait—would positively rate the services' usability. We cannot directly compare our work with previous models because these studies do not group usability goals into one single factor [33, 35]. However, our work offers synergic findings by revealing that extroversion and conscientiousness have a significant effect on the usability construct (composed of efficiency and ease of use).

Limitations of the Study

As described in the methodology section, the conceptual model from Figure 2 was validated using data from 603 subjects living in Mexico with an age range of 18-35 years old, who had a pre-paid cellphone, and were using mobile services for at least six months (calls, messages and basic GPRS/3G related services). Our findings can therefore be safely generalized to this sample profile only ($CL = 95\%$; margin of error: $\pm 4\%$). Note that pre-paid mobile phone services are predominant in developing economies, but it is not in the developed world. Future work should verify whether the model also holds for smartphone users with unlimited data plan.

FROM THEORY TO PRACTICE

The conceptual model validated in the previous section contributes to our understanding of how software engineers and HCI practitioners could improve customers' satisfaction based on more clear usability goals. For example, the perceived usability of the basic mobile phone services used by our participants was the most important factor when explaining customer satisfaction. Moreover, the concept of usability was mostly characterized by efficiency ($R^2 = .67$) rather than ease of use ($R^2 = .38$), thus highlighting an important trend for satisfaction. Note that saving people's time is a recurrent result from our research as mobile phone usage had a significant negative effect on satisfaction. Next, we propose three design solutions:

First, project managers in charge of developing new mobile communication services should focus their efforts on designing more efficient solutions that minimize disruption of the users' routine. For instance we can think about leaving the possibility to request statements of the monthly bill or performing operations on the contract such as enabling (or disabling) options of the call plan via SMS or email instead of requesting the customers to go through call centers that too often require an enormous effort from their side. In terms of minimizing routine disruption, the user's contextual information could be leveraged in order to identify the most suitable periods of the day for sending them notifications or contacting them.

Second, personalized services could be created to help users with low scores on the extroversion and conscientiousness traits better manage their time when overusing mobile phones. For example, less organized people could overuse mobile services during a certain time period without planning much for the additional costs and end up with an unpleasant surprise when receiving their monthly bill. Mobile services with personality-based user models could help these "less organized" users by sending them periodic feedback on how much they have spent with phone calls and text messages, and how close they are to their preferred maximum expense. Recent work by Cherubini *et al.* [9] has revealed that the lack of personalization is actually one of the biggest barriers for the adoption of today's mobile phone contextual services. Although related mostly to basic mobile phone services, our findings are in agreement with these conclusions and further identify new opportunities for personality-based personalization. We expect their practical relevance to increase as tech-

niques for the automatic assessment of personality are more accurate and pervasive [24, 12].

Finally, mobile services should identify and provide awareness of the user's saturation point when consuming mobile phone services. Persuasive techniques (*e.g.*, social support, reminders, *etc.*) are relevant in this context towards preventing mental/physical stress and hence low satisfaction.

ACKNOWLEDGMENTS

Telefonica Research participates in the Torres Quevedo subprogram (MICINN), cofinanced by the European Social Fund, for researchers recruitment.

REFERENCES

1. Ipip 50-item big five questionnaire. Retrieved from: `ipip.ori.org/New_IPIP-50-item-scale.htm`. Accessed on September 2012.
2. Ajzen, I. The theory of planned behavior. *Org. Behavior and Human Decision Processes*, 50 (1991), 179–211.
3. Alsajjan, B. A. How the big five personality dimensions influence customers trust in uk cellular providers? *Int. J. of Global Business 3*, 1 (June 2010), 102–116.
4. Anderson, J., and Gerbing, D. Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin 103*, 3 (1988).
5. Antoniou, A., and Lepouras, G. Modeling visitors' profiles: A study to investigate adaptation aspects for museum learning technologies. *J. Comput. Cult. Herit. 3* (Oct 2010), 7:1–7:19.
6. Arteaga, S. M., Kudeki, M., Woodworth, A., and Kurniawan, S. Mobile system to motivate teenagers' physical activity. In *Proc. of IDC '10*, ACM (Barcelona, Spain, Jun 2010), 1–10.
7. Blunch, N. J. *Introduction to structural equation modelling using SPSS and AMOS*. SAGE, Thousand Oaks, Cal., US, 2008.
8. Butt, S., and Phillips, J. G. Personality and self reported mobile phone use. *Comput. Hum. Behav. 24*, 2 (2008), 346–360.
9. Cherubini, M., de Oliveira, R., Hiltunen, A., and Oliver, N. Barriers and bridges in the adoption of today's mobile phone contextual services. In *Proc. MobileHCI '11*, ACM (Stockholm, Sweden, Aug. 30 – Sep. 2 2011).
10. Cupani, M. El cuestionario de personalidad IPIP-FFM: resultados preliminares de una adaptación en una muestra de preadolescentes argentinos. *Perspectivas en Psicología 6* (Nov. 2009), 51–58.
11. Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly 13*, 3 (Sep. 1989), 319–340.
12. de Oliveira, R., Karatzoglou, A., Concejero, P., Armenta, A., and Oliver, N. Towards a psychographic user model from mobile phone usage. In *EA. Proc. CHI'11*, ACM (Vancouver, Canada, May 7-11 2011), 2191–2196.
13. Devaraj, S., Easley, R. F., and Crant, J. M. How does personality matter? Relating the Five-Factor Model to technology acceptance and use. *Inf. Systems Res. 19*, 1 (March 2008), 93–105.
14. Eckschlager, M., Bernhaupt, R., and Tscheligi, M. NEmESys: neural emotion eliciting system. In *Proc. CHI'05*, ACM (Portland, USA, 2005), 1347–1350.
15. Fishbein, M., and Ajzen, I. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA, USA, 1975.
16. Frøkjær, E., Hertzum, M., and Hornbæk, K. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *Proc. CHI'00*, ACM (The Hague, The Netherlands, April 1-6 2000), 345–352.
17. Goldberg, L. R. The development of markers for the Big-Five Factor structure. *Psych. Assessment 4* (1992), 26–42.
18. Goldberg, L. R. *Personality psychology in Europe*, vol. 7. Tilburg Univ. Press, Tilburg, The Netherlands, 1999, ch. A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models., 7–28.
19. Graziola, I., Pianesi, F., Zancanaro, M., and Goren-Bar, D. Dimensions of adaptivity in mobile systems: personality and people's attitudes. In *Proc. IUI'05*, ACM (San Diego, USA, January 10-13 2005), 223–230.
20. Hendriks, A., Smets, E., Vrielink, M., Van Es, S., and De Haes, J. Is personality a determinant of patient satisfaction with hospital care? *Int. J. for Qual. in Health Care 18*, 2 (April 2006), 152–158.
21. Heo, J., Ham, D.-H., Park, S., Song, C., and Yoon, W. C. A framework for evaluating the usability of mobile phones based on multi-level, hierarchical model of usability factors. *Interact. Comput. 21* (August 2009), 263–275.
22. Hornbæk, K., and Lai-Chong Law, E. Meta-analysis of correlations among usability measures. In *Proc. CHI'07*, ACM (San Jose, USA, 2007), 617–626.
23. Hu, L., and Bentler, P. Cutoff criteria for fit indexes in covariance structure analysis: Coventional criteria versus new alternatives. *Structural Equation Modeling 6*, 1 (1999), 1–55.
24. Khan, I., Brinkman, W.-P., Fine, N., and Hierons, R. Measuring personality from keyboard and mouse use. In *Proc. ECCE '08*, ACM (Funchal, Portugal, Sep. 16-19 2008), 1–8.
25. Lee, K., and Nass, C. Designing social presence of social actors in human computer interaction. In *Proc. CHI'03*, ACM (Ft. Lauderdale, USA, 2003), 289–296.
26. Nielsen, J., and Levy, J. Measuring usability: preference vs. performance. *Commun. ACM 37*, 4 (Apr. 1994), 66–75.
27. Niklas, S., and Strohmeier, S. Exploring the impact of usefulness and enjoyment on mobile service acceptance: A comparative study. *Proc. HICSS* (Jan. 4-7 2011), 1–10.
28. Paunonen, S., and Ashton, M. Big Five factors and facets and the prediction of behavior. *J. of Pers. and Social Psych. 81*, 3 (2001), 524–539.
29. Rogers, Y., Sharp, H., and Preece, J. *Interaction Design: Beyond Human-Computer Interaction*, 3rd ed. Wiley, West Sussex, UK, 2011.
30. Ryckman, R. M. *Theories of personality*, 8th ed. Brooks/Cole, Pacific Grove, CA, USA, 2004.
31. Saati, B., Salem, M., and Brinkman, W. Towards customized user interface skins: investigating user personality and skin colour. In *HCI*, vol. 2 (Las Vegas, USA, 2005), 89–93.
32. Sathish, M., Santhosh Kumar, K., Naveen, K., and Jeevanantham, V. A study on consumer switching behaviour in cellular service provider: A study with reference to chennai. *Far East J. of Psych. and Bus. 2*, 2 (February 2011), 71–81.
33. Sawng, Y.-W., Kim, S.-H., Lee, J., and Young, S. Mobile service usage behavior in Korea: An empirical study on consumer acceptance of innovative technologies. *Tech. and Econ. Devel. of Econ. 17*, 1 (August 2011), 151–173.
34. Turel, O., and Serenko, A. Satisfaction with mobile services in canada: An empirical investigation. *Telec. Policy 30*, 5-6 (July 2006), 314–331.
35. Zhou, T., and Lu, Y. The effects of personality traits on user acceptance of mobile commerce. *Int. J. of Human-Computer Int. 27*, 6 (2011), 545–561.

Effectiveness and efficiency as conflicting requirements in designing emergency mission reporting

Erik G. Nilsson
SINTEF
Pb124 Blindern, 0314 Oslo, Norway
egn@sintef.no

Asbjørn Følstad
SINTEF
Pb124 Blindern, 0314 Oslo, Norway
asf@sintef.no

ABSTRACT
Aspects of usability, such as effectiveness and efficiency, are critical for users' overall experience of an interactive system. In response to the on-going debate on the relationship between different aspects of usability in usability studies, we present an example of a User Interface (UI) design case where the relationship between effectiveness and efficiency should be considered as a requirement or design issue, rather than as variables in usability studies. In the presented case - status reporting from an in-vehicle support system for emergency missions - these aspects of usability were perceived as conflicting rather than as positively correlated. We present various design solutions to the task of status reporting and show how the solutions support effectiveness and efficiency in different ways. Finally, we point out some characteristics of the case that could explain our findings and we suggest how future research may obtain more insight into which types of applications that may possess similar properties.

Author Keywords
Emergency response, usability, effectiveness, efficiency.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Usability is seen as a concept that is included in the broader concept of user experience (UX) [9]. Consequently, the usability of an interactive system is critical for the users' experience, and research on usability is important to extend our knowledge in the field of UX.

The usability of an interactive system is defined as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [7]. Effectiveness is understood as goal achievement, efficiency involves the resources used in reaching the goal, while satisfaction is related to user perceptions.

A current debate is related to the degree of correlation between these aspects of usability [6, 10, 11], where the aspects are perceived mainly as variables in usability studies. In this paper we present an alternative perspective to the correlation debate supported by observations in a particular case; i.e. that in some cases it may be more fruitful to regard effectiveness and efficiency as requirements or design issues, rather than as variables of a UX or usability study that may or may not be correlated.

BACKGROUND
Previous work
The correlation debate
Sauro and Kindlund [10] presented the empirically founded single usability metric (SUM). SUM is based on a quantitative model where the three standard aspects of usability (effectiveness, efficiency and satisfaction) are summarised in one score.

A basic assumption of SUM is that there are fairly high correlations between the three standard aspects of usability. This assumption is controversial within the Human-Computer Interaction (HCI) community, as is seen in a meta-study by Hornbæk and Law [6]. They conclude that correlations between effectiveness, efficiency and satisfaction are generally low, and are lower than was found by Sauro and Kindlund.

In 2009, Sauro and Lewis [11], in response to Hornbæk and Law, reported what they described as strong correlations between the standard usability aspects (r between .44 and .60 at task level measurements) on the basis of data from 90 usability tests. Sauro and Lewis suggested that the higher correlations obtained in their study may in part be explained by their study being more representative of the kind of usability tests typically conducted by usability professionals whereas Hornbæk and Law’s study is more representative of the HCI field at large.

Conflicting requirements and forces in design
Seeing usability aspects as requirements is not new to the field of HCI. Cockton [3] discusses the need to align usability evaluation metrics with stakeholders’ goals and requirements for an interactive system. Jokela has described how to specify usability requirements in call-for-tenders [8].

It is known that requirements to a system under development may be in conflict with each other. Sommerville [12] treats this aspect of requirements engineering as a negotiation during requirements analysis. Such negotiation will typically be revisited throughout the

systems development process as requirements emerge or evolve.

One important aspect of design is to balance conflicting requirements. Design is about making choices [1]. Using prototypes iteratively helps us to make these choices when requirements are not perfect. Within design patterns, conflicting requirements or design constraints may be described as forces to be considered during design [2].

The context: emergency mission reporting

By emergency missions we mean emergency responses by professional personnel, coordinated through a central unit. The particular emergency context in this study is ambulance responses.

The particular task targeted in the present study is the status reporting conducted by the ambulance personnel throughout the mission, where the personnel are required to report when they enter one of a set of predefined statuses. The status values have a natural sequence, but in certain cases one status may be skipped or the rescue task may be cancelled/finished before all statuses are visited.

For the present study, three users or stakeholders are of particular interest: (a) The ambulance personnel as end users of the mobile device, (b) the central unit as receivers of the status reports, and (c) the legislators providing regulatory requirements on emergency health care.

Typically, an ambulance is manned with a driver and a paramedic. The end-users' environment of reporting is highly efficiency oriented. On the road, the ambulance may drive at high speed, the on-board paramedic may be occupied with a patient and at pickup and delivery every second potentially counts in order to save lives.

The requirements regarding the end users' primary task – conducting an efficient emergency mission – may be in conflict with the requirements of the central unit or from the regulatory requirements given by legislators. To the ambulance personnel on a mission, status reporting may be considered to be "noise" that should take as little time as possible. From the perspective of the central unit, in cases of complaints about the response, or in the case of audits on compliance with regulatory requirements, the status reporting should be of high quality: it should never be forgotten, and it should always be reported with correct time stamps.

OBJECTIVE AND METHOD

In the example case studied, the objective was to design the functionality for status reporting for in-vehicle users in ambulances. During our work with the visual prototype, we discovered that this reporting involved an interesting conflict between effectiveness and efficiency requirements. In this paper we want to share these as lessons learnt; as an example of an application with such a conflict. The experiences were attained through the development process, however without the support of a formal research design.

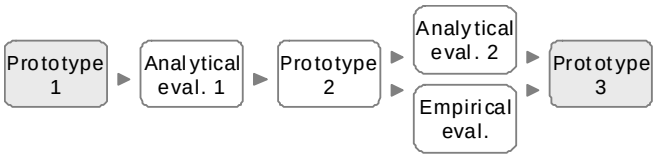


Figure 1. Process steps from Prototype 1 to 3. Non-coloured process steps are covered in the present study.

The visual prototype was developed through a user-centred process. An initial set of requirements had been established prior to the current project. The lessons learnt of this paper are the result of process steps where we conducted an expert evaluation of an initial visual prototype (Prototype 1), refined the prototype (Prototype 2), and finally conducted empirical evaluations and usability inspections on the refined visual prototype. The process steps are visualised in Figure 1.

Prototype 1 was a non-clickable visual presentation of the layout and suggested functionality. The first analytical evaluation was an informal expert evaluation with two independent usability experts (the authors of this paper).

On the basis of the informal expert evaluation, the developer presented a clickable Prototype 2. This prototype was subjected to analytical and empirical evaluations with real users.

Analytical evaluations were conducted as group-based expert walkthroughs [4]. Two sessions were conducted, with four or five ambulance personnel as evaluators in each group.

Empirical evaluations were conducted as an adapted version of cooperative usability testing [5], with alternating phases of interaction and interpretation. Eight ambulance personnel participated in individual testing sessions.

On the basis of the evaluations, the test leaders established overall redesign suggestions and a set of usability predictions.

DESIGN FOR STATUS REPORTING

The main screen of Prototype 2 is presented in Figure 2. Through this screen (presented on an 8 inch touch screen), the functionality of the support system - including status reporting – is available.

One may wonder whether there are many design issues connected to a task as simple as status reporting. The usability evaluations showed that indeed there are. To our surprise, the users were very concerned regarding the needed number of screen taps, the location of different buttons, the layout of the buttons and the labels on certain buttons – issues that are normally more present in the mind-set of usability experts rather than end users.

Figure 2: Main screen of Prototype 2. The status button is at the upper left.

In the following, we focus on three main design alternatives for status reporting: Solutions A, B and C. Solution A was used in Prototype 2; Solutions B and C were suggested during the evaluation of Prototype 2, and were thus not evaluated in the case we describe in this paper.

Solution A: Button opening menu. A button on the periphery of the screen (labelled "Available" in Figure 2) shows the current status. When tapped, a full screen menu is used to change the status. The suggested status button menu is presented in Figure 3. The buttons for passed statuses are passive, showing valuable information like the time stamp for the status change and distance travelled since the change. Clicking on one of the status buttons closes the menu, updates the status information on the current status button and returns to the screen on which the current status button was tapped.

Solution B: Toggling button (one-click status update). When the button that shows current status is tapped, the status changes to the next status in the "natural" sequence. Thus there will be no submenu, and the status may be changed with one tap on the top left-hand button on the main screen (Figure 2).

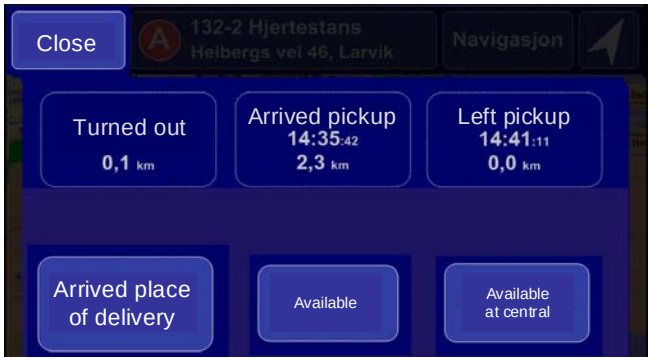


Figure 3: Suggested status reporting menu in Solution A

Solution C: Automatic reporting. This means using some criteria that make it sufficiently likely that a status change has occurred to update the status automatically, thus requiring no user interaction (as there is no interaction, this solution is not illustrated). One example of this is when an ambulance has been notified and has driven for a certain distance above a certain speed, the status should be changed to "started driving". In the same way, when the status is "started driving", the ambulance is in a certain vicinity of the emergency site and the ambulance has been standing still for a certain period of time, the status should be changed to "arrived at incident".

Effectiveness perspective

To support the effectiveness goal of making sure that the reported statuses are correct, a solution that minimises the chances for making errors is needed. Because of the small screen, and especially when operated while driving, the precision of taps on the screen may be fairly low. For the same reason, unintentionally tapping more than once on the screen may easily happen.

Solution A supports correct status reporting best. By presenting the possible status values, the user will both get a degree of consciousness with regards to statuses and reporting them, and by presenting them in a menu the user must make a conscious choice for the new value. As the choices are explicit and organised in the "natural" sequence, the risk of making an incorrect choice is reduced.

Although it may seem that Solution B supports correct reporting in "normal" cases, it increases the risk of making errors, either by unintentionally tapping more than once on the button (and thus doing two status changes) or by tapping on the status button unintentionally, for example while wanting to tap on one of the buttons next to the status button. The former error may be avoided by inducing a forced delay between subsequent status changes. The latter error is difficult to avoid and will introduce the need for functionality for correcting the status – functionality that will anyway be needed to do "unnatural" status changes.

To what degree Solution C supports correct reporting depends on the quality of the automatic reasoning, but there is always a risk that a false status change is reported. This may, for example, cause the central unit to believe that an emergency mission has been accepted by an ambulance, when, in fact, it has not. This is an argument for only using such reasoning for reminding users about status changes, not for automatic reporting, alternatively forcing the users to confirm automatic status changes.

To support the effectiveness goal of making sure that the status changes are reported at the correct time, none of the solutions are optimal. The importance of assuring that reporting is indeed performed, may point to Solution C or a reminder combined with Solution A or B, but the automatic reasoning about status changes requires that the ambulance has been driving for a while before the status change is

detected. Thus, the time that is reported for the status change, which is important from a legislative point of view, will be incorrect. This could be compensated for by setting the time for the status change to the time when the ambulance started driving, but there may also be cases where this is not correct.

In summary, the "best" solution from the effectiveness/control perspective seems to be Solution A with reminder functionality.

Efficiency perspective

The users in the ambulance focus on the main task of rescuing lives at an emergency site. They know what the status is, and the sequence of status changes is identical or very similar in all emergency missions, so reporting status changes is of little value for them. Thus, an important goal of the users inside the vehicle is to perform this task as efficiently as possible; i.e., using as few screen taps as possible and reducing the need for reading items on the screen.

Seen from this perspective, Solution C is best suited, as it requires no actions by the users. The variant requiring confirmation by the users also seems well suited, although such confirmation may come at very unsuitable moments. As the users may be performing a highly attention-requiring task, a reasonable design solution is that the users should choose the appropriate time to perform user interactions. Such confirmations violate this principle, but may still be a usable compromise.

Solution B also supports efficiency to a large degree. For "normal" emergency missions performed by experienced users knowing the sequence of the possible statuses, updating the status may be done with one tap on the screen. A possible solution to correcting errors and handling "unnatural" status changes is to have the status button as a split button like the back and forward button in most web browsers, that may be used both for doing direct operations and for opening a menu. This is a solution that works well on a desktop computer, but that requires a level of precision when tapping that is neither anticipated nor desired on a touch screen solution used in a vehicle while driving.

