International Workshop on the Interplay between User Experience and Software Development (I-UxSED 2010)

NordiCHI 2010, 17th October 2010, Reykjavik, Iceland

Effie L-C Law

University of Leicester, UK/ ETH Zürich, Switzerland

Silvia Abrahão

Technical University of Valencia, Spain

Jan Stage

Aalborg University, Denmark

Towards the integration of transectorial IT design and evaluation

ICO904 TwinTide

Title: Proceedings of International Workshop on the Interplay between User Experience Evaluation and Software Development (I-UxSED 2010)

Editors: Effie L-C Law, Silvia Abrahão, and Jan Stage

ISSN: 1613-0073

The CEUR Workshop Proceedings Series
(http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/)

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

ACKNOWLEDGEMENTS

First of all, we are very grateful to the NordiCHI 2010 workshop chairs, Prof. Kasper Hornbæk and Prof. Klaus Marius Hansen, 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:

- Tiziana Catarci, Università degli Studi di Roma 'La Sapienza', Italy
- Asbjørn Følstad, SINTEF, Norway
- Morten Hertzum, Roskilde University, Denmark
- Rune Thaarup Hoegh, ETI, Denmark
- Kasper Hornbæk, University of Copenhagen, Denmark
- Nuno Jardim Nunes, University of Madeira, Portugal
- Jair Leite, Universidade Federal Fluminense, Niteroi, Brazil.
- Marianna Obrist, University of Salzburg, Austria.
- Virpi Roto, Nokia Research Centre, Finland.
- Jean Vanderdonckt, Université catolique de Louvain, Belgium
- Arnold P.O.S Vermeeren, TU Delft, the Netherlands

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 Software Development: Challenge and Outlook

Effie L-C Law
University of Leicester/ETH Zürich
LE1 7RH/CH-8092
Leicester, UK/Zürich, Switzerland
elaw@mcs.le.ac.uk/law@tik.ee.ethz.ch

Silvia Abrahão
Universidad Politécnica de Valencia
46022 Valencia
Spain
sabrahao@dsic.upv.es

Jan Stage
Aalborg University
DK-9220 Aalborg
Denmark
jans@cs.auc.dk

Theme and Goals

Following up the tradition established in the I-USED (International Workshop on the Interplay between Usability Evaluation and Software Development) series of workshops[1], this workshop is aimed at bringing together Human-Computer Interaction (HCI) and Software Engineering (SE) professionals and researchers interested in discussing recent trends and perspectives of the role of usability in software development. With respect to I-USED, I-UxSED (http://users.dsic.upv.es/workshops/i-uxsed10/) extends its scope to target the broader concept of User Experience (UX) in software development.

This change of focus was mainly motivated due to recent advances in mobile, ubiquitous, social, and tangible computing technologies that has moved HCI into practically all areas of human activity. This has led to a shift away from usability engineering to a much richer scope of user experience where user's feelings, motivations, and values are given as much, if not more, attention than ease of use, ease of learning and basic subjective satisfaction (i.e., the three traditional usability metrics). To accommodate the shift, evaluation approaches need to respond in a way that is sensitive to increasingly diverse use contexts, user goals and roles, and new interaction styles [8].

A range of emergent design and evaluation approaches such as experience-centred design (e.g. [6]), worth-centred design (e.g. [10]), and ethnography-informed design (e.g. [11]) have been developed. These new approaches deal with issues such as emotion, affect, aesthetics and longitudinal user-artefact relationships that entail augmentation of some maturing usability models and methods [17].

Among others, four challenges engendered by the new focus of UX are particularly relevant to software development: (i) definition of UX; (ii) modeling of UX; (iii) selection and application of UX evaluation methods; (iv) interplay between UX evaluation feedback and software development. These issues are closely related. Each of them entails quite some space to elaborate. Here we summarize the main arguments involved in the first three and discuss their impacts on the fourth one – the theme of this proposed workshop.

The concept of UX is commonly understood as *subjective*, *context-dependent* and *dynamic* [18], and these key attributes seem not conducive to measurability. In contrast, the 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* – suggests that UX can be measured in a way similar to the behavioural and attitudinal metrics of usability. Different attempts have been undertaken to demarcate or even dismiss the boundary between usability and user experience at the conceptual as well as operational level. A significant implication of this definitional issue is what can be considered as valid measures of UX, which enable professionals to benchmark competitive design artefacts and to select right design options.

Modelling users' experience - as a basis for producing design guidance - is especially important. First, measurement models are required to provide a sound basis for UX measures with desirable properties (e.g. reliability, validity, sensitivity). Second, structural models are needed for the purpose of understanding, predicting and reasoning about processes of UX with consequences for software design. Despite some visible progress (e.g. [12]), a number of issues pertaining to UX modelling remain to be resolved [19]. Furthermore, it is very important to develop practical guidelines for selecting evaluation methods and an associated set of measures to meet requirements specific to the context of interest. Currently, research efforts have been invested in collecting, consolidating and categorizing UX evaluation methods (e.g. [23]). 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.

Presumably, the aforementioned work pertinent to the three challenges (i.e. defining UX, modelling UX, and selecting UX methods) can contribute to the resolution of the fourth one (i.e. interplay between UX evaluation and system development), which, as far as we know, is only explored to a limited extent.

[1] The I-USED workshops were successfully held at HCSE 2008 and INTERACT 2009 ([1], [2])

We understand the relationship between UX and usability as the latter is subsumed by the former. While usability evaluation methods (UEMs) and metrics are relatively more mature, UX evaluation methods (UXEMs), which draw largely on UEMs [26], are still taking shape. It is conceivable that feeding outcomes of UX evaluation back to the software development cycle and instigating the required changes can even be more challenging than doing so for Usability Evaluation (UE). Several concerns are as follows:

- UX attributes are (much) more fuzzy and malleable, what kinds of diagnostic information and improvement suggestion can be drawn from evaluation feedback. For instance, user-based evaluation of fun - one of the critical qualities that have triggered the shift of attention from usability to UX ([5], [9]) – can involve subjective data with interviews or scales and objective data with psychophysiological measures [20]. A game can be perceived by the same person as a great fun on one day and a terrible boredom the following day, seemingly 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 to fix this experiential problem (cf. usability problem) and how can they know that their fix works?
- 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?
- Irrespective of whether formal or informal evaluation approaches are applied to traditional HCI phenomena like usability or emerging ones like UX, it is the *persuasiveness* of empirical evidence that is ultimately the test of its worth. Indeed, earlier research (e.g. [22]) 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?
- Software Engineering (SE) community has recognized that usability does not only affect the design of user interfaces but the software system development as a whole. In particular, efforts are focused on explaining the implications of usability for requirements gathering [16], software architecture design ([3], [4]), and the selection of software components [21]. Can such recognition and implications be taken for granted for UX, given that the evaluation methodologies and measures of UX could be very different (e.g. artistic performance)?