Figure 4: "Carousel" version of status reporting menu

Solution A is the least efficient one - in the evaluations, a number of users found the menu unnecessary. It requires a number of clicks and, at least for inexperienced users, a bit of reading to find the correct button to press.

The efficiency of this solution also depends on the layout of the menu choices. Prototype 2 presented the status choices in their natural sequence, with the buttons in fixed positions. An alternative design proposed during the informal usability expert walk through was to organise the buttons as a "carousel", always showing the next natural choice as the topmost choice, and using different sizes of buttons to illustrate how "natural" it was to choose a given status, as illustrated in Figure 4. This solution is potentially highly efficient for handling normal status changes, but is a bit "unstable", in the sense that the same menu choices appear at different positions in different contexts. Although not presented as part of the prototype used in the evaluations with the end users, other findings from these evaluations showed that the users have a strong urge to have consistent locations of screen elements.

In summary, the "best" solution from the efficiency perspective is Solution C, and if augmented with a confirmation function, it is probably equal to Solution B, depending on how the unsolved issues with regards to this solution are resolved.

DISCUSSION

Designing status reporting

The end users' needs for efficiency in the reporting task indicate that reporting should preferably be performed automatically. If forced to perform reporting, the user interface for doing this should require as few taps and as little reading as possible. During the evaluation activities, the end users communicated a desire for being able to operate the routine parts of the reporting task almost "blindfolded". Taking the effectiveness perspective, this desire is risky, as the chances of performing erroneous reporting increase when the user is not reading text on the screen.

Although such use is a special risk for Solution B, it should also be mentioned that both layout choices for the status reporting menu in Solution A invite "blindfolded" use for experienced users. Confirmation of "unnatural" choices is one way of reducing this risk. Another way of compromising between the two perspectives is to use aural feedback to confirm the choices. This may be well suited in all three solutions, but maybe most important in Solutions B and C. A drawback of using sound is the noisy environment in an ambulance.

Generalising our findings

Although it is often the case that effectiveness and efficiency correlate positively [11], our example shows that this is not always the case. As is foreseen by Sommerville, multiple stakeholders typically imply conflicting requirements. It should therefore be no surprise that

requirements concerning the usability aspects may also be in conflict. Though efficiency in reporting may be more highly prioritised by the ambulance personnel, effectiveness may be seen as more important from the perspective of the central unit.

In our view, it is important to be able to identify cases with conflicts between the two, as this may have important implications for the usability – and, by extension, the UX of the interactive system. In this section we point out some possible reasons why the conflict occurs in the given case. We assume that other cases with similar characteristics may experience the same conflict.

(1) Conflicts between stakeholders. The effectiveness needs from the central unit and the need for compliance with regulatory requirements, conflict with the end users' needs of being effective and efficient when performing the emergency mission, making efficiency in the reporting task of prime importance. In other application areas, different stakeholders often have similar interests, e.g. to make a purchase process as smooth as possible in an eCommerce system.

(2) Nature of application area. The users in the ambulance experience that their primary task of saving lives conflicts with the secondary task of reporting their status. The task conflict is accentuated in the given application area as the prime task is highly attention-demanding. Other application areas, characterized by the primary task being conducted in the application, may not observe such conflicts.

(3) Strong legislative requirements. The strong legislative requirements make correct reporting much more important than in cases where incorrect information would at worst lead to a package being delivered to a wrong address, or a small economic loss. It should also be mentioned that conforming to legislation is also in the interests of the users in the ambulance, thus raising a conflict of interest for these users independently of other stakeholders.

We perceive our findings and the possible reasons for them as a relevant input to the correlation debate. When the discussion – as it seems to be at present – is oriented towards correlation of usability aspects as a general phenomenon in usability studies, we may lose sight of the most important place for considering the relationship between effectiveness and efficiency; namely, in the requirements and design phases.

Our findings may also serve as basis for discussions about the applicability of SUM. In cases where the standard usability aspects can be seen to contain conflicting requirements, some caution may be needed when applying SUM. However, it may well be that if conflicting requirements are well-managed throughout design and development, SUM may still provide an adequate single estimate of overall system usability – though valuable details about the standard aspects of usability may also be needed.

FUTURE WORK

It is risky to make general conclusions based on observations in only one example case and we do not claim that the reasons for the conflict between efficiency and effectiveness in our study stem only from the possible reasons that have been pointed out. Neither may we conclude that all other cases with similar characteristics will display the same conflict. But we hope that the observations and discussion may serve as inspiration for discussions on the relationship between effectiveness and efficiency as aspects of usability and UX; in particular with reference to the characteristics of cases where a conflict between these two aspects is likely to occur.

ACKNOWLEDGMENTS

This work has been done within the research projects FLAMINKO and EMERGENCY, both financed by the Norwegian Research Council VERDIKT program and the participating partners. Thanks to Locus (www.locus.no) for their participation in the case.

REFERENCES

[1] Beadouin-Lafon, M., and Mackay, W.E. Prototyping tools and techniques. In Sears, A., Jacko, J., A. (Eds.) *The human computer-interaction handbook. 2. ed*. Lawrence Erlbaum Associates, New York, NY (2008), 1017-1040.

[2] Borchers, J. A Pattern Approach to Interaction Design. In *Proc DIS '00*. ACM Press (2000), 369-378.

[3] Cockton, G. Putting Value into E-valu-ation. In eds. E. Law, E. L-C., Hvannberg, E., Cockton, G. (Eds.) *Maturing Usability: Quality in Software, Interaction and Value*. Springer (2008), 287-317.

[4] Følstad, A. 2007 Group-based Expert Walkthrough". In Scapin, D., Law, E. (Eds.) *R^3UEMs: Review, Report and Refine Usability Evaluation Methods. Proceedings of the 3. COST294-MAUSE International Workshop* (2007), 58-60. http://cost294.org/upload/522.pdf

[5] Frøkjær, E., Hornbæk, K. Cooperative Usability Testing: Complementing Usability Tests with User-Supported Interpretation Sessions. In *Proc. CHI 2005*. ACM Press (2005), 1383-1386.

[6] Hornbæk, K., Law, E. L-C. Meta-Analysis of Correlations among Usability Measures. In *Proc. CHI 2007*. ACM Press (2007), 617-626.

[7] ISO. *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability*, International standard ISO 9241-11:1998(E).

[8] Jokela, T. Determining usability requirements into a call-for-tenders: a case study on the development of a healthcare system. In *Proc. NordiCHI 2008*. ACM Press (2008), 256-265.

[9] Law, E., Abrahão, S., Stage, J. Proceedings of international workshop on the interplay between user experience evaluation and software development *(I-UxSED 2010)*, editorial introduction. *The CEUR workshop proceedings series*, 656 (2010). http://ceur-ws.org/Vol-656/

[10] Sauro, J., Kindlund, E. A Method to Standardize Usability Metrics into a Single Score. In *Proc. CHI 2005*. ACM Press (2005), 401-409.

[11] Sauro, J., Lewis, J.R. Correlations among Prototypical Usability Metrics: Evidence for the Construct of Usability. In *Proc. CHI 2009*. ACM Press (2009), 1609-1618.

[12] Sommerville, I., Sawyer, P. (1997) Requirements Engineering. John Wiley & Sons, Chichester, UK.

The Craft of Programming Interaction

Rikard Lindell
Mälardalen University
Box 883
72123 Västerås Sweden
rikard.lindell@mdh.se

ABSTRACT
The creation of useful artefacts with rich experiential qualities required quality driven interaction designers and programmers with the ability to simultaneous problem setting and problem solving. Interaction design is a design practice that defines the appearance and function of digital artefacts. Bridging interaction design and engineering is problematic because design and engineering have different epistemology. Designers are trained to see a plethora of future designs for a situation and explains the phenomena of a context. Engineering focus on problem solving and depends on agreement about ends. In this paper I suggest that the poor state of designers and programmers who are not standing together can be avoided if we give up the claim that software development should be engineering or science, and instead see it as a quality-driven craftsmanship.

Author Keywords
design; interaction design; experience design; highly interactive prototypes; programming; material; craft

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI):

General Terms
Human Factors; Design

INTRODUCTION

This paper provides a normative view on how to bridge the practices of interaction design and software engineering.

The creation of useful artefacts with rich experiential qualities required quality driven interaction designers and programmers with the ability to simultaneous problem setting and problem solving. People use interactive software, websites and mobile applications in different contexts for different purposes. Boehm shows a focus shift in software engineering to usability and that requirements of interactive artefacts cannot be defined a priori [1].

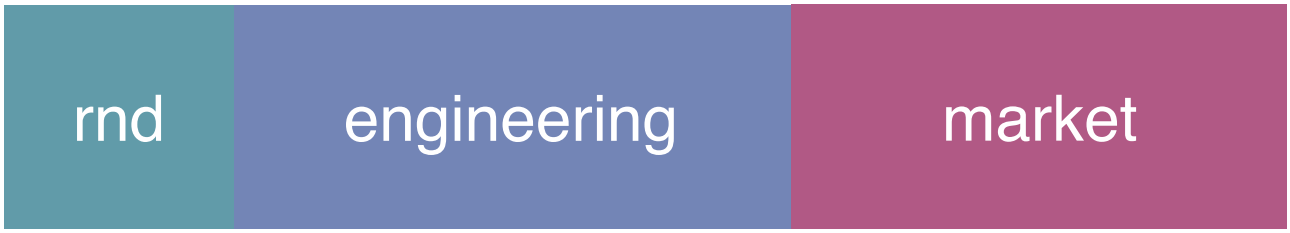
Interaction design has indulged itself in being a design practice that tries to define the appearance and function of digital artefacts [2]. Sketches, storyboards, videomatics, and interactive prototypes depict the appearance and functionality, and at best convey requirements to software engineers [3,4,5]. The result from a design process is rich in clues to the finished product. But, the material in the design process is different from the code that implements the design into a working artifact [6].

There is a big problem in how a development project runs between the phases of interaction design and engineering [7]. These two activities have different epistemology; interaction design is a design practice [2], while software engineering is struggling to describe itself as engineering and science [1]. Designers are trained to see a plethora of future designs for a situation. Design explains the phenomena of the context. It's about framing the problem space of the context, cut into a search tree of plentiful design proposition to reach the right user experience design of a future artefact [4,8]. Design is the exploratory use of malleable tactile materials and provides suggestions for possible future solutions [2,4,8]. The goal of the design process is that as much as possible frame the problem for an engineering process to take over to solve.

Sketches, storyboards, and paper prototypes works in design situations where the designer experiments with known interaction idioms. Users, design colleagues, and programmers fill the gaps and imagine the user experience for the finished artefact based on their experience with these idioms. To get talk-back from the interaction design it is necessary to create interactive prototype programs. The design process does not stop when the programming start, on the contrary, programming is a vital part of the design process.

BACKGROUND

Schön discuss how faith in rational, scientific, and technological solutions are spread because of how they were successfully applied during World War II, where the solution to a problem was to supply more resources [9]. The point he makes is that engineering is close to science. “They began to see laws of nature not as facts inherent in nature but as constructs created to explain observed phenomena, and science became for them a hypothetico-deductive system. In order to account for his observations, the scientist constructed hypotheses, abstract models of an unseen world which could be tested only indirectly through deductions susceptible to confirmation or disconfirmation by experiments. The hart of scientific inquiry consisted in the use of crucial experiments to choose among competing theories of explanation.” This quotation describes how



Figur 1. Buxton's image of the organisation for the engineering driven product development [4].

belief in deductive reasoning disconnecting the explanation of the world from the material to be explained. A scientific approach allows the engineer to deduce, analyse and define problems in a rational way; the positivist epistemology of science [9].

Boehm describes in his expose of just over a half-century of software engineering how the field evolved, mainly that we increasingly focusing on usability and value [1]. Software engineering has realised the problems with the top-down waterfall development model and introduced iterative models [10]. These models deal with changes in the problem space by development iterations. Each iteration in the spiral model or in the rational unified process (RUP) is basically a waterfall model. The foundation is still the technical rationality epistemology.

Buxton describes how an engineering-driven organisation is organised in a simple diagram [4], see Figure 1. First you do research and development, then you do engineering, and finally hand the product to the sales organisation. This type of organisation requires an agreement on ends. At changes and difficulty in clearly defining the problem dissonance arises in the organisation: “Technical Rationality depends on agreement about ends.” In the citation Schön delineates how technical rationality does not address situations where the result is uncertain and where there is no ready-defined problem to solve.

Reflection-in-action and interaction design

Technical rationality and focus on ends has a different epistemological dimension than Reflection-in-action - Schön's term for the reflective practitioner way of thinking and acting. The reflective practitioners have practical knowledge (knowledge-in-practice), they can be aware or unaware of this knowledge regardless of guild. Reflective practitioners deal with problem setting and unique and complex situations, mainly through reflection-in-action (reflection-in-action). Reflection-in-action can be summarized in three phases that are repeated: (1) Frame the problem, assess the situation, and understand the working material. (2) Perform moves over the situation. These moves are parts of the practitioner’s repertoire. They are small experiments with the intentional result, but often with unintended effects (both positive and negative). (3) Reflect and evaluate the consequences of action in conversation with the situation. Practitioners take in and reflect on how the situation responds (talk-backs). The conversation happens in what Schön calls the medium's language. After this phase the process starts over again.

Design problems are often vague, complex, and contradictory [11]. In the problem setting phase interaction designers name the phenomena that they will pay attention to and work with. They create design concepts through various design techniques: sketches, mood boards, storyboards, or paper prototypes to better understand and frame the problem. Concept design are evaluated and refined through introspection, criticism, and user studies, such as Wizard of Oz method [4]. The design work will also increase the understanding of the situation and context. Sketching interfaces and designing paper prototypes will also learn interaction designers more about the context [12]. Figure 2 shows a design driven organisation and how the design team follows the design through the entire process. Such an organisation also understand that fellows close to the market can provide feedback from users.

Interaction Designers have a repertoire of interaction styles that they can apply for different problems [3]. To be able to design great interfaces interaction designers should master programming. It is part of being conscious of the design material [5, 13]. While interaction designers can implement a design by composing software, they must not be seduced by technologies for technology's sake.

Interaction Designers create architecture for interactive artefacts and their spatial and temporal properties. They design the artefact topology, the artefact appearance on the screen or in the space and how artefact change over time because of interaction. Interaction Designers understand the consequences of different designs and have a feel for how a design can be realised. Similarly, interaction designers build interactive prototypes for technical substantiate and in full understand what they designed. Its about material consciousness [14]. The difference between the architect and interaction designer is that the latter build their model in full scale, albeit quickly, and at times chaotic, but it is a model and not a product.

Craft

The profession, the knowledge, and ability to design interactive artifacts is a creative craft. McCullough discusses the craft related to interactive technology use and how an artisan approach can enrich interaction design [15]. According to McCullough, there is a wide gap between the design of digital artifacts, and computer science and software engineering.

That craftsmanship has not been highly regarded is not new. Within software engineering is sometimes used craftsmanship derogatory to describe careless

Figure 2. Modified figure of Buxtons model of a design-driven organisation [4].

programming. Boehm for instance, uses the notion of craftsmanship as analogy for the 1960s, lack of professional discipline and careless "cowboy programming" [1]. However, negligence has nothing to do with craft. On the contrary, describes Sennett the craftsman as a quality-driven bordering on manically busy perfecting his/her work [14]. The craftsman must be patient and not tempt to do quick fixes. But, the craftsman's commitment is to do a good craftsmanship for its own sake.

"Craftsmanship may suggest a way of life that waned with the advent of industrial society–but this is misleading. Craftsmanship names an enduring, basic human impulse, the desire to do a job well for its own sake. Craftsmanship cuts a far wider swath than skilled manual labor; it serves the computer programmer, the doctor, and the artist; parenting improves whet it is practiced as a skilled craft, as does citizenship. In all these domain, craftsmanship focuses on objective standards, on the thing in itself. [...] And though craftsmanship can reward an individual with a sense of pride in work."

Sennett describes the craftsman's ability to simultaneously identify problems and solve them. This is consistent with Schön's ideas about refelektion-in-action, discussing problems qualifying in difficult situations. Sennett says that problem setting and problem solving has a rhythm that relates subconscious and conscious reflection-in-action.

"Every good craftsman conducts a dialogue between hand and head. Every good craftsman conducts a dialogue between concrete practices and thinking; this dialogue evolves into sustaining habits, and these habits establish a rhythm between problem solving and problem finding." [14].

Material

Information technology, according to Löwgren and Stolterman, is a material which it has no recognisable features [3]. This view combines interaction with "traditional" design trades and crafts.

The similarity between the industrial designer and architect on the one hand and the interaction designer however, lies in creating technology. But, the industrial designer and architect's material is concrete as opposed to interaction designer material that is intangible. We distinguish between these disciplines from one another by the material they are working in and what they create, but they have similar practices, methods, and approaches to design. IT is on the surface visual, auditory, or haptic, but this is an illusion created by calculations and represented in ones and zeros and described with programming languages. Media and language for interaction designers are sketches of the interface's appearance, creating paper prototypes and to write computer programs that embody digital artefacts' behaviour in working prototypes.

If we do not consider the development of software such as software engineering, which qualities are in the development process for the continuing development work performed as a work of reflective practitioners and quality driven craftsmen?

PROGRAMMING IS A CRAFT

"How can we make sure we wind up behind the right door [good code or bad code] when the going gets touch? The answer is: craftsmanship." [16].

In Martin et. al quote there is a notion of pursuing the mastery of craftsmanship. Gaining experience through a dialogue between tacit knowledge and explicit critique, and relying on their mastery in their practice. [14].

Empirical Findings

The Manifesto for Agile Software Development and later the Manifesto for Software Craftsmanship provide both empirical evidence supporting the idea of programing as a craft. Manifesto for Agile Software Development was written as a critique of a rigid approach to requirements specification, analysis, design and documentation and to put focus on creating useful artifacts with rich user experience. The manifesto reads: Individuals and interactions over processes and tools, Working software over comprehensive documentation, Customer collaboration over contract negotiation, and Responding to change over following a plan [17]. The manifesto reflects the programmer's frustration that spend most of the time to document and manage the project instead of writing code.

As I write this, the Manifesto for Software Craftsmanship[1] – Raising the bar has been signed by over 9,000 people (9,410) with the constant rising number of signatures. The manifesto reads:

"As aspiring Software Craftsmen we are raising the bar of professional software development by practicing it and helping others learn the craft. Through this work we have come to value: Not only working software, but also well-

[1] http://manifesto.softwarecraftsmanship.org/

Figure 3. Design and Programming as a craft facilitates the transition in the design work's change of materials and technology – from paper to pixels from sketches to code.

crafted software. Not only responding to change, but also steadily adding value. Not only individuals and interactions, but also a community of professionals. Not only customer collaboration, but also productive partnerships"

The development of software - programming - is an activity with a wide range of intrinsic properties that are closer to craft than science or engineering. Sennett describes the Linux programmer as the modern craftsman [14].

"People who participate in "open source" computer software, particularly in the Linux operating system, are craftsmen who embody some of the elements first celebrated in the hymn to Hephaestus, but not others. [...] The Linux system is public craft. The underlying software kernel in Linux code is available to anyone, it can be employed and adapted by anyone; people donate time to improve it. Linux contrast to the code used in Microsoft, its secrets until recently hoarded as the intellectual property of one company. During these two decades, the software industry has morphed within its brief life into a few dominant firms, buying up or squeezing out smaller competitor. In the process, the monopolies seemed to churn out even more mediocre work." [14].

Martin et al. press the importance of quality-driven and disciplined practice in the programming craft. The programmer must carefully name functions, classes, interfaces, methods [16]. Martin et al. focus on the code, to carefully write clean code based meticulous attention on the principles and guidelines for the scope of a function or method, of responsibility for a class, how test-driven development is pursued, how concurrency is best implemented, etc. Above all, Martin et al. show that the problem cannot be solved at once but a problem can be explored by writing tests and constant iteration of possible improved solutions.

Agile Development with XP and Scrum in particular is a big step for software engineering in the direction of focusing on service qualities and user experience as opposed to non-agile development models, such as RUP, the spiral model, and waterfall model. But, despite the Agile Manifesto have XP and Scrum and other iterative development models still a clear plan-implement-evaluate cycle oriented that extends over a longer period, as at least weeks, but in practice longer. A common feature for these methods is agreement about ends.

In recent years, the Kanban method attracted attention by providing even greater freedom for adaptation [19]. "Scrum is less prescriptive than XP, since it doesn't prescribe any specific engineering practices. Scrum is more prescriptive than Kanban though, since it prescribes things such as iterations and cross-functional teams. ... Kanban leaves almost everything open. The only constraints are Visualize Your Workflow and Limit Your WIP. Just inches from Do Whatever, but still surprisingly powerful." [19]. This quotation shows how Kanban can be a support for an agile development process in constant change. Kanban allows the goal of a work in progress (WIP) change during the process. This means that a WIP can have an open end. Thus, Kanban a radically different approach than all the earlier development models; Scrum, XP, RUP, Spiral model and Waterfall model included.

The open-endedness of Kanban stands out and allows practitioners reflection-in-action. Dealing with messy situations and continuous problem setting and problem solving important becomes pillars of programmers' work. This makes the interaction designer and the programmer standing on common ground; craftsmanship epistemology. The ongoing design process turns into a software craftsmanship process, see Figure 3.

The main difference between interaction designers and programmers is that the material of interaction design has slightly different characteristics than the material for programming. The transition between design and programming is in this situation regards the knowledge exchange between practitioners of different repertoires.

DISCUSSION

According Buxton et al the problem setting should be done without writing code. However, programming is a good tool for a design that is difficult to portray on paper; for example, collaboration, pliable, or highly interactive features. Innovative interaction techniques require interactive prototypes. But, exploratory programming allows various designs to be explored and in retrospect transform the code into clean code [15]. One way to explore a design is to propose solutions by writing tests. By first writing tests explores and sets the problem while the programmer simultaneously solves the problem [15].

Buxton notes that there is still a division between design and engineering and suggests how it can be bridged [13]. But, is still an open question for research in design and software engineering [7, 13].

We need to use methods of agile software development with a different approach. The development model Kanban

contains characteristics that allow an artisanal approach. The Kanban development model does not prescribe specific roles, and is designed to accommodate continual change. Programmers and designers can simultaneously be doing problem setting and problem solving.

One way to bridge the design process and implementation here is to let the designer be part of the development team. Instead of user stories as a description of the work in progress (WIP), the concrete material from the design process is used - mood boards, sketches, storyboards, videomatics etc. Initial WIP uses an explorative programming approach to continue the design process and explore the problem space. As the artefact takes shape, the development process can adopt a more pragmatic approach.

Kanban is a relatively new model in software engineering but has since become popular in game development. It is no coincidence, since game development is focused on highly interactive experience and game play. But, to use Kanban artisanal manner, the participants in the project need to have the craftsmanship epistemology.

A practice oriented epistemology and ontology bridge the designing and constructing activities within interaction design and programming. An artisanal approach facilitates the design and development of innovative and highly interactive digital artifacts that have novelty and relevance.

REFERENCES

1. Barry Boehm. *A view of 20th and 21st century software engineering*. In Proceedings of the 28th international conference on Software engineering (ICSE '06). ACM, New York, NY, USA, 12-29. (2006)
2. Daniel Fällman. *The Interaction Design Research Triangle of Design Practice, Design Studies, and Design Exploration*, Design Issues, 24.3, p. 4-18, MIT press (2008).
3. Jonas Löwgren, Erik Stolterman. *Desgin av informationsteknik - materialet utan egenskaper*, Studentlitteratur, ISBN 91-44-04203-5, (2004)
4. Bill Buxton, *Sketching User Experiences - getting the design right and the right design*, Morgan Kaufmann, ISBN 978-0-12-374037-3, (2007)
5. Rikard Lindell. *"Jag älskar att allt ligger överst" – En designstudie av ytinteraktion för kollaborativa multimedia-framträdanden*. Mälardalen University Press Dissertations: 72, ISBN 978-91-86135-24-9. (2009)
6. Anna Vallgårda, Tomas Sokoler. *A Material Strategy: Exploring Material Properties of Computers.* International Journal of Design. Vol 4, No 3 Dec 21. (2010)
7. Thomas Memmel, Fredrik Gundelsweiler, and Harald Reiterer. *Agile human-centered software engineering*. In Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 1 (BCS-HCI '07), Vol. 1. British Computer Society, Swinton, UK, UK, 167-175. (2007)
8. Klaus Krippendorf. *The Semantic Turn*. CRC Press, Taylor & Francis Group. ISBN10: 9-415-32220-0, (2006)
9. Donald A. Schön, *The Reflective Practitioner - how professionals think in action*, Basic Books, ISBN 0-465-06878-2, (1983)
10. Boehm, B.W. 1985. "*A Spiral Model of Software Development and Enhancement,"from Proceedings of an International Workshop on Software Process and Software Environments,* Coto de Caza, Trabuco Canyon, California, March 27-29, (1985)
11. Erik Stolterman. *The Nature of Design Practice and Implications for Interaction Design Research*. i International Journal of Design, 2(1), p 55-65. (2008)
12. Barbara Tversky, *What do sketches say about thinking?,* in Proceedings of AAAI Spring Symposium on Sketch understanding, sid 148-151, (2002)
13. Bill Buxton.-*On Engineering and Design*: An Open Letter. Businessweek. April 29, 2009. http://www.businessweek.com/innovate/content/apr2009/id20090429_083139.htm
14. Richard Sennett. *The Crafsman*. Penguin Books, ISBN-13: 978 0 141 02209 3, 2008
15. Malcom McCullough, *Abstracting Craft - the practiced digital hand*, MIT Press, ISBN 0-262-13326-1, 1998
16. Robert C. Martin, Michael C. Feathers, Timothy R. Ottinger. *Clean Code: A Handbook of Agile Software Craftsmanship*. ISBN 978-01-3235-088-4
17. Kent Beck, Mike Beedle, Arie van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, James Grenning, Jim Highsmith, Andrew Hunt, Ron Jeffries, Jon Kern, Brian Marick, Robert C. Martin, Steve Mellor, Ken Schwaber, Jeff Sutherland, Dave Thomas. *Agile Manifesto*. http://agilemanifesto.org/. (2001)
18. Mikael Lindvall, Vic Basili, Barry Boehm, Patricia Costa, Kathleen Dangle, Forrest Shull, Roseanne Tesoriero, Laurie Williams and Marvin Zelkowitz. *Empirical Findings in Agile Methods*. Ed: Don Wells, Laurie Williams. *Extreme Programming and Agile Methods — XP/Agile Universe* 2002 Lecture Notes in Computer Science, Volume 2418/2002, 81-92 (2002)
19. Henrik Kniberg, Mattias Skarin. Kanban and Scrum - making the most of both. C4Media inc, ISBN: 978-0-557-13832-6. (2010)

A Cross-Disciplinary UX Evaluation of a CRM System

Marcin Sikorski
Gdansk University of Technology,
Faculty of Management and Economics
ul. Narutowicza 11/12
80-233 Gdansk, Poland
Marcin.Sikorski@zie.pg.gda.pl

ABSTRACT
This paper presents a case study of what was intended to be a qualitative usability evaluation of a CRM (Customer Relationship Management) system but finally ended as a cross-disciplinary service design innovation workshop. This text presents evaluation framework and main categories of obtained results, discussed from the viewpoint of redesigning the CRM system as an e-service for internal customers. Discussion of key success factors and lessons learned from this study conclude the paper.

Author Keywords
usability, User Experience, User-Centred Design, Service Design, collaborative design, Intranet

ACM Classification Keywords
H.1.2. Human factors; H.5.2. User interfaces; H.5.3. Group and Organization Interfaces;

INTRODUCTION
Usability of business IT systems has been a topic of numerous studies since the beginnings of HCI [5, 6, 7, 11]. Usability of company Intranets and other back-stage IT systems still has a big impact on work efficiency. These systems are today an essential part of each digital workplace [1], serving as corporate information repositories and facilitating internal communication, teamwork and workflows.

Research perspectives concerning interactive systems in recent years evolved a lot: systems engineering perspective so dominant tree decades ago has been replaced User-Centred Design (UCD) perspective now. In recent years also User Experience (UX), Value-Based Design and Service Design perspectives brought research methodologies closer to a real social and economic context in which contemporary interactive systems have been actually used. Social interactions on-line and – in general – human behaviour on-line have become new, intriguing research issues, regarding both private and business life.

PROBLEM DESCRIPTION

Problem background
A multi-modular CRM (Customer Relationship Management) system has been used by a large Polish financial company, but in the focus of this evaluation there was included only the CRM module used by call-centre operators for serving daily hundreds of customers by the phone.

This usability evaluation project was undertaken mainly due to systematic complaints arriving from the call-centre operators, who were claiming that poor system usability dramatically slows down the customer service. Moreover, recently there have been incoming signals that customers are getting increasingly irritated by time-taking call-centre procedures even in small matters. As a result, after reaching some critical mass, these operator complaints were seriously taken and finally the CRM usability improvement project has been launched.

Evaluation framework
The company so far has not had their own usability staff, so an evaluation team has been formed of:

- two external usability consultants,
- four employees: the CRM system "owner" from the IT department and three senior call-centre operators (department leaders), very experienced in dealing with different types of financial products.

In order to streamline the teamwork, following evaluation procedure was accepted:

1. Crowdsourcing method will be used at first for gathering by e-mail all observed complaints from front-line operators in the call-centre.
2. Collaborative expert review of typical operator procedures will be performed for major operational paths.

3. Complaints collected from front-line operators will be aggregated with evaluator's comments as to their relevance and feasibility for planned usability improvements.
4. Supplementary expert evaluation (checklist and heuristic) will be applied for assessing the user interface compliance with HCI guidelines.
5. Final report (PowerPoint presentation to be discussed with the IT department and the executives) will be prepared, showing prioritized recommendations and their projected impact on system usability.

Evaluation context

The team worked over a week several hours a day, thoroughly analyzing a live demo of on-the-phone customer service and watching literally each step performed by senior operators. The system was operated from a laptop in a training room, with live CRM picture projected onto a big screen so as all team members could have a good visibility of the spots where the usability problems were identified. The demo was accompanied with narrative "user stories" by senior operators explaining the purpose and meaning of each action performed in a call-centre conversation context.

During the demo presentation front-line operators' remarks and suggestions from crowdsourcing have been reviewed and supplemented by senior operators' comments on the possible impact a specific flaw might have on the customer service speed and quality.

It seemed noteworthy that senior operators often referred to the fact that the conversation flow with the customer on-the-phone was strictly regulated by the company procedures. However, because of different reasons on the side of the customer the default conversation flow often must be adapted on-the-fly to the context - and the CRM system should be flexible enough to let the operator work that way.

During the teamwork we could observe *gradually changing focus of attention* from usability of the CRM system to analyzing user experience of an operator. In the background, however, we have been also considering the user experience of the customer on-the-phone; it is indirectly affected by perceived service quality, resulting from the combination of the CRM system usability and the momentary UX of a call-centre operator.

EVALUATION RESULTS

Usability and UX aspects

Despite many usability flaws have been detected, in general in this CRM system using tab-based web interface with plenty of editable forms, operators basically met no problem in finding a suitable navigation path matching the actual needs of the customer on-the-phone.

However, it turned out that the most important operator UX discomforts with the CRM system were caused by some other factors, like:

- necessity to frequently quit the CRM system in order to find information available only in other modules (e.g. off-line contact history data), or
- necessity to verify currently displayed data in other sources.

The issues of sub-optimal visual design, demanding manual control or inconsistent data fields labelling have been also raised, and later confirmed in the expert evaluation review.

While the team approached identifying dimensions of user experience, it also turned out that operators were very creative in finding various workarounds to overcome existing usability problems because their actual performance was very much affected by the bonus system, which was fed by the data from automatic monitoring of operator's actions in the CRM system. These observations helped to understand actual operators' work habits, motivations and attitudes, bringing important ethnographic insight to the scope of this evaluation study.

Organizational aspects

During evaluation sessions the team discussions very often evolved from pure usability towards user experience (UX) issues, interpreted in twofold manner:

(1) *Operator experience*, covering a set of emotions resulting from the CRM system behaviour and simultaneously, from the customer behaviour on the phone line;

(2) *Customer experience*, covering the set of emotions resulting from the perceived quality of specific on-the-phone service.

When discussing the screens and procedures, the team members realized that the CRM system usability problems must be seen as a part of overall service quality landscape, also relevant to the way how operators actually do their best with the existing CRM system (trying to earn their bonus, though).

As a result, a set of guidelines was proposed for the final evaluation report, covering issues such as:

- visual design and interaction flow improvements,
- software improvements (technical quality),
- better formatting of usability specifications for external software vendors.

More importantly, a set of classified recommendations was made, aimed at improving operators' trust to the CRM system and operators' relationship with the company brand, as the employer.

Other outcomes

Apart from usability- and UX-relevant outcomes, other key findings of this study were important:

- negative operator's UX resulting from suboptimal usability of the CRM system is likely to affect the quality of service offered to the on-the-phone customer; therefore improving usability of the backstage CRM is a good investment for enhancing the quality of serving the customer by the call-centre;
- in this project company managers experimentally decided to gather usability comments from CRM operators by open internal crowdsourcing, and also by encouraging other staff members to contribute to the project; it produced surprisingly fruitful outcomes and resulted in creating a unique cross-departmental cooperation around this project;
- front-line operators turned out to be highly motivated to deliver their comments in crowdsourcing and to participate in further redesign process of the CRM system, which is the main tool in their work environment; this attitude may suggest the premise of positive relationship with the employer, reflected here in their commitment.

Finally, during subsequent evaluation sessions a cross-disciplinary perspective was developed in the project team, which seemed to contribute much to the project success. Otherwise it wouldn't be possible to embrace the complexity of discovered problems: evaluation viewpoints that were very diverse at the start, have been gradually negotiated and aggregated during evaluation teamwork, at the end usually resulting in a set of balanced and feasible recommendations.

POST-EVALUATION REMARKS

Key success factors

At this point, after completing the evaluation part of this project, some key success factors could be identified:

A. Staff commitment

The first success factor - already mentioned - was very *productive crowdsourcing*, which delivered dozens of valuable comments and suggestions from the front-line.

Consequently, *senior operators and the CRM owner (IT)* - used their expertise to frame collected suggestions into a specific task context and were very active in searching for feasible solutions.

In both cases it was visible the staff was aware how the usability flaws affect the service quality for external customer, despite natural motivation to improve operator's experience and comfort as well.

B. Flexible teamwork

In this project creating an *ambient evaluation environment* was also very important for facilitating effective teamwork: a round table configuration, circular information flow, ongoing visual contact, a wall-size projected CRM screen as a central focus of attention - all these elements all helped to stimulate group dynamics in this project.

The next important success factor was *agile-like evaluation cycle* which formed the canvas for the analytic part of the project. This cycle was repeated regularly for each discovered usability problem and consisted of following sequence:

1. executing step-by-step specific task situation in the CRM system, accompanied by "user stories",
2. reviewing situation-relevant comments and suggestions from crowdsourcing,
3. locating and classifying user interface problems,
4. brainstorming for possible solutions[1],
5. searching for the problem cause and origin,
6. problem diagnosis and reference to the procedures or local organizational context,
7. documenting proposed solution (or a set of).

This cycle was iterated for each detected problem and it allowed conducting unstructured analysis. Iterative conversational method, asking "naive" questions and refining answers through the unrestricted creation of ideas have finally led to developing solution proposals.

In this cycle "the art of asking right questions" to the senior operators also played some role; it was essential for focusing attention on important UX aspects and for creative exploration of problem space.

Finally, the *integrating role of senior operators* was crucial during evaluation sessions: they enabled putting the operators' complaints *into the screen context* and *into the task/organizational context*, both essential for external usability experts for proper interpreting high-level interaction design principles to a specific screen or conversation scene.

Novel evaluation elements

Despite of direct outcomes aimed for the CRM system redesign, in this project some novel elements emerged:

A. Usability evaluation converted into innovation workshop

When developing proposals for improving the operator UX, both individual creativity and team-discussed refinements

[1] brainstorming for possible solutions was intentionally located in this cycle *before* finding the problem cause

were combined, using spontaneous brainstorming and also analytic conceptual refinements.

Starting from visions of specific screens with improved interaction elements, the amount of creativity input was growing so fast, that it gradually converted usability evaluation sessions into a sort of innovation workshop. The list of proposed improvements and innovations was long, and they could be sorted into two groups:

- ideas relevant to UX, user interface and the CRM system, aimed at improving operator UX with the CRM system;
- ideas relevant to various organizational improvements related to the back-stage activities.

B. Forced multipoint analysis

Due to sensitivity of this project, invited external usability experts were able to operate the CRM systems only via an authorised senior operator.

Paradoxically, the apparent shortage of direct experience from "feel" of the system resulted in more extensive discussions, because domain experts (senior operators) had to explain in more detail the meaning/purpose sense of each click and each operation.

In seems that forced restrictions in access to the system apparently facilitated developing a multi-point, cross-disciplinary evaluation perspective for team members.

C. CRM system as an internal e-service

A cross-disciplinary evaluation perspective has finally led to putting the CRM system in the wider context of the call-centre services offered to customers. From the external customer viewpoint everything is a service, and from the operator viewpoint everything what is provided to facilitate his/her work can be also considered a service (on-line or off-line, respectively).

As such, the CRM system actually is an internal e-service aimed at operators who are internal customers. Analogically, the other part of the system (voice interface with an operator) is the front-stage e-service aimed at external customers.

Treating a CRM system holistically as kind of e-service (twofold: internal and external), helped to identify complementary values produced for internal and for external customers. In general, this perspective seems useful also for prospective evaluations of other IT systems in this company.

KNOWLEDGE MANAGEMENT ASPECTS

The teamwork performed in this project can be divided into three parts:

1. analytic - typical evaluation, based on general HCI and usability evaluation methodologies [6, 7, 11],
2. creative - brainstorming and evaluating solutions, based on Double Diamond model [2],
3. constructive - documenting redesign recommendations, to be implemented later in another project.

In both analytic and creative parts knowledge-intensive tasks have been performed, involving cross-disciplinary knowledge diffusion among team members. Knowledge transfers typical for usability consulting have been described in [10], and they again appeared in this CRM

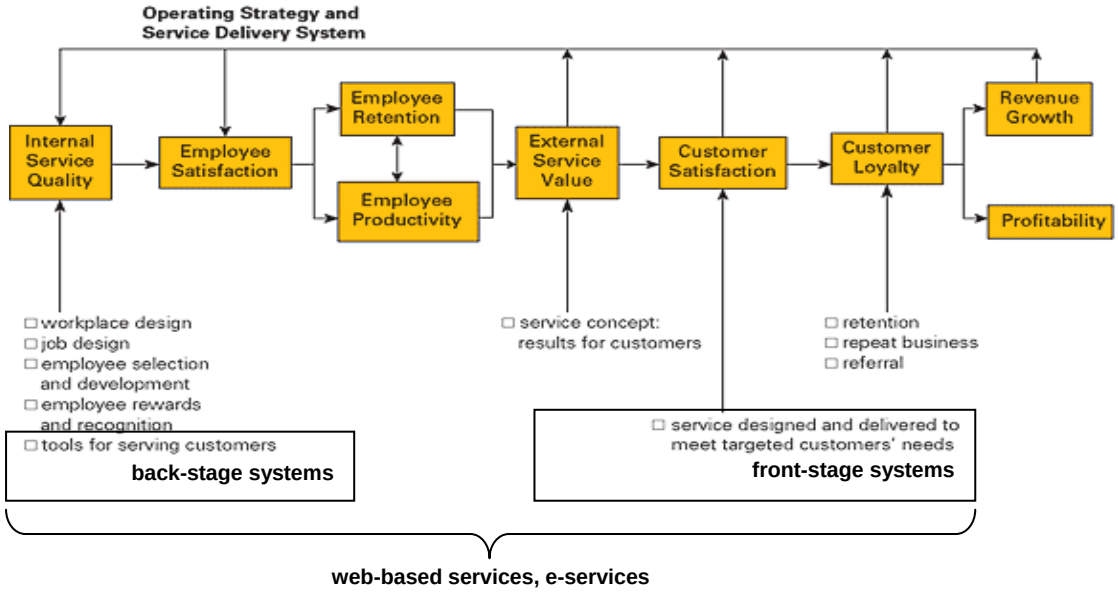


Fig. 1. Value chain in service systems, adapted from [4]

system case. In this project cross-disciplinary knowledge transfer resulted in:

- novel understanding of the CRM system as e-service (with a direct impact on UX of internal customers, indirect on UX of external ones);
- converting usability evaluation framework into a sort of innovation workshop, aimed on developing creative solutions for improving customer service
- converting HCI design focus into service design perspective, adopted for further developments in CRM redesign project.

Finally, during the final report presentation there was the knowledge transfer between the evaluation team and the project sponsors - company executives.

SERVICE DESIGN PERSPECTIVE

Starting from a routine usability study, this project has eventually raised the significance of broader UX evaluation focus, namely treating the *interactive system as a service system*, which produces value for internal and for external customers.

This perspective is coherent with the concept of service value chain proposed by Heskett [4], which argues that internal service quality (incl. tools for serving customers) affects employee satisfaction and job commitment. Consequently, in this case of CRM system the operator UX has an indirect impact on customer UX and on future relationships with the work environment as a part of the internal branding.

Fig.1. (in the lower part) shows the parts of the service value chain included in this evaluation, but also organizational issues, which should be included as internal service quality factors.

Adopting service value chain perspective may result in remarkable redefining the role of HCI in current IT projects:

- while IT these days is often merely a vehicle for launching specific on-line services (internal or external), HCI and interaction design are often expected to build UX-competitive advantage and deliver value to users (customers);
- possibly better UX results may be achieved if an interactive system is designed as a service system (IT-based), aimed to offer value for specific group of customers.

Service design perspective involves the issue if *value co-production*:

- in on-line *service systems* value for customer is co-produced in part by quality of human-computer interaction, but in the other part by quality of human-socioeconomic relationships relevant to actual system usage, like convenience, cost-saving, community etc.
- in on-line service *design process* value is also co-produced by participating clients/users (Value Co-Creation), what extends the current scope of User-Centred Design and UX design closer to increasingly popular the Service Design approach [12].

Developing profitable on-line relationships, involves mutual sharing of values produced by specific business model.

In case of on-line service systems this perspective places current HCI design practices much closer to economics, especially if the user is a conscious consumer (external, internal) willing to consume, abut also willing to co-produce value in a specific business context relationship.

CONCLUSIONS

This evaluation study produced several novel outcomes, unexpected at the beginning of this project: effective use of crowdsourcing, use of narrative "user stories" ethnographically presenting operators' work habits, as well as using elements of Co-Design and Value Co-Creation, characteristic for the Service Design approach.

This project also led to a deeper understanding that:

- in e-business systems projects HCI has many touchpoints with service design,
- many interactive systems can be designed as IT-based service systems, producing value for both internal and external customers,
- in usability evaluation and UX design users/customers should be involved as value co-producers, what extends their role in the current UCD approach.