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

The goal of this workshop is to bring together researchers and practitioners from the HCI and SE fields to identify challenges and plausible resolutions to optimize the impact of UX evaluation feedback on software development. Presentations of new ideas on how to improve the interplay between HCI & SE to the design of *usable, pleasurable* and *desirable* software systems should be based on empirical studies. Within this focus, topics of discussion include, but are not limited to:

- *Which artifacts of software development are useful as the basis for UX evaluations?*
- *How do the specific artifacts obtained during software development influence the techniques that are relevant for the UX evaluation?*
- *In which forms are the results of UX evaluations supplied back into software development (including the UI design)?*
- *What are the characteristics of UX evaluation results that are needed in software development?*
- *Do existing UX evaluation methods deliver the results that are needed in user interface design?*
- *How can UX evaluation feedback be integrated more directly in user interface design?*
- *How can UX evaluation methods be applied in emerging techniques for user interface design?*
- *How can UX evaluation methods be integrated to novel approaches for software development (e.g., model-driven development, agile development)?*

Relevance to the Field

The main contribution of the workshop is the understanding of state-of-the-art about the interplay between UX evaluation feedback and system development and the identification of areas for improvement and further research. The HCI field includes a rich variety of techniques for UX evaluation and user interface design. However, there are very few methodological guidelines for the interplay between these key activities; and more important, there are few guidelines on how to properly integrate these two activities in a software development process.

Accepted Submissions

Based on the results of the systematic peer review process, nine submissions have been accepted for the workshop, representing a spectrum of views on the theme on interplay between user experience evaluation and software development. Here below we highlight the main arguments of each submission and our reflections on them.

Følstad succinctly puts forward a stimulating proposition: "*Complex models, underpinning complex measures, are likely to be valuable to the advancement of UX theory. However, in order to advance the interplay between UX and SD, simplified models and measures may be required*." A concomitant query can be: While simplicity tends to yield receptivity, could simplification have a similar property? Simple is different from simplified: If something is inherently simple, it is likely to be accepted. However, if something is inherently complex, simplifying it may lead to misrepresentation and thus confusion.

Jääskeläinen and Heikkinen have conducted a national survey to identify the differences between professionals and end-users in their understanding of UX definitions and attributes. A host of intertwined factors can contribute to the differences observed. Methodologically it is challenging to isolate the respective impacts of these factors.

Tim and Huang present inspiring proposals for formalizing trust requirements with notations such as extended UML with which software developers are likely familiar. The use of such a boundary object [24] to facilitate communications between designers (or evaluators) and developers seems promising. The authors' proposal of using physiological measures to triangulate cognitive metrics sounds exciting as well.

Jokela provocatively argues that there are two gaps instead of one: between UX and interaction design and between interaction design and software development. The argument can be boiled down to the basic issue of the very nature of UX. Indeed, some researchers and practitioners tend to synonymise UX with interaction design. Besides, Jokela's JFunnel user experience life-cycle model seems built upon usability. Debates on how to demarcate UX from usability and the other related concepts are ongoing [24]

Karahasanović and Obrist extend the issue of downstream utility of usability evaluation methods (UEMs) to user experience evaluation methods (UXEMs). The applications of interest are social media. The authors scope seven UXEMs against eight UX factors. Six recommendations are derived from the scoping exercises. Empirical validations of these recommendations are called for.

Wolkerstofer and his colleagues aim to bridge the cultural gap between the two communities: HCI and eXtreme Programming (XP). The observation that XP developers are resistant to persona coincides with similar findings in some other non-XP projects. It is intriguing to explore which factors contribute to such resistance.

Müller, Law and Strohmeier address the issue of persuasiveness – a significant notion of downstream utility, which is related to the work of Karahasanović and Obrist in this volume. Müller et al. map the constructs of the two traditional models in the domain of Information Systems to UX attributes and then compare whether usability-oriented ones are more persuasive than UX-based one in enhancing developers' problem-fixing tendency. The study serves as a precursor to a more ambitious investigation of the actual fixing behaviours of developers.

Alsos studies the notion of indirect user experience in the context of hospital with physician being primary users and patients indirect ones. User experience of the former can have influence that of the latter, or vice versa. The work may shed light onto the issue of co-experience or vicarious experience, which entail further conceptual and practical analysis.

Vold and Wasson investigate the ever prevailing phenomenon of participatory culture of learning, thanks to the advent of social software applications. The authors put emphasis on the role of fun in terms of playful feedback in enhancing learners' user experience in an online community. Their work, like Alsos', can contribute to the deeper understanding of social experience.

In summary, the nine workshop papers address some basic as well as applied research questions in the domain of User Experience, which is still being defined and scoped. With the notion of UX being somewhat fluid, it is deemed especially challenging to analyse and engineer the effect of UX evaluation feedback on software development.

We would like to say a few words about the picture on the cover of the workshop proceedings. It has been generated by feeding the main bodies of the nine papers (i.e. without abstract, the other front matters or references) into a software application TagCrowd (http://tagcrowd.com/). It visualizes individual words extracted from the submitted text with different shapes and shades of blue according to their relative frequencies. The top 50 words thus identified have further been fed into another similar application Wordle (http://www.wordle.net/), which beautifies the word cloud. Not surprisingly, the words *UX, Design, Experience, Users, Evaluation*, and *Feedback* are salient ones.

References

[1] 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

[2] 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

[3] Bass, L., & John, B. (2003). Linking Usability to Software Architecture Patterns through General Scenarios. *Journal of Systems and Software, 66*(3), 187-197

[4] Bass, L., John, B., & J. Kates, (2001), "Achieving Usability through Software Architecture," Technical Report CMU/SEI-2001-TR-005, Software Eng. Inst., Carnegie Mellon Univ.