Consequently, service value chain concept may be applied for many corporate IT systems, which should be treated as e-services designed jointly with User-Centred and Service Design approaches.

ACKNOWLEDGEMENTS

This work was partly supported by the Polish National Science Centre under the contract No. 4591/B/H03/2011/40. The author would like to thank the three anonymous reviewers for their valuable comments on first version of this paper.

REFERENCES

1. Bernard R. (1998). The Corporate Intranet. Wiley and Sons, New York.
2. British Design Council. URL: http://www.designcouncil.org.uk/designprocess

3.Dourish P. (2006). Implications for Design. CHI 2006 Proceedings, 22-27 April, 2006. ACM. 541-550.

4.Heskett J.L., Jones T.O., Loveman G.W. (1994). Putting the Service-Profit Chain to Work. Harvard Business Review, March-April 1994. 164-174.

5.ISO 9241: Ergonomic requirements for office work with visual display terminals. Parts 10-17

6.Landauer, T. K. (1988). Research methods in human-computer interaction. In M.Helander (Ed.), Handbook of Human-Computer Interaction (pp. 905-928). Elsevier, New York.

7.Hix D., Hartson R. (1993). Developing User Interfaces: Ensuring Usability Through Product and Process. John Wiley & Sons, New York.

8.Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. Organization Science 5 (1). 14-37.

9.Probst G., Raub S., Romhardt K. (2000). Managing Knowledge: Building Blocks for Success. Wiley.

10.Sikorski M., Garnik I., Ludwiszewski B., Wyrwiński J. (2011). Knowledge Management Challenges in Collaborative Design of a Virtual Call Centre. In: Koenig A., Dengel A., Hinkelman K., Howlett R., Lakhani C.J. (eds): Knowledge-Based and Intelligent Information and Engineering Systems. KES 2011, Part II, Springer LNAI 6882, 2011. 657–666.

11.Shneiderman B., Plaisant C. (1994). Designing the User Interface: Strategies for Effective Human-Computer Interaction. Addison-Wesley, Reading, MA.

12.Stickdorn M., Schneider J. (2010). This is Service Design Thinking. BIS Publishers, Amsterdam.

Identifying User eXperiencing factors along the development process: a case study

Marco Winckler
ICS-IRIT
Université Paul Sabatier
winckler@irit.fr

Cédric Bach
ICS-IRIT
Université Paul Sabatier
cedric.bach@irit.fr

Regina Bernhaupt
ICS-IRIT, Univ. Paul Sabatier
RUWIDO
Regina.Bernhaupt@ruwido.com

ABSTRACT
Currently there are many evaluation methods that can be used to assess the user interface at different phases of the development process. However, the comparison of results obtained from methods employed in early phases (e.g. requirement engineering) and late phases (e.g. user testing) of the development process it is not straightforward. This paper reports how we have treated this problem during the development process of a mobile application called Ubiloop aimed at supporting incident reporting in cities. For that purpose we have employed semi-directive requirement interviews, model-based task analysis, survey of existing systems and user testing with high fidelity prototypes. This paper describes how we have articulated the results obtained from these different methods. Our aim is to discuss how the triangulation of methods might provide insights about the identification of UX factors.

Author Keywords
Incident reporting systems, UX factors, development process

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

1 INTRODUCTION
Incident reporting is a very well-known technique in application domains such as air traffic management and health, where specialized users are trained to provide detailed information about problems. More recently, this kind of technique has been used for crisis management such as the hurricane Katrina [1]. Such self-applications are aimed to be accessible by the general public with a minimum or no training. In the context of the project Ubiloop, we are investigating the use of mobile technology for allowing citizens to report urban incidents in their neighborhood that might affect the quality of their environment. We consider urban incidents as any (micro)events, perceived by a citizen, that might affect the quality of his urban environment (e.g. hornet nest, potholes, broken bench, tags,…). By reporting incidents, citizens can improve the quality of life by influencing the quality of their environment. Figure 1 illustrates the overall scenario of our case study.

Figure 1. Overview of incident reporting with Ubiloop: users report incidents like potholes, tagging, or broken street lamps to the local government using a mobile phone application.

Despite the fact that incident reporting systems using mobile technology are becoming more common, little is known about its actual use by the general population and which factors affect the user experience when using such system. In order to investigate which user experience factors must be taken into account when designing the interface of mobile application for incident reporting, we have employed several evaluations methods (including semi-directive requirement interviews, model-based task analysis, survey of existing systems and user testing with high fidelity prototypes) along the development process of the application Ubiloop (developed in the context of the eponym project). Hereafter we report how, using several evaluation methods, it was possible to:

- Identify which (and in what extension) UX factors affect mobile incident reporting systems;

- Associate UX factors and artifacts that are aimed to support the design and implementation of systems;
- Determine how users value the incident reporting systems (in terms of UX factors) in both early and late phases of the development process.

The first two sections of this papers provide an overview of the development process (section 2) and the methods employed (section 3) in the Ubiloop project. Then, at section 4 we describe how we have articulated the results in order to provide a bigger picture of UX factors and artifacts used during the development process. Finally we discuss the results and lessons learned.

2 OVERVIEW OF THE PROCESS

We have followed a user centered design approach. Our first goal was to identify how user experience factors are important to the users when they are performing tasks such as reporting, monitoring and sharing with other citizen's information about urban incidents. We firstly address the following dimensions: perceived quality of service, awareness of perceived user involvement with reported incidents, perceived effects of mobile technology for reporting incidents, trust, privacy, perceived usefulness, usability and satisfaction with incident reporting systems in urban contexts. These dimensions are articulated around four main research questions:

- How citizens perceive and describe urban incidents as part of their quality of life?
- How does the choice of communication to digitally report incidents in a mobile context influence the overall user experience? If so, what dimensions of user experience are important for such an incident reporting application?
- How does social awareness affect the user experience when interacting with incident reporting systems?
- What contextual factors are important for incident reporting and which interaction techniques better assist user in reporting incidents?

These questions were investigated along the development process by the means of different evaluation methods as shown by Table 1.

Table 1. Methods employed during the development process of the application Ubiloop.

Design phase	Methods employed
Requirement analysis	Survey of existing applications Semi-directive requirement interviews Model-based tasks analysis
Design	Prototyping
Evaluation	User testing

Figure 2 shows the articulation between methods and artifacts produced. Notice that the dashed arrows indicate the relationships ensuring cross-consistency between artifacts and results obtained from the methods employed.

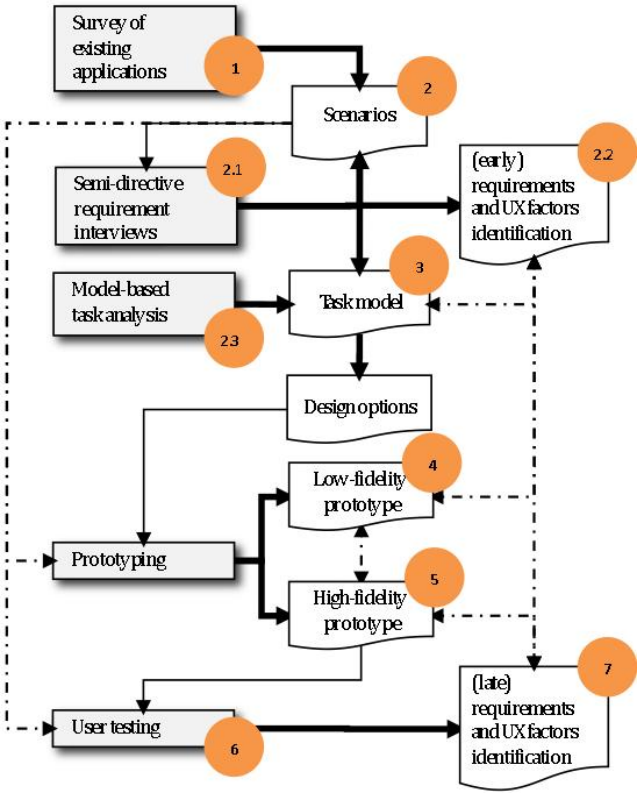


Figure 2 Articulation between artifacts and methods employed. Thick lines indicated artifacts produced; thinner lines indicate input for the method; dashed lines are used to show compatibility checking between artifacts.

In more general terms, this design process started by (1) benchmarking existing applications in order to provide a coverage of the application domain. From this step we have extracted (2) generic and representative scenarios that were used to organize an (2.1) interview with (18 potential) future end-users of the Ubiloop application. These requirement interviews allowed us to identify new scenarios (some of them not covered by existing applications), expectations (that we name here early requirements) and (2.2) UX factors that are associated to the scenarios. By (2.3) analyzing a set of 120 scenarios it was possible to identify a task pattern that was then specified by using a task-model notation. This (3) task model was used to check the coherence of the design with respect to the previously identified scenarios. Then, design options supported by the task model were (4) (5) prototyped and subsequently tested with end users. During (6) user testing, we have assessed (7) UX factors that were then compared with those collected earlier during the (2.2) interviews.

3 METHODS EMPLOYED AND MAIN FINDINGS

In this section presents the methods and key findings.

3.1 Survey of Existing Systems

In order to analyze the actual support provided by existing applications, we conducted an analysis of existing services for incident reporting in urban contexts. This study focused on the front office (i.e. reporter tools). Applications for incident reporting were first identified from the set of tools ranked by Web search engines (i.e. google.com). Then, only those that were available for remote testing were selected for further analysis.

Fifteen applications were selected covering international reporting services. What we found to be specific for the area of incident reporting is the broad diversity of features for reporting urban incidents (more than 340). Nonetheless, these incident reports seem to share similar characteristics which can be used for helping users to locate on the user interface the service that better suits to the type of incident s/he wants to report in a given context of use. Despite the fact that these applications address the same problem of reporting incidents in urban context using mobile technology, none of them was implemented following the same scenario; which might be explained by cultural difference that affect the user experience with this kind of applications. For example, in some countries the identity of the citizen reporting the incident is always mandatory whilst in other countries it was mainly optional or only requested in specific types of incidents (that could be perceived as denunciation).

From the analysis of existing systems we have extracted a set of generic and representative scenarios that should be supported by our application. We could not find in the literature any work describing UX factors addressing this specific application domain.

3.2 Semi-directive requirement interviews

In order to understand users expectations and requirements for the future system, two series of semi-directed interviews were conducted. The first one, called *general interview*, focused on how users perceive their environment and how they formulate general requirements for reporting incidents using a smartphone. The second one, called *scenario-based interview* was designed to investigate how users react to different situations that would be subject of an incident report. Each series of interviews involved nine participants.

During the *general interview*, participants were prompted to report about: how they perceive places and their environment; negative experiences in terms of environmental quality; personal involvement with problems; preferred system design; and dimensions they think important.

In the *scenario-based interview,* participants were introduced to 7 scenarios (one at once, in random order) and then asked to explain how they would envisage reporting incidents using a smartphone. The scenarios included to report a *broken street lamp*, a *pothole*, a *missing road sign*, a *bulky waste*, a *hornet nest*, a *tag/graffiti*, and a *broken bench in a park*. These incidents were selected from the set of scenarios supported by existing applications. Moreover, each scenario was designed to highlight a specific point, for example: a *broken lamp* points out an incident that is difficult to illustrate with a picture, whilst a *hornet nest* focus on the perceived danger. Every interview included a short questionnaire on demographics and technology usage. All sessions were recorded and then transcribed by a French native speaker. The transcriptions were analyzed accordingly to the grounded theory approach [3][6]. A corpus of 92 240 words was analyzed resulting in 11 classes/codes with 1125 segments of text. The coding was supported by the MaxQDA 10 software [8].

The interviews provided two key pieces of information: i) scenarios for reporting incidents, which can be associated to a task that must be supported by the system; and, ii) qualitative attributes that could be interpreted as UX factors associated to the given scenario. For an example, let assume the following segment given by participant P2: *"…Besides going to report your* [own] *idea, you could ask if there are other ideas* [proposed by other]... [that are] *close to your home..."* From this passage, the participant clearly states a UX factor (*stimulation* as described by Hassenzahl [4]) that could influence him to perform the task (*report* [an incident]). These two requirements interviews provided evidence for identifying the following UX dimensions: *visual & aesthetic experience*, *emotions* (related to negative experience of the incident and positive experience to report it – joy / pride), *stimulation*, *identification (through their personality, their own smartphone, their sensibility to specific incidents)*, *meaning and value*, and *social relatedness/co-experience*.

3.3 Model-Based Task Analysis

From the analysis of existing applications and interviews we have identified 120 possible scenarios that could be generalized as a user task pattern consisting of: *(1) to detect the incident, (2) to submit an incident report* and *(3) to follow up on an incident report*. This pattern was modeled using the task notation HAMSTERS [6] which feature a hierarchical graph decomposing complex tasks into more simple ones as shown by Figure 4. Tasks are depicted accordingly to the actors involved in the task execution (i.e. user, system or both). It also integrates operators for describing dependencies between tasks (i.e. order of execution). As this task model does not impose any particular design for the system it can accommodate all the scenarios identified during the analysis of existing applications. By modeling user tasks it was possible to identify aspects such as optional/mandatory tasks associated to incident reporting, inner dependencies between tasks, as well as pre- and post-conditions associated to tasks execution.

3.4 Prototyping

In previous work [2] we have found that information related to incidents includes: **what** the incident is about, **when** it

occurs, **where** it is located, **who** identifies the incident and the expected **outcomes** leading to its solution. These dimensions include optional and mandatory elements that characterize incidents. For example, the dimension **what** can include a combination of either a textual description, a picture of the incident, or just an indication of the incident category. Based on these early findings and the generic task model described above we developed a low-fidelity and then a high-fidelity prototype (see Figure 3). The prototype takes full benefits of currently embedded technology available in smartphones such as video camera and global positioning systems (GPS). GPS makes the user's task of locating incidents easier and photos attached to the description of incidents provide contextual information and in some situation might be used as evidence of its occurrence.

a) b) c)

Figure 3 Ubiloop protoype featuring: a) main menu page; b) textual description of incident; c) location on an interactive map.

The user interface of the Ubiloop prototype supports all the user tasks previously identified. The prototype was also designed to support the early requirements expressed by users. Moreover, the prototype was designed to create a positive user experience that could be also inferred from the results of the semi-directive requirement interview. For example, to enhance the UX factor experience we deploy the prototype in *a smartphone* (whose technology is perceived as a stimulation for using the application), we include *categories of incident* (as users said they are more likely to report an incident if they could see example of categories on incidents) and allow users *to see reported incidents in the neighborhood* (as suggested by the participant 2, see section 3.2).

3.5 User testing

A user testing with high-fidelity prototype was designed to explore how users report urban incidents with Ubiloop. The study was held at the campus of the University of Toulouse during the summer 2012. Thinking aloud protocol was used during the experiment. Users were asked to wear glasses embedding a video recording system, so that it was possible to determine where they were looking at whilst using the prototype. The recording apparatus also included a logging system and a screen recorder embedded into the smartphone.

Users were trained during 5 minutes on how to report a simple incident (i.e. a *Broken street lamp*) with a smartphone embedding Ubiloop. Participants were then asked to follow a predefined route in the campus and any report incidents found in the way. The route was populated with tags prompting users to report fake incidents that refer to the scenarios presented in section 3.2. In addition to these predefines tags, users were free to report any other incident he could see in the campus (and the route had many real incidents such as *potholes*, *tags*, *public light open during day*…). In addition to these tasks users were asked to fill in a demographic questionnaire, an AttrakDiff questionnaire [5] and a debriefing interview.

Nineteen participants, ranging from 21 to 52 years old, took part in the experiment. All participants successfully complete the tasks. The analysis of data concerning UX factors took into account the answers provided by the AttrakDiff questionnaire, the users tasks and the comments provided by users whilst performing the tasks. Again user's comments were transcribed and analyzed accordingly to the grounded theory approach. At this time the segments were coded accordingly to the actual tasks performed by the users during the experiment.

One of the findings is that all UX factors previously identified during the semi-directed requirement interviews (see section 3.2), were reported again during the user testing. Nonetheless, due to space reasons, we illustrate the description of findings to two factors, *stimulation* and *identification to incident*, that we have found out to be key UX factors to engage the process of reporting (when user decides to report the incident s/he identified in the environment).:

- *Stimulation* was evaluated during the user testing through a question of the post-test interview: "Did you discover some incidents on the University campus that you could report with the prototype?" This UX factor can was also detected during thinking aloud technique and the Attrakdiff questionnaire.
- *Identification to incident* was evaluated with another question of the post-test interview: "Are the incidents you declared during the experiment candidates to be really reported by you to the Ubiloop service"?

Furthermore, the evaluation of Identification to incident reveals that a strong proportion of UT participants declare to be ready to report some of the mandatory incidents (90 % for Broken bench and Hornet nest; 75% for the Broken street lamp; and 45% for the Heap of rubble). And individuals are mainly ready to declare the incidents they spontaneously discover during the experiment (according that the declaration is easy to perform and useful). In other words, the applications seem to be able to increase both *Stimulation* and *Identification to incident*.

4 TRIANGULATION OF METHODS

To answer the research questions on what user experience (UX) dimension should be taken into account when designing incident reporting systems for urban contexts; we

have triangulated the results of the three methods used in this work, as follows:

- During semi-directive requirement interviews users expressed requirements and expectation for reporting incidents by the means of personal stories that were interpreted as possible scenarios of use. These scenarios were then used to revise our original task model for incident reporting systems.
- By using a model-based task analysis, it was possible to remove ambiguities present in the discourse of participants and then to formalize users' requirements. Moreover, model-based task analysis provided an accurate description of user tasks. This step is extremely important for future development of incident reporting systems. As described in [7], tasks models do not only improve the understanding of user tasks but they also can be used to assess if an incident reporting system was effectively implemented to support the specified set of user tasks.
- In order to make sure that tasks identified in the semi-directive requirement interviews and model-based tasks analysis are representative we compare them with a survey of existing systems. The results confirm that our analysis is exhaustive because our task model covers all tasks supported by surveyed systems and these systems do not implement any task that is not described in the task model.
- The analysis of transcripts of semi-directive requirements interviews also supported the identification of UX dimensions associated to user scenarios. By combining UX dimensions and user scenarios it was possible to extrapolate the results in a single task model as shown by Figure 4 where user tasks are decorated with UX dimension (e.g. [ID] for identification] so that the above could be read as follows: *"I am passing by at this park every Sunday and this bench has not been repaired for weeks [ID]. It is time now to report that, so it will get fixed. It is not really a problem or unsafe, but the bench is simply not usable in the current state [MV]. [: detect/recognize the incident:]. It seems important now to make sure that the appropriate person is informed about that bench [CX], I think I should use the application to report the incident, because I want to be a good citizen [ID]. I think it is a good idea to send them a photo so they can see that the bench is really broken and that the wood has to be replaced. And when they see the photo they see that it is really there and so they will not need my contact information to have a proof that the broken bench really exists. [MV] [:describe the incident:]"*. This example shows how user tasks are interrelated to the UX dimensions.
- The prototypes were building accordingly to the task models. Once implemented, the prototype was cross-checked in order to make sure that it can effectively support the scenarios early identified. Thus, every presentation unit (ex. screens and widgets) can be easily associated with an element of the task model. By extrapolation with the results from requirement interviews we could extrapolate a tuple consisting of *user interface elements + user tasks + UX factors*.

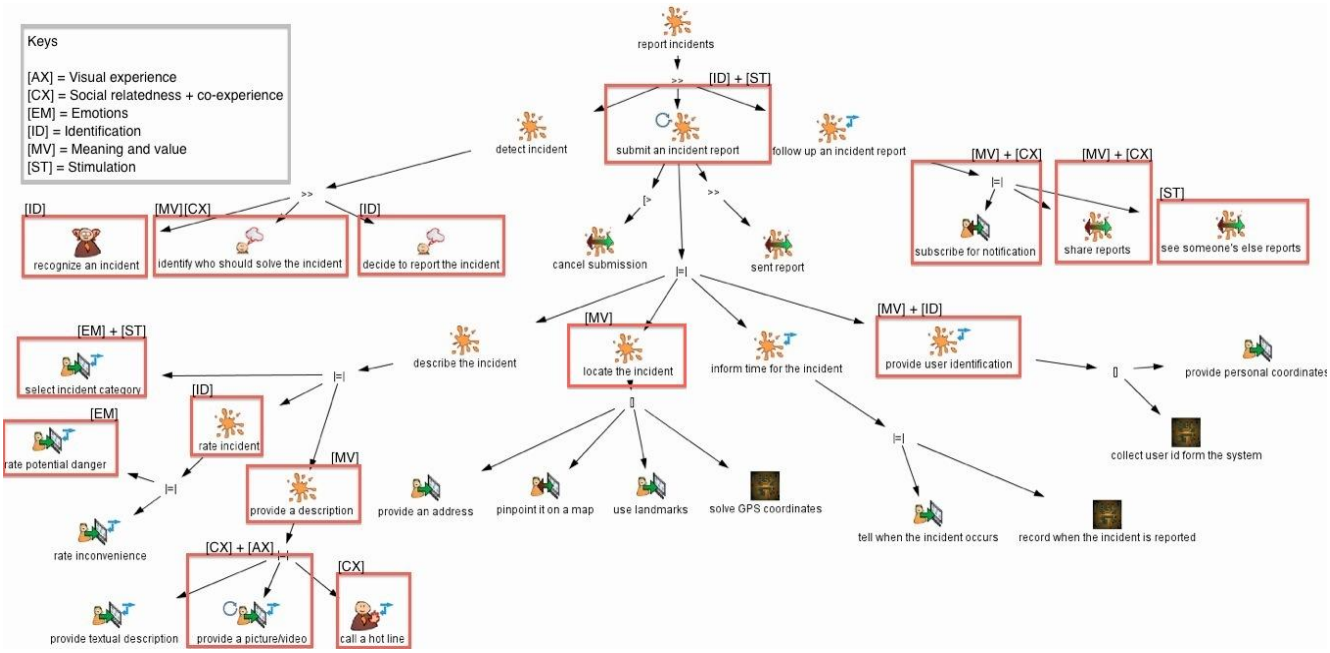


Figure 4 Generic task and most important UX dimensions for each sub-task.

- During user testing is was possible to identify UX factors during the execution of the tasks with the prototype. It is interesting to notice that the scenarios supported by the prototype were the same used during interviews so it was possible to correlate the results found in early and later phases of the development process. Thus, we have found that the UX factor *stimulation* reported during interviews to the tasks *to*

find incidents occurred again when the users use the prototype to complete the same task. This confirms the value of early identification of UX factors with requirement interviews. Moreover, when counting the number of segments of user testing reporting the UX factor *stimulation*, we have found that this factor is more frequent and even distributed along tasks. We also have compared the categories of incident reported by users during the thinking aloud and during the debriefing; we have found that the distributions of incidents across categories are more important in the requirement phase (72 citations/42 categories) than in user testing (80 cites/19 categories). Indeed, during the requirement interview participants had difficulties to identify/remember urban incidents whilst during user testing participants had more ease to identify incidents along the route of the experiment.

- Before the participants of the requirements interviews had strong difficulties to identify, remember or imagine urban incidents. It's not the case (or less the case) when users can interact with the mobile application.
- Others examples come from the responses to the post-test interview question about the Stimulation factor.

  "I never thought to report this kind of incident [a public garbage with a broken top] before [to use the application], but that true this is would be quickly a serious problem of squalor."

  "That's funny because this application gives me the opportunity to discover my own environment with a new eye."

5 DISCUSSION AND LESSONS LEARNED

Unfortunately, we don't have room for providing a comprehensive description of the results collected by the different methods. Nonetheless, the results given in this paper illustrate that UX factors can be detected both in early and later phases of the development process. Moreover, in some extend, such results can be correlated.

One of the challenges was to determine the importance of UX factors when they are collected in different phases of the development process. In the present work we have been using a simple counting method (number of segments) and distribution of UX factors across users' tasks. Using this simple method we found some differences that require further analyses. Nonetheless, it prompts by a case where it would be interesting to have quantitative metrics of UX for comparing them.

It is important to associate the identifying UX factors with the artifacts used to the design. In our study, we have found that scenarios and task models works as a *lingua franca* for mapping user requirements and UX factors. However, it is worthy to notice that this might be specific to a certain types of interactive systems that can be successfully described by tasks models. We can just wonder if this approach could work in application domain such as game where user activity is harder to represent by the means of task models. Further work is required to determine if other design artifacts and evaluation methods can also be used to provide such as articulation.

We have deliberated performing the user testing with high-fidelity prototypes. We have found in the requirement interviews that the use of the device smartphone is *per se* a stimulating element. For the purpose of the project, it was more important to test the high-fidelity prototype in a situation of mobility than a paper-based mockup. However it would be interesting to assess the impact of mockups on the identification of UX factors.

Acknowledgement

This work is part of the Ubiloop project partly funded by the European Union. Europe is moving in France Midi-Pyrenees with the European Regional Development Fund (ERDF). Genigraph and e-Citiz are partner of this work.

REFERENCES

1. Moynihan, D. P. (2007). From Forest Fires to Hurricane Katrina: Case Studies of Incident Command Systems. IBM Center for the Business of Government.
2. Bach, C., Bernhaupt, R., Winckler, M. Mobile Incident Reporting in Urban Contexts: Towards the Identification of Emerging User Interface Patterns. In 5th IFIP's WG 13.2 Workshop PUX. Lisbon, Portugal, September 5th 2011.
3. Glaser, B.G., Strauss, A. L. (1967) The discovery of Grounded Theory: Strategies for qualitative research, Adline: Chicago.
4. Hassenzahl, M. The thing and I: understanding the relationship between user and product. In Funology: From Usability to Enjoyment, M. Blythe, C. Overbeeke, A. F. Monk, & P. C. Wright Eds Dordrecht: Kluwer, 2003, pp. 31-42.
5. Hassenzahl, M. (2002). The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction, 13, 479-497
6. Lazar, J., Feng, J. H., Hochheiser, H. Research methods in Human-Computer interaction, John Wiley & Sons: UK, 2010.
7. Martinie, C., Palanque, P., Winckler, M. Structuring and Composition Mechanisms to Address Scalability Issues in Task Models. In Proc. of INTERACT, (3) 2011, pp. 589-609. Springer LNCS.
8. MaxQDA [online] software available: www.maxqda.com
9. Matyas, S., Kiefer, P., Schlieder, C., Kleyer, S. (2011). Wisdom about the Crowd: Assuring Geospatial Data Quality Collected in Location-Based Games. Entertainment Computing – ICEC 2011 Lecture Notes in Computer Science, 2011, Vol. 6972/2011, 331-336.
10. Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211–221. doi:10.1007/s10708-007-9111-y.

The Big Picture of UX is Missing in Scrum Projects

Marta Kristín Lárusdóttir
Reykjavik University
Menntavegur 1, 101 Reykjavik
marta@ru.is

Åsa Cajander
Uppsala University
Box 337, 751 05 Uppsala
asa.cajander@it.uu.se

Jan Gulliksen
Royal Institute of Technology
Lindstedsv. 3, 10044 Stockholm
jangul@kth.se

ABSTRACT
The Scrum development process has gained increasing popularity during the last decade. At the same time user experience (UX) has emerged as an important quality feature. However, the integration of UX related activities into Scrum projects has not been without problems, and this area needs to be further examined. This paper describes the results from two in depth interviews with knowledgeable UX specialists working in Scrum projects in the product development industry. It describes their ways of working generally with UX, their experiences from UX evaluations and the challenges encountered from their UX work. The main concern when working with UX in Scrum projects is that the big picture of UX is often lacking. Finally, the paper discusses the differences and similarities between the experiences from the UX specialists.

Author Keywords
User experience, user experience evaluation, usability, agile software development, Scrum.

ACM Classification Keywords
H.5.2. User Interfaces-User-centered design. General Terms: UCSD, Design, Human Factors.

INTRODUCTION
The international standard ISO 9241-210 defines user experience as: *"a person's perceptions and responses that result from the use or anticipated use of a product, system or service"* [8]. The standard extends the concept of usability from the ISO 9241-11 standard in several ways [7]. User experience (UX) deals with much more than the effectiveness and efficiency that is the main focus of usability measurements. UX addresses satisfaction in its widest possible application, from the hedonic feelings about a product before it has even been unpacked to the feelings raised that goes far beyond the very task-oriented nature of the usability focus.

In software development the need to focus on UX keeps increasing as products and services become more competitive and need to function in a much broader context than previously. Recently development processes of a more agile nature have emerged that put emphasis on team work and production rather than on structure and documentation. One of the most popular agile software development processes is Scrum [14]. Many people associate Scrum with UX, but there is nothing in the process saying that user experiences are taken into considerations automatically. Hence the need to study how Scrum development projects address and manage user experience and the development of effective and agile ways of addressing usability and UX is much sought after.

User Experience Measures and Evaluation
Researchers agree that UX is a complex concept, including aspects like fun, pleasure, beauty and personal growth. UX focuses on the more emotional aspects of user interactions, shifting the focus on how the users feel while and after using the software, the sensation, and the meaning as well as the value of such interactions in everyday life. Evaluating the UX has been a challenge for IT professionals.

UX is described by Hassenzahl as having pragmatic and hedonic attributes [6]. The pragmatic attributes are task related. The main two pragmatic attributes are ease-of-use, described by effectiveness and efficiency and usefulness, described by words like clear, supporting and controllable.

The difference of usability and UX measures has been discussed extensively, especially the difference of user satisfaction and UX [11]. One of the main discussion topics is that user satisfaction is more quantitative and UX is more qualitative. Moreover, it has been pointed out that usability measures may hint to a particular problem and sometimes to a solution of it, whereas UX measures are more general. According to Law, this makes the usability measures more useful and persuasive for the IT professionals [11].

Many methods have been suggested to evaluate the different aspects of UX. The usage of 96 UX evaluation methods in various software development activities was studied in a recent study [20]. The methods were analyzed according to at what stage in the development process the method could be used. Most of the methods could be used in the implementation and testing stages, and around one-third could be used either in requirements analysis or design stages.

UX in Scrum Projects

The agile process Scrum has gained popularity in the software industry in the Nordic countries in recent years. One third of IT professionals in Iceland used this process in 2009 [9]. In Scrum, self-organizing and well compounded teams are emphasized, typically with six to eight interdisciplinary team members [14]. In Scrum, the projects are split up in two to four week long iterations called sprints. At the end of each sprint, a potential shippable product is delivered to the customer, meaning that it should be functioning for the users.

The Scrum development process has been criticized for not involving real users in the software development and for not adequately addressing their usability needs [15]. One of the main conclusions in an extensive literature survey on the integration of the end user needs into agile processes is that these have not yet been sufficiently included in the agile development processes [17]. Because of the short sprints and the emphasis on completing a particular part of the software during each sprint, the IT professionals do not have much time in their development for involving users and for conducting UX evaluation [10].

Some researchers have suggested that some human-centred activities are conducted before the actual implementation in the project starts in order to address usability from a more holistic perspective. This is also the method used in the organization described by Sy [18] where a strategic phase before the project begins, contains specified human-centred activities to understand the context of use for example. Additionally, other researchers have recommended that activities related to UI design should be performed before the actual implementation starts [4, 21].

METHOD

This workshop paper presents the experiences that two much knowledgeable UX specialists had from integrating UX evaluation into Scrum projects. These two UX specialists were interviewed in a large interview study focusing on the integration of a wider concept, namely User Centred Design in Scrum practice made in 2010. For the purpose of this workshop paper these interviews haveconseque been re-examined for data regarding UX evaluation and Scrum and these two interviews were found.

The two interviews were semi-structured and an interview template was used. The interviews were carried out on site and lasted for about an hour. Two researchers conducted the interviews. One researcher was taking notes and the other asked the questions. The interviews have been transcribed verbatim. The quotations provided in the text are however not always verbatim, but sometimes slightly rephrased in order to be more readable and representative.

In the data analysis three predefined categories were used, as described for example by Silverman [16]. The categories are: 1) The UX specialists way of working, 2) their remarks on UX evaluation and 3) the challenges they have encountered when working with UX and Scrum. The interviews were read through and coded by two researchers according to the predefined categories. The writing of this workshop paper was also a part of the analysis. Data was discussed and interpreted as a part of the writing like in Wolcott [22].

The male UX specialist is a 46 years old man who has 13 years of experience from working in different consultant companies. In his present employment he works for one of the largest IT companies in the Nordic countries with about 10 000 employees and 14 000 customers. His job title is usability designer, and he holds a PhD in Human Computer Interaction with the focus of adding a usability and user experience perspective in software development. He has worked with the integration of Scrum and UX in several different projects in industry. These projects range from "public interfaces to internal systems" but his main focus during the last years has been on public applications.

The female UX specialist is a 35 years old woman with a Master-degree in media technology science with the specialization towards human computer interaction and sound. She has worked in industry for four years. Her present employer is a large Swedish product development company founded in 1994. The company has offices in eight countries and clients from all over the world. During her four years as a UX specialist she has made much progress in her company, and she has managed to establish UX as a core activity. Her formal role is a user experience manager, and she is in the middle of a process to hire ten members for a UX team where she will be the manager. The products she is working on are adaptable custom products related to social media and the web.

RESULTS

In this section the experiences made by the UX specialists are presented. Each person's experience is categorized according to the three predefined categories: Ways of working, experience from UX evaluations in Scrum and challenges encountered. The experiences that each person has had are presented separately with the help of quotations from the interviews.

The Male UX Specialist

Way of Working – Importance of UX vision

The male UX specialist describes that the strategic vision of the product is very important for him and that he uses that as a starting point in his work with prototypes in the development: *"What I usually do when I work with products like this is to look at the vision. The strategic vision for the product is stated and then I describe that in terms of prototypes and develop it from that. So usually I work both on a strategic level and in the actual production (in the development of the product). "*

The UX specialist explains that his way of working is an adaptation of his UX work made for Scrum. The strategic

vision and the UX goals are necessary to define before the actual project starts, ie before the sprints, but also to have in mind during the whole project when defining what to do in the different sprints. He stresses that he and the team work with the UX vision before the project as well as during the project and that the vision and the development work needs to run in parallel. *"We work in the strategic level usually both before the sprints, before the project starts, but also during the project you need to both develop a vision to get the big picture, basically about the whole user experience and then from here we can decide that okay here's a chunk of work that needs to go into production. Then it goes into the Scrum project. It's not like first we do a lot of work beforehand and then suddenly the project starts and we do nothing more. Because I think it needs to be developed in parallel."*

When asked about if he is a member of the team or outside a team he answers: *"A bit of both. I was a member of the teams, but at the same time I become more, almost like, since I worked more on the requirements part of the development I was a bit of both, you need to be both on the requirement side and also part of the actual production to make it work."*

The UX specialist explains this double role of the development of the vision before the project, and the use of the vision in the development: *"You need to get the big picture, but you also need to be involved in the actual production to be sure of that what's actually produced is what was decided on in the first place."*

UX Evaluation – Common Understanding is Vital

When asked about how the UX is evaluated the UX specialist explains that a common understanding of the UX experience is crucial in Scrum projects since he is often not directly involved in the work during the sprints. Hence, he often has meetings before the sprints with developers and testers to set the requirements together, and to decide what to include in the different sprints. The goal of these meetings is to have a common understanding in the team regarding the product and the UX. Note that the UX specialist uses the word testing while he is actually is talking about evaluation of the user interface. *"When you have the vision clear, you can make sure that this user requirement is going to be implemented in this sprint and before the sprint starts you make sure that all the detailed requirements are set. And then we walk through it. The detailed requirements are something that we do together, some members of the team, and make sure that they are in place before the actual sprint starts. And we usually have meetings together with some developers and testers. The testers can make sure that they have the test cases in place based on this. Developers can make sure that they know what to do before the sprint starts. And then during the sprint you can be there for ad hoc discussions when a case needs to be straightened out. But most of the work is done before the sprint starts and some of the work is done during the sprint. Then of course when they (the team) have something to show, you can actually test it by walking through it yourselves."*

The UX specialist explains the importance of doing UX evaluation before the actual development starts in order to have a good vision of the UX experience during the project: *"I think it's more important to do user testing before production (development) starts and then every now and then on the actual products to make sure."*

The UX specialist stresses the importance of doing UX evaluations on big chunks of functionality when working in Scrum projects. He maintains that doing tests on small pieces of functionality is unimportant, as it is the big picture that adds to the UX. The timing of the UX evaluations depends on the progress in the project, and in his experience user tests should be done as soon as there is enough to evaluate. He describes the timing of user evaluation by saying: *"That could be anytime when you have a decent chunk of functionality to test. Again this is because I think it's more important for us as usability people to test the big picture, to get the full, it's not like okay I know we are able to log in, but it's not so interesting; the interesting part is when you have the big picture in place. And then you can test maybe a number of things at the same time."*

Challenges Encountered –UX vision Difficult to Maintain

The UX Specialist describes that despite his work and experience with the UX vision it is especially difficult to get an overview of the UX in Scrum projects: *"The drawback in Scrum is that it's so feature oriented and the problem is that you don't have a big picture of the whole user experience."*

The UX specialist describes the challenge further, and maintains that the Scrum process that focuses on delivering small pieces of functionality suits most programmers perfect. The programmers have a responsibility to deliver a small piece of the software, but they often do not feel responsible for the UX or the whole system: *"I guess for a programmer it's perfect to get a small piece of work that you can work on and deliver. But the problem is that there is now no one that actually is responsible for putting this piece of functionality into the big picture. So there is no one responsible for the actual full user experience. That's the problem."*

The UX specialist describes that it can be difficult to maintain the UX vision in Scrum projects: "*After a while you have added so many features that you don't know where to put them anymore. And if you don't have the vision clear in your head or on paper it's starting to get quite difficult to know what to do with this piece of functionality and then you do something, just to squeeze it in. And that's the reason for that I think it's so important to do a thorough pre-study before prototyping and testing. Because if you have that it's so much easier to prioritize*

and say okay say that from this vision we have decided to do this piece now and that piece then. At least we know where it all fits in. And then of course this vision will change all the time, because the market changes or whatever. But still you can work on the vision then and know where to put the pieces."

The Female UX Specialist

Way of Working – UX is a Part of the Whole Project

The second UX specialist explains her way of working and it is noticeable that she works together with developers as well as managers, in other words she is both working in and outside the development team. She does different kinds of UX activities throughout the project:*"Right now I'm a user experience manager. Basically it's the role of an interaction designer, but I don't do the visual design at all. We have designers doing that. My main part is to come up with low-fi prototypes and wireframes and stories and those kind of things. I'm looking at the information architecture, the interaction architecture and then I handle it over to people who design them and implement them. In the end I also do testing as well and I take care of the focus group."*

The specialist also explains how she prepares her work. She works in a very strategic way and influences the people that have informal power in the project. She sees to it that her solutions are presented to the developers by someone they listen to, and she does not do the presentation herself:*"When I come up with a solution I don't do it myself. I always consult their leaders. You pick the developer that they are listening to. You kind of work around with them (the development leaders), then they are the ones telling the developers that this is technically possible and it suits our platform and it's definitely best way to go."*

The UX specialist explains further that in her experience the UX activities need to be a part of the whole project, from idea to testing and the UX people need to work in parallel to the developers:*"Right now the company is really focusing on user experience and have that as a mission to enhance it and provide, well as they say in the business goal, the business strategy to have an exceptional user experience so we're going from a team of one (me) to ten people I think… we're going to have a team, user experience team that I'm going to be managing with I think three interaction designers, two web designers and four developers as well. Because we want the whole chain. We don't want interaction focusing on one thing and then handing it over to development, and then implementing and then testing. We want it to be within the same team, all the expertise."*

The UX specialist explains the motivation for this change is that the company is selling UX rather than features:*"The company has noticed that it makes money. I think that's the main force. I think they were selling features, now they are selling experience rather than features."*

UX Evaluation – The Value of Social Skills

The UX specialist describes the UX evaluation of one particular feature, and explains that most of the evaluation was done in pre-studies. Note that the UX specialist uses the word testing while she actually is talking about evaluation of the user interface: *"We just did a huge task and we worked for three weeks doing the prototypes and testing the prototypes and those kind of things, before the development actually started. I started small defining it and then we tested out and then we came up with a concept and we presented it for developers and the product managers. This is what we think it should look like, feel like, these features are what we need in this system to be able to support it and all the motivation around the concept. Then we changed it a little bit then we got that input."*

The UX specialist describes that often she gives feedback to the developers. Here it is noticeable that this UX specialist manages to give severe critique to the developers without them becoming really irritated. Sometimes they react when she gives comments, but she manages to solve the situation by joking and smiling. In the following the UX specialist describes how she managed to keep good attitude in the team:*"My bosses are saying you're too diplomatic. You should be more strict. You should point with your hand and say this is wrong. But I don't believe in that. I've gained respect from not doing that, so that's what I'm telling them. If I had come in and starting doing that in the beginning, I mean I don't think it would work but right now you know they (the developers) don't feel threatened. It's been more of collaboration and I've told them that this is how I am as a person as well. I could have pointed and said do it like this. It's a give and take."*

Furthermore, the UX specialist describes a very informal approach to UX evaluation where paper prototypes are used. She usually tests her paper prototypes on developers working in the company, and invites users over lunch to make them evaluate the prototype:*"We tested these prototypes in-house and with two contacts outside the company that I can test quickly with because they know who I am. Then it's a little bit more simple to say like: 'Hey let's take a lunch and you can come here and test the product, rather than making such a big deal out of it. Because it's just a paper-prototype so it's quite hard to get people motivated to come here and assign an hour and leave their job." "I think that's the biggest struggle that we have getting people motivated."*

Challenges Encountered –Timing of UX Evaluation is Hard

The UX specialist explains that it is hard to find a good timing for the UX evaluation in Scrum. She explains that evaluation too early in the project is difficult since the different features are too small to be relevant to evaluate the full user experience with users. If sufficient amount of features have been developed to evaluate then it is difficult to make large changes on the product because some parts of it have already been delivered to the customers and there is

little time to evaluate the remaining part before the delivery. She explains: *"When one back-log item (user requirement) is done you can't really test that separately. Because it's just a component within the whole feature you know. So that's really a little bit hard because then you have to wait for all the components to be ready and that can take, it took two months I think to get it ready and then I could test it again. But then you can just tweak it a little bit, you can't do big changes there."*

DISCUSSION

The Big Picture of UX is Missing in Scrum Projects

Both our UX specialists mention that Scrum is feature oriented. One of them stresses that a consequence of this characteristic is that the big picture of the user experience is often missing in Scrum projects. One of their challenges is to keep their vision of the user experience of the whole software, while small pieces of the software are developed in each sprint. Salah et al. [13] argue that the HCI community and agile community do not share the same understanding of how much design and how detailed it needs to be before the actual implementation starts. The agile developers argue that UX designers want "big design up front", meaning that the design needs to be complete and documented, but UX design iterative in nature [1]. The developers' concern is that the requirements will change so much, so designing big parts of the software up-front will be a waste of time because some parts of the design will never be used. Successful projects that have significant user interaction have found that some level of design is necessary before the implementation [19]. It has also been argued that the fundamental requirements from the users do not change that substantially, so designing the fundamental user interaction up-front will not be a waste of time [1]. A vision of the user experience needs to be made before the implementation starts, but it needs to be iterated during the whole Scrum project, like one of our UX specialists stresses. It seems like the HCI and the agile communities do not agree on how much and how detailed design is needed before the actual implementation starts.