[5] Blythe, M. A., Overbeeke, K., Monk, A.F.., & Wright, P. C.(Eds.) (2004). *Funology: from usability to enjoyment*. Kluwer Academic Press.

[6] Blythe, M., Wright, P., McCarthy, M., Bertelsen, O.W. (2006). Theory and method for experience centered design. In *Extended Abstracts Proceedings of CHI'06*, Montréal, Québec, Canada, ACM Press.

[7] Bødker, S., Kensing, F., & Simonsen, J. (2004). *Participatory IT Design. Designing for Business and Workplace RealitiesI*

[8] Bødker, S., (2006). When second wave HCI meets third wave challenges. In *Proc. NordiCHI 2006*, (pp.1-8), 14-18, October 2006, Oslo, Norway.

[9] Carroll, J.M., & Thomas, J.M. (1998). Fun. *ACM SIGCHI Bulletin, 19*(3), 21-24.

[10] Cockton, G., 2008. Designing Worth: Connecting Preferred Means with Probable Ends. *Interaction, 15*(4), July+August issue, 54-57.

[11] Dourish, P., 2006. Implication for design. In Proc. of the Conference on Human Factors in Computing Systems (CHI 2006), Montréal, Québec, Canada, April 22-27, 2006, pp. 541-550.

[12] Hassenzahl, M., 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, *19*, 319-349.

[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] ISO 9241-110, 2006. Ergonomics of human-system interaction - Part 110: Dialogue principles, ISO Copyright Office, Geneva, Switzerland.

[15] Juristo, N., Moreno, A. M., Sanchez-Segura, M. (2007), "Guidelines for Eliciting Usability Functionalities", IEEE Transactions on Software Engineering, Vol. 33, No. 11, November 2007, pp. 744-758.

[16] Juristo, N., Windl, H., & Constantine, L. (2001), "Introducing usability", IEEE Software, 20-21

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

[18] Law, E. L-C., Roto, V., Hassenzahl, M., Vermeeren, A., & Kort, J. (2009). Understanding, scoping and defining user experience. In *Proc. CHI 2009*, 4-9 April 2009, Boston, USA

[19] Law, E.L-C., & Schaik van, P. (2010). Modelling UX – an agenda for research and practice. *Interacting with Computers*, *22*(5), 312-323.

[20] Nacke, L., Lindley, C.A., Stellmach, S. (2008). Log Who's Playing: Psychophysiological Game Analysis Made Easy through Event Logging. In *Proceedings of Fun and Games 2008 (LNCS 5294)*, (pp. 150-157), Eindhoven, the Netherlands.

[21] Perry, D., & Wolf, A. (1992) "Foundations for the Study of Software Architecture," ACM Software Eng. Notes, vol. 17, no. 4, pp. 40-52, Oct. 1992.

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

[23] Roto V., Obrist M., Väänänen-Vainio-Mattila K., 2009. User experience evaluation methods in academic and industrial contexts. In *Proceedings of Workshop UXEM'09*, CHI'09, Boston, USA.

[24] Roto, V., Law, E. L-C., Vermeeren, A., & Hoonholt, J. (2010). Demarcating User Experience. Dagstuhl Seminar. Accessible at: http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=10373

[25] Star S. L., & Griesemer J.R. (1989). "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39". *Social Studies of Science* **19** (4): 387–420.

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

The Relevance of UX Models and Measures

Asbjørn Følstad
SINTEF
Po. Box 124, Blindern
0314 Oslo
+47 22067515
asf@sintef.no

ABSTRACT
Two approaches to research on UX models and measures are discussed on basis of experiences from the field of usability research and an ongoing case of user involvement in software development (SD) by way of social media. It is suggested that simple measures and ad-hoc models, rather than complex models and measures, may be beneficial to the relevance of UX research for SD practice.

Categories and Subject Descriptors
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Keywords
User experience, software development, model, measure

INTRODUCTION
A key objective of research on user experience (UX) models and measures should be to influence the practice of software development (SD). However, to reach this objective, UX research need to address research relevance as well as rigor.

A useful discussion of rigor and relevance in research was made by Lee [1] within the context of information systems (IS). He argued that knowledge produced by IS research emulating the rigorous natural sciences does not by necessity imply relevance. Rather, for the research to be relevant it needs to generate *"knowledge about how to intervene in the world and change it in order to satisfy real world needs"* (p. 29).

In this paper, I will discuss two distinct approaches to research on UX models and measures, and their implications for research rigor and relevance. The first approach, *complex models and measures*, is what I see as a likely trend given that relevance is not prioritized in UX research. The second approach, *simple measures and ad-hoc models*, is suggested as an alternative.

The first approach is paralleled in previous work of the related field of usability research. In order to learn from the experiences made in this more mature field, my argument will be supported by reference to research on usability models and methods.

The second approach represent an intermediary position to what has been referred to as *phenomenological/pragmatist* vs. *inspired by experimental psychology* [2]. This approach is exemplified by a case from an ongoing research project.

The intended contribution of the paper is to serve as a starting point for discussions on the relevance of UX research.

APPROACH 1: COMPLEX MODELS AND MEASURES
A recurring theme of UX research discussions is the components of UX [2]. Which model components are needed for measurement and systematic UX improvement? Given the comprehensiveness of the UX concept [3] the set of model components is likely to be voluminous, indicating that future UX models well may be complex. Suggested components include, for instance, motivation, trust, aversion, hedonics, and fun [2].

This trend towards complexity mirrors parts of the usability research during the 80'es and 90'es. A large number of usability components were suggested, including Nielsen's efficiency, satisfaction, learnability, memorability, and error rate [4]. A similar complexity is found in measures such as SUMI [5]. ISO 9241-11 [6] decomposes usability in effectiveness, efficiency and satisfaction.

The complexity of usability models and measures is beneficial for the rigor of usability research. However, complexity may be detrimental to the relevance for usability models in SD. For SD practitioners, complex generic models may be impractical. The generic model may include several components that seem irrelevant to the development project at hand. Also, seemingly important aspects may not be handled by the model. One may speculate that the prevalence of homegrown usability measures at the expense of standardized measures [7] may be a consequence of the complexity of the underlying usability models and their mismatch with the SD practitioner's understanding of which aspects of usability that are relevant.