Designing One Sprint Ahead of Implementation

Both our informants describe the need of designing the user experience some days before the implementation of one particular feature starts. Some researchers have suggested this [12,13,18] but there is a conflict in their guidelines on when the design and evaluation of the user interface should take place. Sy et al. suggests that design happens one sprint ahead the implementation and the evaluation one sprint after the implementation [18], but Salah et al. suggest that the particular UI is designed two sprints before implementation and evaluated one sprint ahead [13]. Both our informants seem to design and evaluate before the implementation of particular feature. One our UX specialist in our study stresses that getting a clear vision of the user experience at the very beginning of a project is vital. When designing during the project the UX specialist uses the vision as a reference point. If requirements change the vision is changed too. It is also noticeable that low fidelity prototypes are used by the UX specialist as a means to evaluate UX early on in the project.

The Collaboration - UX Specialists and the Team

Both our informants believe that UX specialists need to work closely with the developers. It has been suggested that the UX design should happen in a parallel track to the development track [12,13,18]. Still, the UX specialists should view themselves as a part of the team because the team needs to include everyone necessary to go from idea to implementation [3]. According to our informants the real life situation is not necessarily as black and white as described in the literature. They describe their roles as being both in and outside the development teams. Our informants use informal ways of collaborating with the developers, like what is practiced in general in agile development. Both our UX specialists stress that maintaining good co-operation with the developers is vital. It is also noticeable that the UX specialists do only mention a few documents in their way of explaining their work. It seems that most collaboration is informal and oral.

Responsibility for UX in Scrum is Complex

One of our UX specialists explains the big picture of UX is missing also because of the lack of responsibility for UX in Scrum projects. One interesting aspect in software development is defining the responsibility for particular activities. Responsibility here may refer to either the state of having a duty to deal with something, or the state of being accountable or to blame for something. This can be seen as either a rule based view of responsibility, or a consequence based view, as in Gotterbarn [5]. This problem can also be found in other system development processes, as is reported in Boivie et al. [2]. The notion of responsibility for UX is closely related to discussions of responsibility generally in social science in relation to groups. Here phenomena such as "the diffusion of responsibility" and the notion of "somebody else's problem" are interesting to investigate. Diffusion of responsibility is a social phenomenon, which might occur, in larger groups, where no one in the group takes responsibility for phenomena. When a task is placed before a group of people, there is a tendency for each individual to assume someone else will take responsibility for it—so no one does. This is a negative outcome that might occur in groups where responsibility is not clearly assigned. Previous research in the area have indicated that the diffusion of responsibility might have negative effects in systems development [5].

CONCLUSION

It is hard to make any general conclusions from our study because we only analysed the interviews with two UX specialists. Still, it can be concluded that working on

projects using the software development process Scrum affects the UX specialists' way of working and their possibilities to conduct UX evaluation. Furthermore, the challenges that these two UX specialists are facing while planning, conducting and describing the results of UX evaluation are considerably affected by the overall values of Scrum especially that Scrum is feature oriented, and informal co-operation in the team is emphasised.

ACKNOWLEDGMENTS

We would like to thank COST Action IC0904, Twintide, for the financial support that they have provided. Also, we would like to thank the two informants in the study that took their time to participate.

REFERENCES

1. Beyer, H. User-Centered Agile Methods. *In:* Carrol, J. M. ed. *Synthesis lectures on human-centered informatics,* Morgan & Claypool Publishers (2010).
2. Boivie, I., Gulliksen, J. and Göransson, B. The lonesome cowboy: A study of the usability designer role in system development. *Interacting with Computers,* 18(4), (2006) 601 – 634.
3. Cohn, M. *Succeeding with agile*, Addison-Wesley, USA, (2010).
4. Detweiler, M. Managing UCD within agile projects. *Interactions,* 14(3), (2007), 40-42.
5. Gotterbarn, D. Informatics and professional responsibility. *Science and Engineering Ethics* **7**(2), (2001), 221-230.
6. Hassenzahl, M. The thing and I: Understanding the relationship between user and product. In: K. Blyth, P.C. Overbeeke, A.F. Monk, P.C. Wright eds., *Funology: From usability to enjoyment.* Kluwer Academic Publishers, (2003) 1-12.
7. ISO IS 9241-11. *Ergonomics of human system interaction - Part 11: Guidance of usability.* ISO, Switzerland, (1998).
8. ISO IS 9241-210. *Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems,* ISO, Switzerland, (2010).
9. Larusdottir, M.K., Haraldsdottir, O. and Mikkelsen, B. User involvement in Icelandic software industry. In: *Proc. I-Used 2009 workshop at INTERACT 2009,* (2009), 51 – 52.
10. Larusdottir, M. K., Bjarnadottir, E., Gulliksen, J. The Focus on Usability in Testing Practices in Industry. In: *Proc. HCI Symposium at WCC 2010,* (2010), 98-109.
11. Law, E.L.-C. The measurability and predictability of user experience. In: *Proc. 3rd ACM SIGCHI symposium on engineering interactive computing systems,* ACM Press (2011).
12. Miller, L. Case study of customer input for a successful product. In: *Proc. Agile 2005,* IEEE Computer Society (2005), 225-234.
13. Salah, D. and Petrie, H. Towards a framework for integrating user centered design and agile software development processes. *In: Proc. Irish CHI 2009,* (2009).
14. Schwaber, K., 1995. Scrum development process. *In: SIGPLAN Notices,* 30(10).
15. Singh, M. U-SCRUM: An agile methodology for promoting usability. *In: Proc. AGILE '08,* IEEE Computer Society, (2008).
16. Silverman, 2011. Qualitative Research, 3rd edition, Sage Publications Ltd., UK.
17. Sohaib, O. and Khan, K. Integrating usability engineering and agile software development: A literature review. In: *Proc. ICCDA 2010,* (2010).
18. Sy, D. Adapting usability investigations for agile user-centered design. *Journal of usability Studies,* 2(3), (2007), 112-132.
19. Takats A. and Brewer N. Improving Communication between Customers and Developers. In: *Proc. Agile 2005,* IEEE Computer Society, (2005).
20. Vermeeren, A.P.O.S., Law, E.L.-C., Roto, V., Obrist, M., Hoonhout, J. and Väänänen-Vainio-Mattila, K. User experience evaluation methods: Current state and development needs. In: *Proc. NordiCHI 2010,* ACM Press (2010).
21. Williams, H. and Ferguson, A.The UCD perspective: Before and after agile. *In: Proc. Agile 2007.* IEEE Computer Society, (2007) 285-290.
22. Wolcott, H. 2009. Writing up Qualitative Research. California, Sage Publications Inc.

The early phases of UX: Why they are important (more than evaluation), and what they are?

Timo Jokela
Helsinki, Finland
timo.jokela@gmail.com

ABSTRACT
Evaluation is a key activity in developing high level UX. This paper argues, however, that the early phases form the basis for UX, and evaluation should be seen only as a supportive role in ensuring UX. Four (4) main early activities are identified and their challenges briefly discussed.

INTRODUCTION
Good *user experience (UX*) and *usability* are key factors for successful products and systems.

For developing good UX and usability, a general paradigm 'user-centered design' (or 'human-centered design') is being established over the last few years. A well-known reference is ISO 9241-210 (ISO/IEC, 2010).

ISO 9241-210 identifies four main activities: (1) understanding and defining the context of use, (2) specifying the user requirements, (3) producing design solutions, and (4) evaluating the design. These activities are of general nature, and more or less included in other models of UX development.

Much of the research and pragmatic work around UX – and also usability – is around the evaluation activity (for example this workshop). Although evaluation important, this paper argues that there are several reasons why the earlier activities are probably even more critical for successful UX.

Four such early UX activities are identified, and there challenges briefly discussed.

EVALUATION IS ALWAYS A LATE, REACTIVE ACTIVITY
Before any evaluation can be made, some design solutions need to be produced. Moreover, the design solutions need typically be developed to be 'working ones' in the sense that a user can try to use them.

The only way to make evaluation effective is to make changes to the design solutions, based on the results of the evaluation.

Making changes to design solutions is always a reactive activity where resources are needed. The more and bigger problems are found in the design solutions, the more work is needed for redesign and possibly for many sets of iterative design – evaluate cycles.

Further, as Cooper (Cooper, 2003) argues, evaluation is useful for correcting small problems. If major design problems are found, their redesign is always a big challenge.

If the product has ambitious UX targets, the less effective evaluation-driven development will be. The author argues that in-depth 'thinking' is needed for the generation of design solutions with high level UX. Evaluations reveal which design solutions work and which do not; but evaluations are not 'design solution generators'.

Overall, it is more effective if the design solutions would be of 'good quality' before any UX evaluation is started. This would lead to less need for changes and redesign during evaluation.

Well-thought and elaborated design solutions reduce the need for redesign. But how one can achieve high-level design solutions, before evaluation? In the following, key activities are identified.

KEY PRE-EVALUATION ACTIVITIES OF UX
What should then be done, to produce design solutions of good quality form the beginning, before any UX evaluation?

In the following, four interrelated early activities are identified that can, and should be done. But each of them is challenging.

1. Defining the desired business impact of UX
This is a key, fundamental activity. Before any project starts, one should define what do we want to achieve with good UX in the first place? What is the desired *business impact* of UX?

This is a business issue, and is dependent on the specific business context of the product/ system. The desired UX impact should be defined in a measurable way.

As an example, in one of the author's projects, a desired usability impact was defined as to reduce 90% of users' support calls, compared with the old system.

The author argues that similar impact targets should be defined for UX, too. As said, this activity is very much business related: the appropriate impact measures and target values are business related and a business decision.

This is an important activity for guiding and resourcing UX work. The more important UX is for business, the more resources one can be expect from the business management for UX work.

2. Understanding the system's and users' world

Understanding *users' world* is a well-known activity. It is called the *definition of context of use* in ISO 9241-210: know users' goals, tasks and environments of use. Well-known techniques for understanding user's world (work) are interviews and contextual inquiry (Holzblatt, 1993). This is naturally an important activity.

The author introduces another, complementary activity: defining *system's world*. This is even a more profound one to be defined. System's world is about defining what is to be built.

The background for this activity is the author's experience in consulting work. When joining system development projects and trying to understand 'what system' is to be built, the case is always that no one in the project team can explain it in a systematic and analytical manner. Not even persons who have worked in the domain for many years.

The author argues – although has not done literature studies or such – that this important activity is not generally recognized. In this paper, the author does not give a more elaborated definition for what is 'system's world' – because the author does not have it. The author has experience on carrying out this activity and modeling the results (system's world), with absolute excellent customer feedback. But when asking the customers to describe, "what did we exactly produce", they cannot find any name or term to describe it.

It is obvious that the designers need to understand 'what is to be developed' to be able to produce good design solutions. If this knowledge is weak, it is likely that the design solutions include (major) problems.

3. Defining measurable UX targets and giving incentives for achieving them

This activity is to define 'how good UX' we want to achieve. This activity transforms the desired UX business impact and the understanding of users' and system's world into concrete, measurable UX design targets.

Further, it is useful if the design team gets some incentives for achieving the defined UX targets. The more challenging the targets are, the higher incentives the business management should consider.

The usefulness of UX targets with incentives is that the targets drive the design team for good solutions from the very beginning of the project (project teams anyway always have limited budgets and tight time scales). If the UX targets are ambitious, the design team knows from the beginning that 'any design' would not be acceptable.

For measurable UX targets, one needs to define what is the *measure*, what is the *measuring instrument*, and what is the *target value*. For example, a measure may be a SUS (System Usability Scale (Brooke, 1986)). Measuring instrument defines how SUS evaluation is exactly conducted (e.g. with how many and what kind of users, and in what kind of context). The target value defines the desired level of UX, e.g. the target may be '90' of the average SUS results.

A key challenge here is: how to define the appropriate UX measures and target values?

4. Designing high-quality design solutions

This is the ultimate and decisive activity. The designers need to transform their in-depth knowledge of users' world (activity 2) into design solutions that meet the UX targets (activity 3).

This is obviously dependent on the designers' talents, creativity and knowledge of HCI. But at more detailed level, the big question is, how to do this? How to transfer a UX target such as "the average SUS score must be at least 90" into a design solution?

The "trial and error" – i.e. design and evaluate with users - approach might work. But the author has a hypothesis – but no evidence - that in-depth 'thinking' when creating the design solutions before evaluations, is required as a basis.

IMPLICATIONS FOR RESEARCH

The author's understanding is that these four activities are remarkably less in the focus in UX research than UX evaluation.

The author has some experience and solutions for these activities in cases where usability has been in the development focus. The challenge is not easy, and for developing good UX the challenge may be even more demanding.

The activity 2 – understanding the system's and users' world – probably is quite the same, no matter whether usability or UX is the design focus. But defining valid measures for the UX business case, and valid measurable UX targets may be very challenging.

In summary, the author proposes that following activities are key ones for designing good UX, and more research is

needed for how to do these in an effective and efficient way:

1. How to define the desired business impact of UX?
2. How to get an understanding on the system's and users' world?
3. How to define UX targets for design?
4. How to transform this knowledge into design solutions?

REFERENCES

Brooke, J. (1986). SUS - A "quick and dirty" usability scale. Digital Equipment Co. Ltd.

Cooper, A. (2003). *About Face 2.0: The Essentials of Interaction Design*. Wiley.

Holzblatt, K. (1993). Contextual Inquiry: A Participatory Technique for System Design. In D. Schuler & A. Namioka (Eds.), *Participatory Design: Principles and Practices*. Lawrence Erlbaum Associates.

ISO/IEC. (2010). 9241-210 Human-Centred Design Processes for Interactive Systems. ISO/IEC 9241-210: 2010 (E).

AHP Supported Evaluation of LMS Quality

Bojan Srđević
University of Novi Sad
Trg D. Obradovica 8,
21000 Novi Sad,
Serbia
bojans@polj.uns.ac.rs

Matija Pipan
Jozef Stefan Institute
Jamova cesta 39,
1000 Ljubljana,
Slovenia
matic@e5.ijs.si

Zorica Srđević
University of Novi Sad
Trg D. Obradovica 8,
21000 Novi Sad,
Serbia
srdjevicz@polj.uns.ac.rs

Tanja Arh
Jozef Stefan Institute
Jamova cesta 39,
1000 Ljubljana,
Slovenia
tanja@e5.ijs.si

ABSTRACT
Learning Management System (LMS) provides a platform for an on-line learning environment by enabling the management, delivery, and tracking of the learning process and learners. Selection of the most suitable method is usually prolonged by the time and effort consuming evaluations of numerous features of LMS. To reduce the number of features and at the same obtain a reliable result from an evaluation, we propose a decomposition of this complex problem to more easily comprehended sub-problems that can be analyzed independently through a multi-criteria method called Analytic Hierarchy Process (AHP). To verify the approach, an expert is asked to use AHP on an originally developed reduced hierarchy of the problem of selecting the most appropriate LMS for the student target group. Results of the application are compared with the results obtained by the DEXi multi-criteria model.

Keywords: LMS, Evaluation, Analytic Hierarchy Process

INTRODUCTION
The Organization for Economic Cooperation and Development defined LMS technology as a technology used by instructors to build and maintain courses. It features personal communication via email, group communication via chatting and forums, posting content including syllabi, papers, presentations and lesson summaries, performance evaluation via question and answer repositories; self-assessment tests, assignments, quizzes and exams, instruction management via messaging, grade posting and surveys, and more.

There are many LMS systems on the market that can be obtained for free and are Open Source (i.e. Moodle, Sakai, Claroline, ATutor, etc.) or through payment (i.e. Blackboard, WebCT, Clix, and many others). All of them support many different features which can be used as evaluation criteria and analyzed from different aspects [6]:

1. Pedagogical aspect
2. Learner environment
3. Instructor tools
4. Course and curriculum design
5. Administrator tools and
6. Technical specification.

Pedagogical criteria can, for example, include [15]: Learner control, Learner activity, Cooperative/ Collaborative learning, Goal orientation, Applicability, Added value, Motivation, Valuation of previous knowledge, Flexibility and Feedback.

On the other hand, Kurilovas [12] groups technical criteria as follows:

1. Overall architecture and implementation: Scalability of the system; System modularity and extensibility; Possibility of multiple installations on a single platform; Reasonable performance optimizations; Look and feel is configurable; Security; Modular authentication; Robustness and stability; Installation, dependencies and portability;
2. Interoperability: Integration is straightforward; LMS standards support (IMS Content Packaging, SCORM);
3. Cost of ownership;
4. Strength of the development community (for open source products): Installed base and longevity; Documentation; End-user community; Developer community; Open development process; Commercial support community;
5. Licensing;
6. Internationalization and localization: Localizable user interface; Localization to relevant languages; Unicode text editing and storage; Time zones and date localization; Alternative language support;
7. Accessibility: Text-only navigation support; Scalable fonts and graphics; and
8. Document transformation.

It is obvious that selection of the most suitable LMS is a complex task that involves defining the evaluation criteria and selecting a method for criteria evaluation that will be systematic, comprehensive, easy to use, etc.

Once defined, the criteria can be evaluated using a self-evaluation questionnaire that employs a 7-point Likert scale 1 (strongly disagree) – 5 (strongly agree), 6 (not applicable), 7 (don't know) [7, 13, 14, 15]. Other evaluation tools include MS-Excel spreadsheets application [1], fuzzy logic [6], an expert system shell for multi-attribute decision support DEXi [2], a hybrid Multi-criteria decision-making (MCDM) model based on factor analysis and DEMATEL [21] etc. The number of features for evaluation is usually very high in all these applications (e.g. 57 in Pipan [16]; 52

offered in Cavus [6]). To evaluate such a great number of features, a significant amount of time and effort is required of the evaluator.

We believe that reliable results can be obtained with fewer criteria if the problem is decomposed in order to more easily comprehended sub-problems that can be analyzed independently, i.e. presented as a hierarchy. One of the most popular methods that deal with decision hierarchies is Analytic Hierarchy Process (AHP) [17] and we propose this method for the evaluation of selected LMS products because: (1) it supplies management in both education and industry with a less complex and more appropriate and flexible way to effectively analyze LMSs, (2) it supports their selections of an appropriate product, and (3) achievement of a higher level of e-learner satisfaction [18]. Other advantages of AHP that should be emphasized are that AHP provides a measure of consistency of the evaluator and that it can be used for participative evaluation of LMS product.

To verify AHP applicability, an expert is asked to use AHP on an originally developed hierarchy of the problem of selecting the most appropriate LMS for the student target group. Consistency of the expert is checked throughout the process. At the end, results of the evaluation are compared with results presented in [16].

AHP IN BRIEF

Main features

One of the key issues in decision making is eliciting judgments from the decision maker (DM) about the importance of a given set of decision elements. If a problem can be structured hierarchically, then a certain ratio scale can serve as an efficient tool to enable this hierarchy by performing pair-wise comparisons. The core of AHP [17] lies in presenting the problem as a hierarchy and comparing the hierarchical elements in a pair-wise manner using Saaty's 9-point scale, Table 1.

This way, the importance of one element over another is expressed in regards to the element in the higher level. The AHP is a multi criteria optimization method which creates so-called local comparison matrices at all levels of a hierarchy and performs logical syntheses of their (local) priority vectors. The major feature of AHP is that it involves a variety of tangible and intangible goals, attributes, and other decision elements. In addition, it reduces complex decisions to a series of pair-wise comparisons; implements a structured, repeatable, and justifiable decision-making approach; and builds consensus.

Judgment term	Numerical term
Absolute preference (element *i* over element *j*)	9
Very strong preference (*i* over *j*)	7
Strong preference (*i* over *j*)	5
Weak preference (*i* over *j*)	3
Indifference of *i* and *j*	1
Weak preference (*j* over *i*)	1/3
Strong preference (*j* over *i*)	1/5
Very strong preference (*j* over *i*)	1/7
Absolute preference (*j* over *i*)	1/9
An intermediate numerical values 2,4,6,8 and 1/2,1/4,1/6,1/8 can be used as well	

Table 1: The fundamental Saaty's scale for the comparative judgments

In standard AHP, an eigenvector (EV) method is used for deriving weights from local matrices; the EV is called the prioritization method, and the computational procedure is consequently called prioritization. After local weights are calculated at all levels of the hierarchy, a synthesis consists of multiplying the criterion-specific weight of the alternative with the corresponding criterion weight and summing up the results to obtain composite weights of the alternative with respect to the goal; this procedure is unique for all alternatives and all criteria.

AHP is aimed at supporting decision-making processes in both individual and group contexts. In later cases various aggregation schemes are applicable, e.g. AIJ and AIP [9], as well as various consensus reaching procedures are easy to implement. This issue is out of scope here; namely, the paper deals strictly with an individual application of AHP.

Measuring consistency

The DM makes judgments more or less consistently depending not only on his knowledge of the decision problem itself, but also on his ability to remain focused and to ensure that his understanding of the cardinal preferences between elements will always, or as much as possible, be formalized properly while using a verbal scale or related numerical ratios [20]. For example, if the Saaty's 9-point ratio scale is used, the question could be: will the DM put $a_{ij} = 3$, or $a_{ij} = 2$, if he considers element E_i slightly more important than E_j? Or, if there are seven elements to be compared, then matrix A is of size 7x7, and the question could be: is the DM really capable to preserve consistency while comparing head-to-head 21 times all pairs of elements? How is the DM to override the imposed difficulty with Saaty's scale when he compares elements E_i and E_k, after he has judged the elements E_i and E_j, and E_j and E_k? If he has already made the judgments $a_{ij} = 3$ and a_{jk}

= 4, he should logically put a_{ik} = 12 without any further judging because a simple transitivity rule applies: $a_{ik} = a_{ij}a_{jk}$ = 3x4 = 12. Because the maximum value in Saaty's scale is 9 for declaring the absolute dominance of one element over the other, there is a problem in attaining consistency while judging certain elements. The inconsistencies generally accumulate until the need for their measuring arises.