Sauro and Kindlund [8], a key critic of the relevance of usability research, argued that to increase the practical impact of usability data *"usability metrics need to be easier to use"* (p. 401). To this end they suggested a single summative usability metric.

The position of Sauro and Kindlund is controversial [9] though it recently has been underpinned by substantial empirical evidence from industry projects [10]. Controversial or not, an important lesson may be learnt here for research on UX models and metrics: Complex models, underpinning complex measures, are likely to be valuable to the advancement of UX theory. However, in order to advance the interplay between UX and SD, simplified models and measures may be required.

APPROACH 2: SIMPLE MEASURES AND AD-HOC MODELS

An alternative approach to UX models and measures may be pursued through simple measures and ad-hoc models.

A simple UX measure is a single rating scale, common in social software. Consider for instance the book ratings of Amazon (1-5 stars) or video ratings at YouTube (thumbs up / down). Such scales are typically shunned for the scientific measurement of experiences or attitudes, due to reliability issues. Similarly, Sauro and Lewis argue for composite rather than single item measures [9]. Even so, such measures seem to serve their purpose as practical social navigation tools. Also, research within the field of marketing indicates that single-item measures may hold similar predictive validity as multi-item measures for concrete constructs such as *ad liking* and *brand attitude* [11].

By ad-hoc UX models I mean models developed in response to a given concept, prototype or running system. Instead of utilizing a general UX model as basis for a generic UX measure, relevant UX components may be established on basis of users' responses.

As an exemplification of simple measures and ad-hoc models, I will briefly present an ongoing development case addressing new functionality for mobile phone e-mail clients. Design typically involves the *"simultaneous investigation of multiple alternatives by the same designer or team"* [12, p. 1243]. In the present case 22 ideas were generated and then refined as six early concepts; all across a working period of 60 hours. Following this, the concepts were made available for user feedback.

User feedback was collected through a social software application for sharing audio-visual content, modified for the purposes of design feedback. The participants were 212 regular users of e-mail-clients on mobile phones (use several times a week or more). They were presented for six concepts in sequence. For each concept they were asked to make a rating (1-5 stars) and one or more comments. The comments were made in response to open questions: *How would you use the suggested function? How may the suggested function be improved?*

As part of a research design not to be detailed here, half of the participants were allowed to see the other participants' comments prior to making their own and half of them were not. All participants were allowed to comment on each other's comments. None were allowed to see the others' ratings prior to making their own.

The user feedback provided the following key information:

- **Differentiation between the concepts:** Three of the concepts were rated low, three rated high. The ratings corresponded closely to the frequency of positive vs. negative comments for each concept.
- **Establishment of concept specific issues that may serve as basis for an ad-hoc UX model:** For one of the concepts, detailing functionality for 'Postponed sending of messages', relevant issues were mainly targeting utility. For another, 'Reading aid for long e-mails on small screens', relevant issues were lack of comfort and utility. For a third, 'Context-dependent e-mail receipt', relevant issues were privacy and utility.

The case thus illustrates that simple measures and ad-hoc models may serve as basis for choosing which design alternative to pursue, and enable the establishment of ad-hoc models that may be used to control the development process. In the present case, a UX model for the first concept would need to include only utility in addition to usability, whereas the UX models for the second and third concepts would also need to include comfort and privacy/trust respectively.

CONCLUSION

By this paper, I hope to contribute to a discussion on how research on UX models and measurements should be approached in order to obtain relevance. I hold that complex models and methods will indeed be beneficial for the establishment of UX theory. However, research relevance may require a different approach. Possibly, such an approach may be *simple measures and ad-hoc models*.

ACKNOWLEDGMENTS

This paper was written as part of the research project RECORD (www.recordproject.org), supported by the VERDIKT programme of the Norwegian Research Council.

REFERENCES

[1] Lee, A. S. 1999. Rigor and relevance in MIS research: Beyond the approach of positivism alone. *MIS Q.* 23, 1, 29-33.

[2] Law, E. L.-C., Vermeeren, A., Hassenzahl, M. and Blythe, M. 2007. Towards a UX Manifesto. Proceedings of Towards a UX Manifesto. COST294-MAUSE affiliated workshop.

[3] ISO. 2010. *ISO 9241-210: Human-centred design of interactive systems.* Geneva: International standards organization.

[4] Nielsen, J. 1993. *Usability engineering*. Academic Press, San Diego, CA.

[5] Kirakowski, J. *SUMI Background Reading: The Use of Questionnaire Methods for Usability Assessment.* Available at: http://sumi.ucc.ie/sumipapp.html (downloaded July 6, 2010)

[6] ISO. 1997. *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs). Part 11 – guidelines for specifying and measuring usability*. Geneva: International standards organization.

[7] Hornbæk, K. 2006. Current Practice in Measuring Usability: Challenges to Usability Studies and Research. International Journal of Human-Computer Studies, 64, 79-102.

[8] Sauro, J., Kindlund, E. 2005. A method to standardize usability metrics into a single score. *Proceedings of CHI '05*. ACM, New York, NY, 401-409.

[9] Hornbæk , K., & Law, E. (2007). Meta-analysis of correlations among usability measures. *Proceedings of CHI'07*. ACM, New York. NY, 617-626.

[10] Sauro, J., Lewis, J. R. 2009. Correlations among Prototypical Usability Metrics: Evidence for the Construct of Usability. *Proceedings of CHI'09*. ACM, New York, NY, 1609-1618

[11] Bergkvist, L., and Rossiter, J.R. 2007. The predictive validity of multi-item versus single-item measurements of the same constructs. *Journal of marketing research*, 44, 175-184.

[12] Tohidi, M., Buxton, W., Baecker, R., and Sellen, A. 2006. Getting the right design and the design right. Proceedings of CHI'06. ACM, New York, NY, 1243-1252.

Divergence of User eXperience: Professionals vs. End Users

Anssi Jääskeläinen
Lappeenranta University of Technology
P.O.Box 20, 53851 Lappeenranta
Finland
jaaskela@lut.fi

Kari Heikkinen
Lappeenranta University of Technology
P.O.Box 20, 53851 Lappeenranta
Finland
kmjh@lut.fi

ABSTRACT
User eXperience (UX) is a well known term, but there is a divergence of understanding its meaning between UX professionals and end users. Even among UX professionals definitions and attributes vary from source to source. Therefore it is not surprising that the importance of the different UX attributes between UX professionals and end users also vary greatly. The differences are even bigger, if the personal characteristics are taken into account. This might lead to a situation where the UX designer does not know what an end user of the software would like to have, also vice versa; an end user does not necessarily appreciate what the UX designer has done. This problematic situation forms the basis of the currently ongoing study and the preliminary results from the conducted UX surveys are presented here.

Categories and Subject Descriptors
H.1.2 [**Human factors**]

General Terms
Measurement, Design

Keywords
User experience, survey, professionals, end users, divergence, personal characteristics

INTRODUCTION

Despite user experience (UX) is an important aspect in virtually every modern day business activity and its importance has been acknowledged by scholars around the globe, general consensus about the meaning and understanding of it, is still missing [7], [4]. This is considered as a problematic situation, thus when a word means almost anything or everything, it actually means nothing [1]. How a UX designer, which is assumed to be a professional, can accurately foretell what an end user from a heterogeneous group of all possibilities would like to get, see, hear, feel or smell? This question is justified thus even the consensus about the UX definition is missing among UX professionals. The study behind this paper contributes to this issue by comparing the viewpoints from UX professionals against the viewpoints from end users by taking account the personal characteristics of end users to demonstrate the divergence of understanding different UX related aspects.

SURVEYING USER EXPERIENCE

Results presented in this paper are based on a two phase survey study. First part was conducted with 20 UX professionals, which were chosen by browsing IEEE and ACM digital libraries as well as the Google Scholar with a keyword ‘user experience’. The second part; end user survey was sent to about 15000 university students. Accurate amount of students who received the survey cannot be given since emailing was done by the universities IT support and the submission list was automatically generated from the students.

End user responses were received from the Lappeenranta University of Technology and the University of Oulu. Other Universities in Finland were also asked to participate but those did not participate due to different reasons. UX professionals and end users answered to a virtually same survey, but UX professionals had more open questions and a possibility to suggest new UX attributes for the survey. The professional survey form the qualitative part of the data and end user survey form the quantitative part. Later on the qualitative part will be enhanced with professional interviews. The results gained from the conducted surveys are compared against each other and also against the result gained by the authors of [7], called later on as a baseline survey.

Table 1 presents the common characteristics of the current respondents of our surveys. End users are from multiple different disciplines like IT, business administration, energy, environment, chemistry, medicine, economics and humanities.

	UX professionals	End users
Gender	9 females, 3 males	559 females, 801 males
Age	25-59	18-64
Computer ability	3-5	1-6

Table 1: Characteristics of the current respondents

Most of the end user respondents (72%) are from 20 to 29 years old and 78% of end users have 3-4 level computer ability. Multiple different nationalities were found from the responses, but 95% of all end user respondents are Finnish so the effect of nationality cannot be evaluated, this is also true for professional respondents, 50% are Finnish. Ability to use computers was divided from very poor (1) to a rock solid professional (6). Levels three and four, which contain the majority of the respondents, were defined as follow:

- 3 - I can use, install and update programs
- 4 - I can develop / maintain minor programs, web sites, etc.

Despite large amount of the respondents are 20-29 years old Finnish, this sampling should give somehow reliable picture of what the most potential end users of common applications appreciate. In the future, the effect of nationality and other age groups, like under aged, will be taken into account and those results will be compared with the present ones.

The conducted surveys consist of three different categories; personal characteristics, UX definitions and UX attributes. Characteristics are all commonly studied in the field of social sciences and psychology e.g. [2], [14], [9], [12]. Also the ability to use computer is taken into characteristics questions, since almost every modern day technology is somehow related to interaction between human and computer.

Second part of the survey repeats the UX definitions part from the baseline survey, but with a modest differences. One definition was dropped out since it was too closely related to a company and its' services and products (D1 in a baseline survey). Some definitions on the other hand were added to the survey, e.g. the new ISO 9241-210 definition [6], which is considered as an important step by the authors. Following definitions were included in our surveys:

- d1 - All aspects of the user's experience when interacting with the product, service, environment or facility [6].
- d2 - User experience is a special case of experience, where the person can use a system, with or without a purpose. Using means that the user not only senses the system, but also has the opportunity to manipulate or control the system [10].
- d3 - UX is a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational / social setting, meaningfulness of the activity, voluntariness of use, etc.) [5].
- d4 - The entire set of effects that is elicited by the interaction between a user and a product, including the degree to which all our senses are gratified (aesthetic experience), the meanings we attach to the product (experience of meaning), and the feelings and emotions that are elicited (emotional experience) [3].
- d5 - The quality of experience a person has when interacting with a specific design [13]
- d6 - The value derived from interaction(s) [or anticipated interaction(s)] with a product service and the supporting cast in the context of use [11].
- d7 - A momentary, primarily evaluative feeling (good-bad) while interacting with a product or service [4].
- d8 - All aspects of the end-user's interaction with the company. Its services and its products [8]

The respondents of our surveys were asked to select three most suitable definitions (compared to one in baseline survey); since in our opinion this three point arrangement will give more thorough information about the mutual order of the definitions. End users were also given a possibility to answer "I don't care" or "I don't know", which were included in order to reduce the amount of randomly chosen definitions.

Last part of our survey is using a 5-point evaluation scale, where 1 is the best and 5 is the worst. This same classification is also used in the presented tables and figures. The task was to evaluate 21 plain UX attributes in addition with a possibility to answer "I don't know" or "I don't understand". Respondents were asked to: "Evaluate the importance of the following UX attributes to you when using the software". Earlier in the survey they were asked to pick some software that they should think while answering to the questions. With this procedure respondents had a possibility to approach UX attributes from "What is best for me?" point of view. This had hopefully led to more reliable results than asking only about the importance of general UX attributes.