Consistency analysis of the individual DM can be based on the consistency ratio (CR) defined by Saaty [17], and the total L^2 ED for each comparison matrix. Whichever method is used to derive the priority vector from the given local AHP matrix [19], if it already has all the entries elicited from the DM, measuring consistency is necessary in order to ensure the integrity of the outcomes.

Standard AHP uses EV, the prioritization method, and the consistency coefficient *CR* to indicate the inconsistency of the DM [17]. The other commonly used consistency measures are the total Euclidean distance, and minimum violations measure.

The *CR* is calculated as a part of the standard AHP procedure. First, the consistency index (*CI*) is calculated using the following equation:

$$CI = \frac{\lambda_{\max} - n}{n-1} \tag{1}$$

where $\lambda_{\max}$ is the principal eigenvalue of the given comparison matrix. Knowing the consistency index and random consistency index (*RI*) defined also by Saaty [17], the consistency ratio is obtained:

$$CR = \frac{CI}{RI}\,. \tag{2}$$

Saaty [17] suggested considering the maximum level of the DM's inconsistency to be 0.10; that is, *CR* should be less or equal to 0.10.

EXAMPLE APPLICATION

Problem statement

The problem is stated so as to assess and rank by applicability the three e-Learning Management Systems based on three typical qualitative criteria and a number of qualitative sub criteria. An expert is asked to perform the decision making processes by applying the AHP model.

Hierarchy of the problem

An original hierarchy of the problem [16] consists of five levels: goal – criteria set – sub criteria set (4+4+3 per criterions in upper level) represented by specific groups of attributes – sub sub criterions (24 in total under sub criterions), represented by groups of more detailed attributes – and three alternatives (LSMs). In order to reduce the number of decision elements, the fourth level in the hierarchy (sub sub attributes) is avoided and thus the reduced hierarchy of the problem is created as shown in Figure 1.

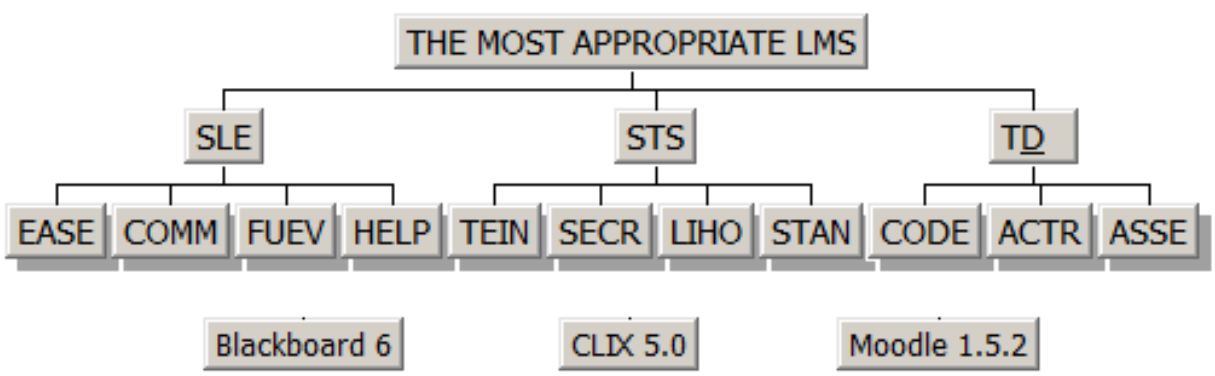


Figure 1: Reduced hierarchy of the decision problem

Identify LSM with the best applicability characteristics

Criteria set (with attributes as sub-criteria)

The set of criteria is the key component of the decision-making model. In creating the model [16], an attempt is made to meet the requirements set by Bohanec & Rajkovič [5] by taking into account the principle of criteria integrity (inclusion of all relevant criteria), appropriate structure, non-redundancy, comprehensiveness and measurability [4]. Comprehensiveness means that all the data about the subject are actually present in the database. Non-redundancy means that each individual piece of data exists only once in the database. Appropriate structure means that the data are stored in such a way as to minimize the cost of expected processing and storage [3].

The criteria set is stated under three main scopes: Student's learning environment, System, technology & standards, and Tutoring & didactics. These three scopes represent the global skeleton of the multi-attribute model with attributes (considered as sub-criterions) associated with each criterion.

(1) SLE (Student's learning environment): The first scope is adopted as the first criterion and declared as the Student's learning environment. It is composed of four basic attributes:

- (**EASE**) *Ease of use*
- (**COMM**) *Communication*
- **(FUEV)** *Functional environment* and
- *(***HELP***) Help*.

(2) STS (System, technology & standards category): The second group of attributes is grouped into the *System, technology & standards* category. These groups of criteria are assessed through four basic attributes:

- (**TEIN**) *Technological independence*. The attribute of technological independence is used for the evaluation of an LMS from the prospective of its technological accessibility, which is a pre-condition that has to be met if we wish to talk about system applicability and efficiency.
- (**SECR**) *Security and privacy. The Security and privacy* criterion focuses on two issues: *User security and privacy* and *security and privacy of an LMS*. User security and privacy should be at the forefront of attention; therefore an LMS must keep communication and personal data safe and avoid dangers and attacks on user computers. Application security and privacy assessment is made using authentication, authorization, logging, monitoring and validation of input.
- (**LIHO**) *Licensing & hosting*. Add description.
- (**STAN**) *Standards support.* It is also important to consider *e-learning standards* – standards for description of learners' profiles and standards for the description of learning resources [11]. In the context of e-learning technology, standards are generally developed to be used in system design and implementation for the purposes of ensuring interoperability, portability and reusability, especially for learning resources as they require for their preparation qualified professionals and are very time [10].

(3) T&D (Tutoring & didactics): Third group of criteria is merged into *Tutoring & didactics*. The tutor's quality of environment is assessed using the:

- (**CODE**) *Course development*,
- (**ACTR**) *Activity tracking* and
- (**ASSE**) *Assessment criteria*.

Activity tracking undoubtedly provides important support to the tutor in the learning process. Here we have focused on monitoring students in the process of learning and the possibility of displaying students' progress, analysis of presence data, sign-in data and time analysis.

Decision alternatives

The multi-attribute decision making model was completed with three learning management systems (LMS):

A1. Blackboard 6 (www.blackboard.com): Blackboard is among the most perfected and complex LMSs on the market. The system offers various communication options (both synchronous and asynchronous) within the learning environment. The Blackboard LMS is designed for institutions dedicated to teaching and learning. Blackboard technology and resources power the online, web-enhanced, and hybrid education programs at more than 2000 academic institutions (research university, community college, high school, virtual MBA programs etc.). Blackboard has 5,500 clients representing 200 million users (2.5 million from its largest, hosted client; 100,000 from its largest, self-hosted client) in 60 countries [8].

A2. CLIX 5.0 (www.im-c.de): CLIX is targeted most of all at big corporations because it provides efficient, manageable, connected and expandable internet-based learning solutions. This scalable, multilingual and customizable software aims at providing process excellence for educational institutions. For educational administrators, CLIX offers powerful features for course management and distribution. Additionally, it provides personalized learning paths for students, a tutoring centre for lectures and a whole bunch of innovative collaboration tools for both user groups, e.g. a virtual classroom. Altogether, CLIX makes planning, organizing, distributing, tracking and analyzing of learning and teaching a smooth and efficient process.

A3. Moodle 1.5.2 (www.moodle.org). Moodle is a free, open source PHP application for producing internet-based educational courses and web sites on any major platform (Linux, UNIX, Windows and Mac OS X). The fact that it is free of charge is especially attractive for schools and companies which always lack resources for the introduction of new learning technologies. Furthermore, the Moodle

system is not only price-efficient – it can easily be compared to costly commercial solutions on all aspects. Courses are easily built up using modules such as forums, chats, journals, quizzes, surveys, assignments, workshops, resources, choices and more. Moodle supports localization, and has so far been translated into 34 languages. Moodle has been designed to support modern pedagogies based on social constructionism, and focuses on providing an environment to support collaboration, connected knowing and a meaningful exchange of ideas. It has nearly 54,000 registered sites (over 9,800 from the U.S.) representing over 200 countries, 44.3 million users, and 4.6 million courses. Moodle's wide spread international use, coupled with its continued growth over the past six years, has made it the leading open source LMS solution.

Evaluation of decision elements

After a brief explanation of basics and concepts of AHP, the expert compared in pairs first criteria versus goal, then sub criteria versus criteria, and finally alternatives with respect to each of the sub criteria. Comparison matrices and related calculated local weights of decision elements are presented in Figures 2-3.

Criteria vs. Goal: CR=0.021

	SLE	STS	T&D	Weights
SLE	1	5	4	0.683
STS		1	1/2	0.117
T&D			1	0.200

Figure 2: Criteria versus goal and their local weights

Sub criteria vs. Criterion SLE *(Student's learning environment):* CR=0.087

	EASE	COMM	FUEV	HELP	Weights
EASE	1	6	4	8	0.631
COMM		1	2	5	0.188
FUEV			1	4	0.135
HELP				1	0.046

Sub criteria vs. Criterion STS *(System, technology & standards category)* CR=0.021

	TEIN	SECR	LIHO	STAN	Weights
TEIN	1	4	7	2	0.501
SECR		1	3	1/3	0.129
LIHO			1	1/6	0.055
STAN				1	0.316

Sub criteria vs. Criterion T&D *(Tutoring & didactics):* CR=0.008

	CODE	ACTR	ASSE	Weights
CODE	1	1/2	2	0.297
ACTR		1	3	0.540
ASSE			1	0.163

Figure 3: Sub criteria versus criteria and their local weights

After the local weights (W) of all decision elements are calculated, a synthesis is performed to obtain composite weights of the alternatives with respect to goal (Table 2).

	Weights
Blackboard 6	0.257
CLIX 5.0	0.590
Moodle 1.5.2	0.152

HCR=0.059

Table 2: Final (composite) weights of alternatives

The alternative with the highest final weight is CLIX 5.0 (0.590) and can be considered as the most applicable LMS for the students. The second ranked alternative is Blackboard, while Moodle 1.5.2 is the least applicable LMS.

It is worthy to mention that the expert was very consistent during the whole evaluation process. Overall HCR is 0.059.

DISCUSSION AND CONCLUSIONS

One of the important problems in the field of e-learning is the selection of an appropriate LMS that will satisfy most of the users' preferences and requirements. The complexity of the problem is increased due to the growing number of LMS each year and also due to the number of features that should be taken into account while evaluating each LMS.

To reduce that complexity and facilitate selection of an appropriate LMS, we propose a decomposition of the problem to more easily comprehended sub-problems that the evaluator can analyze independently. The AHP methodology based on pair-wise comparison of decision elements on one hierarchy level was found to be appropriate for such analysis. Also, the final result of AHP application, which found CLIX 5.0 to be the most applicable LMS, proved that the proposed approach was justified: the reduced hierarchy and use of AHP led to the same result as the one provided by the DeXi evaluation of 57 criteria.

If AHP and DeXi are further compared, it should be also emphasized that:

a) AHP treats consistency of the DM (DMs), DEXi does not.
b) DEXi uses a simplified 3-point scale (linguistic semantic statements such as low, average and high); AHP most commonly uses Saaty's 9-points (fundamental) scale; other scales also in use are geometric (Lootsma's), balanced, Ma-Feng scale etc. In practical implementations the first seems easier, especially if many decision elements have to be considered (assessed). If one has to compare 7 or more elements at a time by using any AHP scale, it can be time consuming and inconsistent (e.g. due to 'short term memory' and/or 'brain channel capacity' limits).
c) AHP produces cardinal information represented by weights at all hierarchical levels of the decision problem; DEXi does it very approximately and with limited theoretical justification.
d) Both AHP and DEXi run easily on any standard PC platform.

Both AHP and DEXi can be used in individual and group d-m frameworks. In group contexts AHP enables the direct application of various aggregation schemes (e.g. AIJ, AIP; different weights allocated to DMs; different consensus reaching procedures) while in the use of DEXi, there are no implemented aggregation schemes.

REFERENCES

1. 3Waynet Inc. *LMS evaluation tool user guide*. Under License to Commonwealth.
2. Arh, T., Blazic, J.B. Application of multi-attribute decision making approach to learning management systems evaluation. *J Comput 2*, 10 (2007), 28–37.
3. Awad E. M., Gotterer M. H.. Database management. Danvers, MA: Boyd & Fraser, 1992.
4. Baker, D., Bridges, D., Hunter, R., Johnson, G., Krupa, J., Murphy, J., Sorenson, K. *Guidebook to Decision-Making Methods*. USA, WSRC-IM-2002-00002, 2002, Department of Energy.
5. Bohanec, M., Rajkovič, V. Multi-Attribute Decision Modeling: Industrial Applications of DEX. *Informatica 23* (1999), 487−491.
6. Cavus, N. The evaluation of Learning Management Systems using an artificial intelligence fuzzy logic algorithm. *Advances in Engineering Software 41* (2010), 248–254.
7. Chen, Y. , Hwang, R., Wang, Ch. Development and evaluation of a Web 2.0 annotation system as a learning tool in an e-learning environment . *Computers & Education 58*, 4 (2012), 1094-1105.
8. Cobb, J., Steele, C. Association learning management systems 2011: *Special Blackboard edition*. Tagoras (2011), 30-31.
9. Forman, E., Peniwati, K. Aggregating individual judgments and priorities with the analytic hierarchy process. *Eur J Oper Res 108* (1998), 165–169.
10. IEEE Computer Society. Learning Technology Standards Committee LTSC, IEEE, Draft Standard for Learning Objects Metadata (LOM) (Tech. Rep. 1484.12/D4.0). Washington, DC, USA, IEEE Computer Society, 2002.
11. Jerman Blažič, B., Klobučar, T. Privacy provision in e-learning standardized systems: status and improvements. *Computer standards & interfaces 27* (2005), 561−578.
12. Kurilovas, E. Methods of Multiple Criteria Evaluation of the Quality of Learning Management Systems for Personalised Learners Needs. *Presented at the EC-TEL'2009 Workshop "Learning Management Systems meet Adaptive Learning Environments"*, September 29 (2009), Nice, France.
13. Lee, J., Lee, W. The relationship of e-Learner's self-regulatory efficacy and perception of e-Learning environmental quality. *Computers in Human Behavior 24*, 1 (2008), 32-47.
14. Naveh, G., Tubin, D., Pliskin, N. Student LMS use and satisfaction in academic institutions: The organizational perspective. *The Internet and Higher Education 13*, 3 (2010), 127-133.
15. Nokelainen, P. An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. *Educational Technology & Society 9*, 2 (2006), 178-197.
16. Pipan, M., Arh, T., Jerman Blažič, B. (2010). The Evaluation Cycle Management - Method Applied to the Evaluation of Learning Management Systems. *Chapter In Integrating Usability Engineering for Designing the Web Experience: Methodologies and Principles* (Eds. Tasos Spiliotopoulos; Panagiota Papadopoulou; Drakoulis Martakos; Georgios Kouroupetroglou), IGI Global, Hershey PA, USA, 2010.
17. Saaty, T.L. The analytic hierarchy process. McGraw-Hill Inc, 1980.
18. Shee, D. Y., Wang, Y. Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications. *Computers & Education 50*, 3 (2008), 894-905.
19. Srdjevic, B. Combining different prioritization methods in analytic hierarchy process synthesis. *Computers & Operations Research 32* (2005), 1897–1919.
20. Srdjevic, B., Srdjevic, Z. Bi-criteria evolution strategy in estimating weights from the AHP ratio-scale matrices. *Applied Mathematics and Computation 218* (2011), 1254-1266.
21. Tzeng, S., Chiang, C., Li, C. Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL. *Expert Systems with Applications 32*, 4 (2007), 1028-1044.

To Measure or Not to Measure UX: An Interview Study

Effie Lai-Chong Law
University of Leicester
Dept. of Computer Science
LE1 7RH Leicester, UK
elaw@mcs.le.ac.uk

Paul van Schaik
Teesside University
School of Psychology
TS1 3BA Middlebrough, UK
P.Van-Schaik@tees.ac.uk

ABSTRACT
The fundamental problem of defining what UX is (or is not) has a significant influence on another challenging question: *to measure or not to measure UX constructs*. The answer of most, if not all, UX researchers and practitioners, would probably be "It depends!" As we were motivated to find out "depending on what", we conducted semi-structured interviews with eleven UX professionals where a set of questions in relation to UX measurement were explored. Participants expressed scepticism as well as ambivalence towards UX measures and shared anecdotes related to such measures in different contexts. To improve the interplay between UX evaluation and system development, a clear definition of UX, combining various data types, and robust education in UX concepts are deemed essential.

Author Keywords
User experience; Measurement; Interview; Feedback loop

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design; Evaluation; Measurement

INTRODUCTION & BACKGROUND
To measure or not to measure concepts of interest? A clear cut "*Yes!*" if this question is raised in the context of physical sciences whereas an ambiguous "*It depends!*" when it is addressed in the context of social sciences in general and the emerging research area of User Experience (UX) in particular. We aimed to explore such stipulations (i.e. 'depending on what') for UX measures and their implications to design and evaluation of interactive systems. To meet this purpose, we conducted an empirical study in which eleven UX researchers and practitioners were interviewed. In this paper we report some main findings of the study that are particularly relevant to understanding the interplay between UX measurement and iterative system redesign. Specifically, we adopt Hand's ([4], p.3) definition of measurement "*quantification: the assignment of numbers to represent the magnitude of attributes of a system we are studying or which we wish to describe.'*

The exploration of the issue of UX measurement was embarked on (e.g. [6]) after another, if not more, thorny issue of UX - its multiple definitions - had been examined [7]. In principle these two foundational issues should be solved in tandem. However, the definitional issue on UX remains unresolved, UX researchers and practitioners tend to select and adapt one of the many definitions out there to serve their particular goals and needs. The recent efforts of deepening the understanding of the theoretical roots of UX [10] can complement the earlier work on UX evaluation methods on the one hand [13] and the current operatonalisation work for UX measurement on the other hand (e.g. [11]). As UX research studies have hitherto relied heavily on qualitative methods [1], the progress on UX measures has thus been slow. A plausible reason is the scepticism about the measurability of UX.

The field of HCI in which UX is rooted has inherited theoretical concepts, epistemological assumptions, values, and methodologies from a diversity of disciplines, ranging from engineering where measures are strongly embraced (cf. William Thomson's [14] dictum 'to measure is to know') to humanities where measures can be regarded as naïve or over-simplistic, especially when the concepts to be measured are ill-defined, leaving (too) much for interpretation [2]. As UX subsumes a range of fuzzy experiential qualities such as happiness, disgust, surprise and love, controversies and doubts about the measurability of UX are inevitable. A main divergence between two major camps of UX researchers is the legitimacy of breaking down experiential qualities into components, rendering them to be measured; it is rooted in the age-old philosophical debate on reductionism versus holism.

INTERVIEW ON UX MEASUREMENT
Instrument
The interviews were semi-structured with 12 questions grouped into three main parts. Part A comprises four background questions (Table 1).

Q1. Gender: Female, Male
Q2. Age: <=20, 21-30, 31-40, 41-50, >50
Q3: I am a: Practitioner, Researcher, Student, Other
Q4. How long have you worked in the area of UX? (Never, <1 year, 1-3 year, 3-5 year, >5 year). Please describe the topic and related work.

Table 1. Background questions

I-UxSED 2012. 14th October 2012, Copenhagen, Denmark

Part B comprises five questions on the measurability of UX qualities (Table 2). The inclusion of Q5 is to know if the respondent's understanding aligns with any of the existing definitions of measurement. For Q6, the rationale underpinning each statement varies. The first one was derived from the classic justification for measurement advocated by Thomson [14]. The second and third ones were two rather extreme views against UX measures expressed in some informal contexts (e.g. group discussion in a workshop). They were aimed to stimulate thoughts and should not be treated as scientific claims. In contrast, the fourth and fifth statements represent views on the potential uses of UX measures. They were deliberately broad in scope to stimulate discussions.

Q5. What is a 'measure'?
Q6. (a) Please rate your agreement with each of the following statements (5-point Likert scale); (b) Explain your ratings ▪ UX measures lead to increase of knowledge ▪ UX measures are insane ▪ UX measures are a pain ▪ UX measures are important for design ▪ UX measures are important for evaluation
Q7. (a) Name a specific experiential quality (e.g., fun, surprise) that is most relevant to your work; (b) Explain the relevance; (c) Do you think the named quality can be measured: If 'yes', describe how; If 'no', describe why.
Q8. (a) Name an experiential quality that you are (almost) certain is measurable; (b) How can it be measured and when (before/during/after interaction)? (c) Why are you so (almost) certain about its measurability? What is your reservation, if any?
Q9. (a) Name an experiential quality that you think (almost) impossible to measure; (b) Why do you think so? What is your reservation, if any?

Table 2. Five main questions on UX measures

The notion of "*experiential qualities"* is central for Q7, Q8 and Q9. In the simplest sense, they are referred to as feelings. In the broadest sense, they are related to the concept of emotional responses, as defined in the Components of User Experience (CUE) model [15], which are influenced by instrumental (i.e. usability) and non-instrumental qualities (i.e. aesthetic, symbolic and motivational). While CUE focuses more on evaluation, in the context of the design the notion of experiential qualities is defined as articulations of key qualities in the use of a certain type of digital artefact intended for designers to appropriate in order to develop their own work [8]. Note that in order to enable open discussion no definition was provided to the interviewees unless requests for clarification were solicited. Part C comprises three questions aimed to simulate in-depth discussion (Table 3).

Q10. Which theoretical arguments (e.g. reductionism) are for or against UX measurement?
Q11. Which methodological arguments (e.g. validity) are for or against UX measurement?
Q12. Which practical arguments (e.g. cost) are for or against UX measurement?

Table 3. Questions for in-depth discussions

Participant and Procedure

An invitation to the interview was circulated in the intranet of a university. Eight participants volunteered to take part in it. The other three participants were recruited by the first author via personal invitation. Their participations were also voluntary. They were designated as P1, P2 and so on. Seven of them were female, five aged between 31 and 40, another five between 41 and 50 and one above 50. All were researchers except P5, who was a practitioner. The job of eight of the participants was predominantly design-oriented, be it practical or theoretical, such as empathic design for house renovation, co-design for persuasive games, and design theories. The other three focused more on UX evaluation of interactive products such as mobile phone. Two of them have worked in UX for less than 1 year, three 1-3 years, five 3-5 years and one for than 5 years. All the interviews were conducted on an individual basis in English, audio-taped and transcribed subsequently.

RESULTS AND DISCUSSIONS

For analysing the data, we developed coding schemes for individual interview questions by applying thematic analysis [3] and the CUE model [15]. Due to limited space, here we do not report results of Q5 (What is a 'measure'?).

Statements on UX Measures

Given the small sample size, no inferential statistics of the ratings are computed. Justifications for the ratings are of higher relevance and the analyses are presented below.

UX measures lead to increase of knowledge (mean = 4.0, range: 2-5). When prompted to specify which kinds of knowledge would be increased, several were mentioned,

- references against which products can be compared;
- the extent to which the development goals is achieved;
- values to be delivered by certain design methods;
- information helpful for future projects;
- experience per se;

Ambivalence was observed, for instance: "There are ways to get knowledge about UX in a more meaningful way rather than using measures, but I still think that they are important." (P6). Besides, the need for including qualitative data as complementary knowledge was emphasized: "We should have both… qualitative is to know what the reason is for user experience and for the related design issue." (P8). Furthermore, conditions for benefiting from UX measures were specified: "It requires people using the measure, understand the measure and what it actually means… There might be people who are not trained to use UX measures, no matter how well we define the measures." (P5). This observation highlights the need for enhancing education and training in UX.

UX measures are insane (mean = 2.0, range: 1-4). A common view was that the insanity lies not in UX measures but rather in what claims to be made about them, especially when people do not understand such measure, intentionally misuse them, are unaware of their inherent limitations (e.g. incompleteness) or over-formalize them. There were also

concerns whether UX measures can explain why people experience something or have any use for design, as remarked by P11 (a designer):

> "… for the purpose of design, measuring variables up to a very high degree and intricate level of measurement might not be that purposeful because you have to translate the numbers back to design requirements, and I am not sure whether that works."

UX measures are a pain (mean = 3.27, range: 1 – 5). Pain inflicted was psychological rather than physical. Reasons for such pain varied with the phase of UX measurement. In the preparation phase, defining valid and meaningful metrics, which entailed deep and wide knowledge of various matters, was cognitively taxing and thus painful. For data collection, participant recruitment and time constraint were a pain for researchers, as illustrated by P4's remark: "We would not use half-an-hour to measure something but rather get some qualitative data out of participants." On the other hand, the intrusiveness and lengthiness of the procedure could be pain for users. For data analysis, statistical analysis was deemed challenging by four participants. This again is a clear implication for the training of UX. Interpretation of UX measures was another common concern: it could be an issue of lack of knowledge, confirmation bias, and attempts to draw implications from exact measures for design.

UX measures are important for design (mean = 4.0, range: 2-5). Participants' stance on this claim was ambivalent. They recognized that UX measures could help identify design constraints and justify design decisions by convincing developers and management, given that numbers could convey a sense of reliability. However, they stipulated the importance of UX measures in design with the need of combining with qualitative data, for instance:

> "I mean they are important, but I'd *not* base my design solely on UX measures... there are lot of things that I don't think that we can measure properly enough yet… it would cause too much work to get really really good measurement that would be our main basis for design… [UX measurement] would only be second; the first being an overall understanding of qualitative views we have found out from users." (P4)

> "If UX measures are clusters that are described through numbers or questionnaires, then they are *not* important for design, whereas if UX measures are, for instance, clusters of qualitative data and users' accounts, then they are important for design" (P11)

Some participants explicitly expressed their doubt about the role of UX measures in design, for instance:

> "I can see relatively little value of applying UX measures, because they don't really link to the product's attributes in most cases… they link it at an abstract level… it is hard to trace what the underlying causes for certain response. It is almost impossible if we just use UX measures without combining them with qualitative data" (P1)

Furthermore, one participant pointed out the differences between usability and UX measures:

> "… sometimes it is difficult to explain why we design like this even when we provide evidence. From usability point of view we can more easily give this measurement that it is better, but designing for UX is problematic. People with technical backgrounds have problems making the difference between UI and UX. They think they are the same thing." (P3)

In summary, the interplay between UX measures, which are common evaluation outcomes, and (re)design is ambiguous.

UX measures are important for evaluation (mean = 4.6, range: 2-5). On this claim the participants were somewhat less ambivalent. Supporting arguments such as justifying decisions, validating design goal, and giving reliability (cf. P2's remark: "If you only use the designer intuition, only use empathic interpretation, it is not very reliable for the rest of the world") were given. Some participants pointed out the time issue: in which development phase UX measures are taken and how much time the process of measuring is allowed, for instance:

> "… in industry-led cases they are more keen on fast phenomenon … the industrial people want to improve the design but not really want to provide input for the academic world in general" (P4)

There are also reservations about the role of UX measures in evaluation, for instance:

> "it's not been proven yet that [UX measures] can make any difference to outcomes…. I mean, they *could* be; certainly if you include traditional usability measures, then persistent task failure for many designs is going to be something you want to know about. But I don't think they're automatically important; they're all hinges around design objects" (P11)

Measurable and Non-measurable Experiential Qualities

In response to Q7, Q8 and Q9 (Table 2), participants identified different experiential qualities (EQ), which we categorized by the adapted CUE model [15]:

- *Instrumental qualities (NQ)* – "the experienced amount of support the system provides and the ease of use" (e.g. controllability, learnability, effectiveness);
- *Non-instrumental qualities (NIQ)* – "the look and feel of the system", including aesthetic, symbolic and motivational qualities ([15], p. 916; [9]);
- *Affective responses (AR)* – subjective feelings, motor expressions, and physiological reactions [12] arising from interacting with the system (NB: It broadens the scope implied by original notion of 'emotional reactions' to accommodate mildly affective responses with an artefact).
- *Evaluation* (cf. system appraisal) – long-term effects of interacting with the system on user affect, attitude and cognition;

Several interesting observations are noted:

i) All three EQs considered as non-measurable fall into the category of Evaluation; it seems implying that long-term effects of interaction are considered not amenable to measurement;

ii) No non-measurable instrumental and non-instrumental qualities were identified by the participants; this is not surprising as instrumental qualities are closely related to traditional software attributes that have explicitly been operationalised and operationlising non-instrumental

qualities such as aesthetic and symbolic has been endeavoured in recent UX research efforts (e.g. [5]);

iii) Fun is the EQ that was dually considered as measurable as well as non-measurable. This is somewhat surprising because game experiences of which fun is an integral part have been one of the hot topics in UX research where different attempts to measure fun have been undertaken (see the review in [1]). This observation underpinned P11's argument for the measurability of fun as it is a well-defined concept. In contrast, P1's counterargument referred to the complexity and multidimensionality of fun; reporting on overall fun after interaction seemed more plausible than on individual sub-constructs;

iv) Several high-level concepts were mentioned: 'hedonic quality' for measurability and 'long-term experience' and 'deep [sub]-conscious experience'; they do not fit into any of the categories.

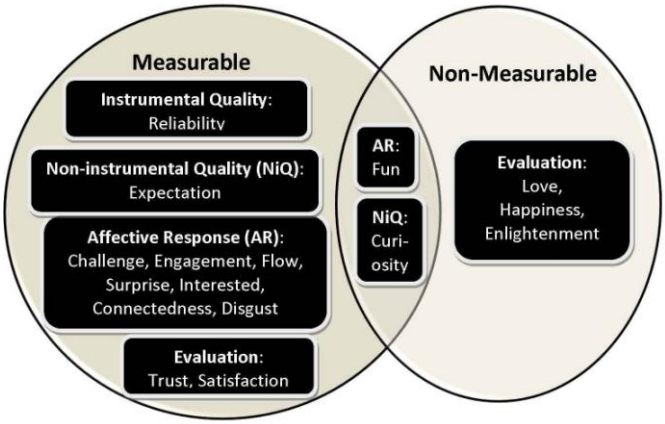


Figure 1: Categorisation of experiential qualities in terms of measurability

Furthermore, the main argument for measurability is that the EQs of interest are well defined and documented in the literature. Two participants, however, could not name any certainly measurable EQ because they considered that qualitative data were better for understanding feelings and that experiential concepts were in general fairly vague. In contrast, the key arguments for non-measurability are the epistemological assumption about the nature of certain experiences and lack of a unified agreement on what UX is. The five participants could not name any certainly non-measurable EQ. They, while assuming that everything can be measured, had the reservations for the validity, impact and completeness of UX measures. Specifically, P9 pointed out the issue of conflating meaningfulness with relevance:

> "I think anything can be measured in a meaningful way; it depends who the audience is… the issues with measurement … are well understood in the psychometric system whether you are really measuring what you think you are measuring. So, and, again you need to distinguish between meaningfulness and relevance… there are things that are irrelevant … but I don't think it's possible for things in this world to have no meaning… people are natural interpreters.

With regard to the question on how to measure EQ, the participants identified a range of known HCI methods, which can be categorized into three major types: *overt behaviour* (e.g., time-on-task, number of trials to goal); *self-reporting* (e.g. diary, interview, scale); and *psycho-physiological* (e.g. eye-tracking, heart rate). Obstacles for implementing measurement were also mentioned, including various forms of validity, individual differences, cultural factors, confidence in interpreting non-verbal behaviour, translating abstract concepts into concrete design property, and consistency of observed behaviour

Anecdotal Descriptions on the Interplay between Evaluation and Development

In responding to the interview questions, some participants described intriguing cases that can well illustrate the challenges of enhancing the interplay between UX evaluation and system development. Subsequently we highlight the challenges and related anecdotes, which are grouped as theoretical (Q10), methodological (Q11) and practical issues (Q12).

Theoretical issues

- *Problem of measuring UX in a holistic way and breaking down into components seems not an ideal solution.*

  > **P3**: When we go through the issues with uses, we observe the whole expression, their comments on certain issues. If we have a lot of things to study, it is more difficult to run this kind of a holistic study; in a lab test where we only study some specific items. In an evaluation session when we study several issues, we can show users some of them and then the whole one. Holistic approach is the way to go, but measures about some specific details help as well.

  > **P4**: I'd say UX is holistic in nature, it is difficult to break it down into very small pieces. From the traditional scientific perspective, the way to measure something, to break it down and separate different factors … The value of the measurement gets lower if you break it down to small pieces... My colleague studied 3D video. She was able to measure objectively some aspects in lab by breaking things down, but when she went to realistic context for certain kinds of arrangement, the results are really different…. Your experience may change dramatically.

- *Memorized experiences prone to fading and fabrication*

  > **P5**: the actual intensity of the moment fades very fast… So it is interesting to see how to recall and how we change the memory of the experience. When we ask people whether they like something or not it depends on the moment you are asking. iPhone, there is so much positive information of that product out there that even if you did not like it, your environment is so positive about it that you are positive as well. It is the same as with reconstructing the memories. … Most people as well as I myself are sure I have memories where I cannot make a difference between the reconstructed and actual memory.

- *UX measures are highly sensitive to timing and nature of tasks*

P2: When to measure depends the duration and complexity of the task. For a small task, we can let people complete it and take measures at the end. For the longer one may need to be interrupted…. I am thinking a lot how much I am manipulating everything when I am organizing a workshop with some tasks how everything would be different if the tasks would be different….

P8: Different measures in different phases of the usethey complement each other if we need long-term evaluation. Sometimes you can get details out of there supporting design. They are more for prioritising the essential issues.… You don't have exact measures for evaluating emotions at the moment. Very momentary info can be useful, but you also need other measures. Even though you can capture all the momentary emotional measures, you don't know how the user interprets the emotion. The interpretation of the person is very important
a negative experience can be interpreted as a positive experience later on.

Methodological Issues

- *Different preferences for qualitative and quantitative data by design- and engineering-oriented stakeholders*

P7: … we are not fond of measures … we have smart design work, something we have emphasized more on qualitative and inspirational aspect of UX. We have something to do with design perspective; kind of measurement only gives basic constraints and do not give directions. It depends where you apply the methods; how they should be interpreted and position the methods. Measures are good background knowledge but we have more unpredictable, qualitative data.

P8: Qualitative data could cover everything, but then how to convince the engineers, that's why we need numbers. Also for research purpose, it could be interesting to find the relationships between factors. I have to measure somehow to find out which is more influential, hedonic or pragmatic quality, on customer loyalty… quantitative data are more convincing, but developers need qualitative data as well because they want to understand the reason for frustration… the developers like videos because they can describe very lively the situation. They can also believe textual descriptions. … It is important to measure both immediate experience and memorable experience. Practitioners are very thrilled by the idea that you can do it afterwards because it is so easy. So the companies are very interested in long-term UX or this kind of retrospective evaluation, they don't mind that, because they are convinced that memories are very important because they are telling stories to other customers; they are loyal to the companies based on the memories. Only the reviewers are criticising the validity of retrospective methods. Practitioners are very interested in it and like the idea.

P10: You have to interpret psycho-physiological data and map these data to one of these experiential concepts and it is very hard to know whether you get it right. You can have a high heart rate because you really love it or you hate it. So may be it also depends on how many categories you have; the more categories you have, the more difficult to find a good mapping.

P11: To see the impact of the goal of the system, how people perceive it. I think that's fine. For the purpose of design, quantitative measures do not make sense. It is a wrong method for the purpose of design.

- *Resource-demanding evaluation with a large number of heterogeneous users*

P4: Our perspective is very design-oriented. My experience in measuring UX in design process is not so much. It is so easy and fast to make the participants fill out AttrakDiff, it really would not make sense *not* to do it. How we analyse the results and get out of it, that's still to be seen. We don't have so many participants that we could see what the different ways of using those results are. Like a backup, we get a general understanding of the situation to compare for making the second prototype, what things to change. When we have the second prototype and we use the same measurement, we can see where the design is going. As measurement depending so heavily on individual participants, it is difficult to make conclusion about the measurements… it is hard to say why there is a difference in the results because of different social groups.

- *Need of sophisticated prototypes for eliciting authentic user experiences*

P7: Difficult, especially housing business … we cannot build only one prototype and then ask people experience it, get feedback and then do it… we need good examples, media we can use to produce our tools, social media, TV, etc to show what kind of solution we might have.. the storytelling method like movie; I'd like to see sophisticated level like what would be done with professional actors, directors, writers, like real life, feeling like real life with different natural mistakes.

Practical Issues

- *Lack of knowledge in exploiting feedback on UX for future system development*

P5: Most people in industry, whether they have backgrounds in economics, engineers or marketing, for them handling qualitative information is very difficult and they even don't know how to use that or they would need that…. We've been criticising the UX evaluation, not about how we measure UX, but how we use the information it in industry. … But there is so much information that people don't bother to read or follow them. We need to make things simple and easy so that people don't have backgrounds they can understand. In fact, the majority of usability people, at least in Finland, have engineering or computer science background but have little about psychology. There are a lot of things natural for psychologists or sociologists during the study handling control vs. experiment. They don't necessarily come to think of; there are experts in company talking about human beings, but they have certain views. It is challenging. This area of UX has the good side of interdisciplinary as well as the negative ones.

P4: Quite often field experiments lead to straightforward results that can be exploited in their design work right away. One project quite a while ago… We had purely lab experiments. We were doing lab test applying Fitt's law with different input devices, we were creating some constants that could be used for evaluating early stages of design to see if input device Design A is better than Design B. The partners were really excited about the results. They were well done, theoretically and practically validated and applicable… Industrial people were quite lost when we were not there. They needed our guidance. Unfortunately we had no choice. We had good results, but no real exploitation of the results since the customer did not know what to do with the results.

- *Lack of standard UX metrics renders redesign decisions prone to personal bias*es

  P5: People make decisions based on their personal beliefs. They just pick from the UX measures the ones that support their existing belief, and ignore the other results that don't support. … They don't even realize it themselves that they are manipulating the results. … People don't know how to use information on human beings. … we had noticed that the same icon did not work for various kinds of notification… We got feedback the people were annoyed… there was a very strong personality in the design lead who said that he did not want the design changes because they look ugly… It is problematic that UX have no commonly agreed definition or no commonly agreed metrics. It allows people to use this kind of argumentation that "I believe that it is better UX". You don't need to justify, it can be a personal opinion even though there are tons of user feedback.

- *Packaging UX measures for decision makers and speaking their language*

  P4: … social TV case we did Attrakdiff questionnaire and industry partner was very interested in that. They saw the potential in that when we had enough data, more convincing, more easily convince their superior of the organization to finance their projects, show the need for working on some aspects further; objective foundations.

  P5: It is not meaningless to measure moment-to-moment experience, but the question is how you use this information… But how to pack the thing and sell the thing to people making product or legislation decisions. In this area we should talk about how we use the information in this domain for the legislation and guiding the decision makers of different countries… Even when I think about from the industry perspective. Strategy management what they are most interested in is that what are the elements that make users buy next devices from the same company as well and what can reduce the number of helpdesk contacts. The first one is related to the future revenue of the company and the second one is related to the cost saving. It is mostly transfer it to money. It is the language that the management understands.

CONCLUDING REMARKS

UX, as an immature research area, is still haunted by the challenges of defining the scope of UX in general and operationalising experiential qualities in particular. Apart from addressing these basic issues, it is necessary for UX professionals to identify plausible means for compromising the difficulties of evaluating UX in a holistic manner with the limitations of adopting the reductionist approaches. Deeper understandings about the relationship between experience and memory and about the temporality of UX are also required. While the utility and necessity of employing both quantitative and qualitative methods is commonly recognized, the concomitant issue of providing appropriate education and training in UX needs to be explored. Specifically, UX researchers and practitioners should be equipped with knowledge and skills to know why certain UX measures are taken and how to use and interpret them in order to inform design and development decisions.

Insights into the issues of UX measures have been gained from the interviews. The study has raised more questions than it can answer. As the number of participants was relatively low with most of them originating from one country, namely, Finland, the views expressed might not be representative. Given this drawback, we have been motivated to expand the investigation on UX measurement with a larger scale survey of which results are documented elsewhere (under review). With a better understanding of the issues about UX measures, especially how they can be translated into new design requirements, insights into the interplay between UX evaluation and design can be gained.

ACKNOWLEDGEMENT

Many thanks should be given to Dr. Virpi Roto, Aalto University, Finland for her generous support in arranging the interviews when Effie Law had her short-term scientific mission in Helsinki funded by COST IC0904 TwinTide.

REFERENCES

1. Bargas-Avila, J.A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. In *Proc. CHI'11*
2. Bartholomew, D. J. (2006) (Ed). Measurement (Sage Benchmarks in Social Research Methods). Volume 1. Sage.
3. Boyatzis, R. E. (1998). Transforming qualitative information: Thematic analysis and code development. Sage.
4. Hand, D.J. (2004). Measurement theory and practice. Wiley-Blackwell.
5. Hassenzahl, M., & Monk, A. (2010). The influence of perceived usability from beauty. *Human-Computer Interaction, 25*(3), 235-260.
6. Law, E.L-C. (2011). The measurability and predictability of user experience. In Proc. of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing System (EICS 2011), Pisa, Italy, June 2011.
7. Law, E. L-C, Roto, V., Hassenzahl, M., Vermeeren, A., & Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach. In *Proc. CHI '09*, 719–728. ACM, 2009.
8. Löwgren, J. (2007). Fluency as an experiential quality in augmented spaces. *International Journal of Design, 1*, 1-10.
9. Mahlke, S., Lemke, I., & Thüring, M.(2007). The diversity of non-instrumental qualities in human-technology interaction. *MMI-Interaktiv*, Nr. 13, Aug 2007, ISSN 1439-7854.
10. Obrist, M., Law, E.L-C., Väänänen-Vainio-Mattila, K., Roto, V., Vermeeren, A., & Kuutti, K. (2011). UX research- which theoretical roots do we build on – if any. In *Extended Abstract CHI'11.*
11. Schaik, van P., Hassenzahl, M., & Ling, J. (2012). User experience from an inference perspective. *ACM Transaction on Human-Computer Interaction*.
12. Scherer, K. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*(4), 695-729.
13. Vermeeren, A. P.O.S. Law, E. L-C., Roto, V., Obrist, M., Hoonhout, J., Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. In *Proc NordiCHI 2010* (pp. 521-530).
14. Thomson, W. (1891). *Popular Lectures and Addresses, Vol. I.* (p.80). London: MacMillan.
15. Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human-technology interaction. International Journal of Psychology, 42(4), 253-264.

European Science Foundation provides and manages the scientific and technical secretariat for COST

COST is supported by the EU RTD Framework Programme