Clause formatted UX statements from the baseline survey were left out since those were considered to be too UX designer-oriented for end users to understand and answer properly, for instance, clause like "We cannot design UX, but we can design for UX". The conducted surveys can be found in the following links:

- https://www.webropol.com/P.aspx?id=356910&cid=56046243 (UX professional)
- https://www.webropol.com/P.aspx?id=404764&cid=117863430 (End user)

1436 end user responses were given to the survey, but 76 responses were dropped out from the final analysis. Most common reason for a rejection of an answer from these analyses was a row of empty values. Second reason for a rejection was a row of I don't know / I don't understand values. As an example this rate for aesthetics in all answers was 17,9% and 9,4% for user interface.

RESULTS

Results presented here do not go into details. Thorough results will be published after in depth analyzing.

UX Definitions

The most important finding from the definition part is that end users do not seem to appreciate definitions (over 25% I don't care answers). This finding is also reinforced by a quite equal amount of support for every listed definition among end users. UX professionals included in this survey on the other hand were fairly consent about the same two definitions, d3 and d4 (over 70%) as the respondents in the baseline survey (50%). It is also noteworthy that d5 remained without support in baseline survey as well as in our professional survey. Baseline survey also showed that there are great differences between UX professionals in academia and in industry [7]. This in our opinion is a downside since professionals might speak about the same thing but actually mean different things. Results in [7] in their own opinion indicate that higher expertise level correlates to lower need for a standardized definition. They also state that UX professionals seem to think that definitions are a communication tool for non-experts.

When the results from the both UX professional surveys are compared against the responses given by end users the difference is huge. This indicates that UX professionals and end users see things differently. The same phenomenon is even enhanced if all three selected definitions are taken into account. It was argued that asking end users about UX definitions is not the right way, which I do agree if the intention is to find out how end users understand UX. But in this case the intention was to compare viewpoints from UX professionals and end users so it was mandatory to ask the same questions from the both groups.

UX Attributes

If the averages of top rated attributes from UX professionals are compared against the same attribute averages by all end users, differences are notable as Table 2 presents.

	Professionals avg.	Diff %	End users avg.
Interaction	1,33	23,75%	2,28
Motivation	1,58	21,25%	2,43
Ease of taking into use	1,33	17,00%	2,01
Usefulness	1,33	12,75%	1,84

Table 2: Difference of top attributes pro vs. end users

On the other hand if top rated attributes by end users are compared against the same attributes from professionals, differences are fairly small as can be seen from Table 3.

	Professionals avg.	Diff %	End users avg.
Stability	1,67	9,00%	1,31
Functionality	1,75	7,25%	1,46
Usability	1,58	3,75%	1,43
Reliability	1,58	3,00%	1,46

Table 3: Difference of top attributes end users vs. pro

Results show that end users do not necessarily agree with UX professionals in all attributes but professionals seem to agree with end users. Generally it seems that UX professionals consider the whole picture which includes environmental and emotional aspects as well, while an end user shows more interest towards something concrete like the actual device or software and its properties.

In the baseline survey [7] authors discovered the un-significance of the background education. We speculate that this result can be explained by the fact that all respondents in their survey were more or less related to the UX field. Instead of repeating this background education step, we studied the effect of the personal characteristics to the answers.

First if the average of all attributes among all respondents was compared, difference is only 0,24 (professionals 2,15, end users 2,39), so from that viewpoint divergence is not an issue. When we moved on to more detailed results, differences were found.

End user results between males and females are close to each other when the average of all 21 attributes is considered (2,38 vs. 2,40). If same comparison is done to UX professionals, results are (males 2,32 vs. females 1,98), but only three male respondents are included so the result

might not be reliable. When attributes are considered individually, interesting differences can be found. End user male seems to be more oriented towards UI and interactive features than end user female, but in UX experts survey the results were opposite as Figure 1 shows. In figure (e) means end user and (p) means professional.

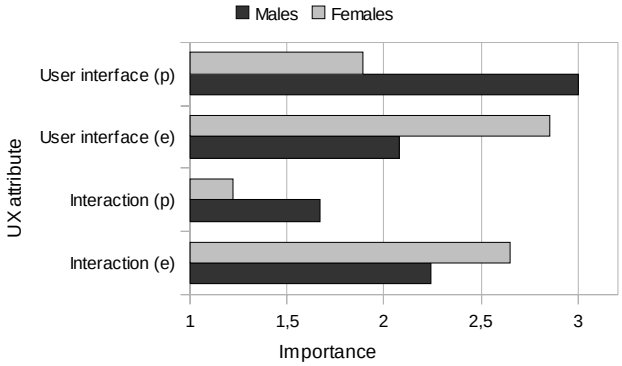

Figure 1: Males vs. females

When age is used as the divider, again the average of all end user attributes is almost the same for all age groups (2,33-2,42). Divergence among professionals is a bit larger (1,98-2,19). Interesting and linear differences are found when individual attributes are considered. Older respondents in both groups seem to appreciate general usability and user interface attributes more than younger respondents. Younger respondents on the other hand in both groups are a bit more positive about pleasure and coolness than older ones. Clearly age is an important affecting factor regardless of the experience in UX.

As the final part the effect of ability to user computers is studied. Respondents were asked to categorize themselves with a 6-step evaluation, where 1 was very poor and 6 was a rock solid professional. Average behaves the same way as before and similar linear effect as in the age was founded in some attributes. As an example, interaction and environment are presented in figure 1. In both cases the upper and longer line is the end user graph and the lower one is the UX professional graph.

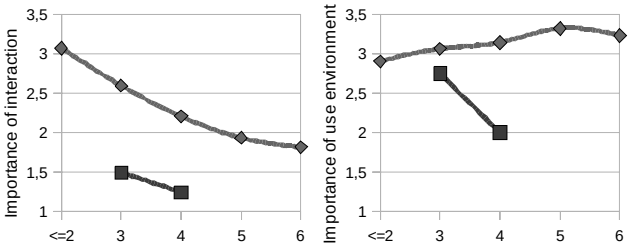

Figure 2: Effect of ability to use computer

In some attributes UX professionals and end users agree, but in some the opinions seems to be opposite.

CONCLUSIONS AND FUTURE WORK

The results of the conducted surveys clearly present the divergence of understanding UX matters between UX professionals and end users. When the personal characteristics of end users are taken into account, the divergence of results is even more amplified. Even among UX professionals there are big differences in answers when their personal characteristics are considered.

Need for clarification of different areas of UX exists and not only from the UX professional viewpoint but from the common end user viewpoint as well. This clarification makes it possible to design and implement better end user experience for devices, software, etc.

Information collected from the conducted surveys is used to create an UX database. This database in co-operation with an UX tool under development will offer a simple way for a software developer to focus on those UX attributes that the selected target group considers as important. UX database with the UX tool will be presented in a future paper.

REFERENCES

[1] Buxton, B. 2007. *Sketching User Experience: Getting the Design Right and the Right Design*., Morgan Kaufman.

[2] Dawson, P. Dobson, S. 2010. The influence of social pressure and nationality on individual decisions: Evidence from the behavior of referees. J. Economic Psychology, 31 (2010), 181-191.

[3] Desmet, P. Hekkert, P. Framework of product experience, J. Design 1, 1 2007.

[4] Hassenzahl, M. 2008. User Experience (UX): towards an experiential perspective on product quality. *Proceedings of IHM 2008*, 11-15.

[5] Hassenzahl, M. Tractinsky, N. User experience - a research agenda. J. Behavior & information Technology 25, 2 (2006)

[6] ISO 9241-210 (2010). Human-centred design for interactive system

[7] Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., and Kort, J. 2009. Understanding, Scoping and Defining User eXperience: A Survey Approach. *Proceedings of the CHI 2009* (Boston, Massachusetts, April 04-09, 2009), 719-728

[8] Nielsen Norman Group http://www.nngroup.com/about/userexperience.html (verified 27.9.2010)

[9] Rosen, M., 2007. Gender differences in reading performance on documents across countries. J. Reading and Writing 48, 4 (2007), 1-38.

[10] Roto, V. 2006. Web Browsing on Mobile Phones - Characteristics of User Experience. Doctoral Thesis.

[11] Sward, D. MacArthur, G. 2007. Making user experience a business strategy. Workshop on Towards a UX manifesto http://www.cost294.org

[12] Sykorova, D. 2009. Age identity. J. Sociologia 41, 2 (2009), 149-167.

[13] The User Experience Network http://www.uxnet.org/ (verified 27.9.2010)

[14] Yildirim, S. Narayanan, S. Potamianos, A. 2011. Detecting emotional state of a child in a conversational computer game. J. Computer Speech & Language, 25 (2011), 29-44.

.

Intangible Trust Requirements - How to Fill the Requirements Trust "Gap"?

Tim French
Department of Computer Science & Technology
University of Bedfordshire
Park Square Campus, Luton LU1 3JU, UK.
Tim.french@beds.ac.uk

Wei Huang
Department of Computer Science & Technology
University of Bedfordshire
Park Square Campus, Luton LU1 3JU, UK.
Wei.Huang@beds.ac.uk

ABSTRACT
Previous research efforts have been expended in terms of the capture and subsequent instantiation of "soft" trust requirements that relate to HCI usability concerns or in relation to "hard" tangible security requirements that primarily relate to security assurance and security protocols. Little direct focus has been paid to managing intangible trust related requirements *per se*. This 'gap' is perhaps most evident in the public B2C (Business to Consumer) E-Systems we all use on a daily basis. Some speculative suggestions are made as to how to fill the 'gap'. *Visual card sorting* is suggested as a suitable evaluative tool; whilst *deontic logic trust norms* and UML extended notation are the suggested (methodologically invariant) means by which software development teams can perhaps more fully capture hence visualize intangible trust requirements.

Author Keywords
Intangible trust requirements; visual card-sorts; deontic norms;

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
The extant trust related research literature is vast and highly diverse. A full literature review of intangible trust in B2C e-service contexts lies outside the scope of this exploratory paper. The interested reader is directed to both [1] and [2] for a comprehensive literature review. One of the difficulties is that the notion of trust is closely related to other concepts such as reliance, competence, trustworthiness and credibility. Deutsch [3] was one of the first modern writers to seek to build a formal model of trust. He defined trust in terms of an individual confronted with an ambiguous path. Further, the path may lead to either an event leading to a beneficial outcome (Va+) to that individual or to an event perceived as being harmful (Va-). This individual perceives that the occurrence of Va+ or Va- is dependent on the behaviour of another human agent. Finally, the strength of Va- is greater than the strength of Va+. Essentially, his view of trust is of a trust relationship in which events are linked to other events, each of which has beneficial or non-beneficial paths. For a trust relationship to occur, the harmful path is more significant than the beneficial path. Risk is an essential property of the environment within which a choice of paths occurs.

The notion that trust building between individuals takes place within information spaces that are both potentially risky to the participants and where incomplete information is available to the human actors has been widely accepted and developed by many subsequent researchers [4]. Within this information space the notion of expectation is central to many writers. For example, Gambetta [5] provides us with a rich and potentially computationally useful definition that encapsulates the notion of trust as expectation: ‘Trust (or symmetrically, distrust) is a particular level of subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action (or independently or his capacity ever to be able to monitor it) and in a context in which it affects his own action’. This idea of trust as an expectation (either rational or affective or a mixture of both) is a closely related concept to that of confidence levels. Confidence can be defined as a conscious or unconscious act or mental state involving the placing confidence in something or someone. The idea that this confidence level can be formalized, hence measured is developed by Marsh in his highly influential PhD thesis [6]. Human actors often invest their trust in a particular object of trust. This object may be another human agent or some artifact (such as a B2C Web-site). Indeed, within our intended scope of enquiry a company web-site acts as a central focal point within an information space (cyberspace) within which actors can choose either to follow "selfish self-interested" paths or act in the interests of others ("benevolent" paths). With the notion of an object of trust in mind, researchers have attempted to explore and develop the notion of trust as credibility, both within off-line and on-line contexts [2]. Kim *et al.*, [2] have variously

identified the sub-components of the credibility of an object of trust as comprising: honesty, expertise, predictability and reputation. Research appears to indicate that *credibility* is driven by the behavioural predictability of a trusted object, for example a web-site [2]. Indeed, many researchers, such as those of [7], define trust as comprising the dimensions of trustworthiness and expertise. These dimensions are closely related to the notion of credibility. Many have stressed the importance of trust building over time (i.e. the temporal dimension of trust). Thus the concept of trust as *reputation* has been developed by those seeking to quantify and measure high-level organisational trustworthiness in businesses and organizations, and most recently in Virtual Organisations [8]. Trust is not only dependant on our past experiences but also on an expectation of reliability and confidence in future events too. These aspects have been incorporated into various formal and informal models of trust building in relation to specific methodologies, such as agile [9-10]. An important aspect of trust building is the degree to which *affective* vs. *rational* components are involved. It is clear from the literature that the affective component has been relatively under researched in comparison to the rational cognitive dimension [1].

TRUST and UX

Hassenzahl and Tractinsky, [11] deem that UX is influenced by a user's internal state (e.g. predispositions, expectations, needs, motivation, affective state), as well as by the characteristics of the designed system (e.g. complexity, purpose) as well as by the environmental context within which the interaction occurs (e.g. organisational, social and cultural setting, and perceived meaningfulness of the activity). With respect to B2C e-shopping user contexts, reinforcing initial user trust building expectations is critical if businesses are to leverage full value from their B2C sites. B2C trust building via the HCI (Human Computer Interface) layer has been extensively studied and many of the trust building factors have been identified. Trust signs, seals, physical address details, rich media *et al.*, all serve to act as trust builders both as affective and rational trust drivers. A review of the relevant literature and a summary of trust determinants can be found in [12]. There is little direct methodological support to B2C system developers for intangible trust requirements capture. Rather, implicitly such systems are enhanced with respect to trust only much later via acceptance testing and usability testing, if at all. Thus, intangible trust requirements in sharp contrast with tangible security requirements are not typically captured early enough within the software development lifecycle.

One of us has successfully used visual card sorts in the context of an SME to improve the UX of their B2C web-site. The method has proven to be useful in revealing to both developer and end-users semi-tacit trust knowledge. Card sorts are not of course a new technique but their use in the context of validating intangible trust designs of for instance B2C e-banking is relatively new and may hold promise for the future [13]. The main advantage is that a cognitive "map" can be created from early (even paper based) web-site visual designs that reflects both the tacit and semi-tacit knowledge of site users. This "map" can aid in probing customer perceptions of UX, including affective and rational trust responses to various visual design features. Card-sorting can also be used to validate early paper-based prototype designs as well as later in the release cycle so as to probe "live" or pre-release user trust site perceptions.

The final site design is however, ultimately only the product of earlier application of methods, tools and requirements gathering activities. Indeed, we go on to argue below that there is an intangible requirements 'gap' within existing methodologies. That is, capture of intangible requirements is often implicit and does not always form an integral part of the normative team design process that ultimately leads to a given trusted or distrusted UX.

A METHODOLOGICAL TRUST 'GAP'?

The familiar software development lifecycle methods used in industry such as agile approaches, XP (Extreme Programming), RAD (Rapid Applications Development) and indeed the typical waterfall models all tackle the issue of requirements in different ways. Agile emphasises face-to-face interactions and iterative development within time-boxes, in which case it may be expected that some informal stakeholder trust expectations may exist during the lifecycle. As a type of agile software development, XP aims to improve responsiveness for requirement changes during the software lifecycle and it normally has multiple short development cycles instead of one long one in order to reduce the cost of changes. Unlike typical waterfall models, which can be treated as "plan-driven" or "predictive" so trust is relatively easier to be built up, agile methods, which have much in common with XP and RAD, are "adaptive" and can respond quickly to requirement changes. Thus trust plays an important role in this kind of "adaptive" method [9]. The following points seek to summarise and compare a number of typical lifecycle methods to see whether, when and how intangible trust issues are considered:

- *Waterfall*: Intangible trust requirements are typically embedded within feasibility study, and in a requirements catalogue as non-functional requirements explicitly agreed by client and developer;
- *Agile (e.g. XP)*: Trust is vital at every stage among developer team members. More specifically, trust is tested during the meetings and at the end of each and every iteration. Trust is built up due to its iterative nature and its primary focus upon interpersonal trust in the development process. Tested software is generated at the end of every iteration and this helps to build a sense of credibility (if the tested software is working as scheduled!);

- *RAD (Rapid Applications Development)*: Often implicit and embedded within evolving software artifact itself (all stages) Trust is most likely to emerge as an "issue" via acceptance testing and system walkthroughs once artifact is well refined. "Look and feel" of software engenders customers to engage through the use of branding, metaphor and narrative;
- *Test-driven development (TDD):* Not widely considered? Some signs of confidence / trust build up when all test cases "pass" (which only may mean that the code meets all the defined/explicit requirements. (Trust is perhaps not widely considered explicitly here partly because TDD is a relatively new technique.)

Ideally trust building aspects should be initiated at the very beginning among all system stakeholders (i.e. at the requirements stage) no matter which lifecycle method is used. Trust requirements are important to final users as well as to other stakeholders such as developers, managers, clients, and system sponsors. Indeed many consider trust amongst Agile teams, tools and techniques for example to be absolutely vital to help generate a credible "win-win" specification. Yet as many industry practitioners acknowledge this is rarely the case in practice due to deep seated cultural differences as between developers and their clients.

It would appear that there is a methodological 'gap' with respect to intangible trust requirements, particularly with respect to aspects of UX that encompass hedonic, emotional aspects - not merely trust as rational decision making and tangible security. There is a methodological trust “gap”, particularly at the requirements stage. None of the notations within the UML (Unified Modelling Language) for example directly support intangible trust requirements - aside from generic use-case and domain models. Rather, the main focus is upon tangible security and assurance aspects. Methods such as MEASUR [10] claim to add value to existing approaches via ontology charting by seeking to capture not only semantic entities and their relationships but also organisational contexts and culture, including normative methods of working, both formal and informal. This may perhaps serve to reveal implicit trust aspects within workgroups or indeed trust expectations concerning the presentation layer of the system. However, MEASUR is not in fact as yet widely used outside academia, despite many years of effort. This lack of adoption limits its potential impact and relevance to addressing the trust gap, despite recent efforts to formalize, align and integrate MEASUR with modern component based design principles [10].

HOW TO FILL THE TRUST "GAP"?

Dyadic trust between an e-service provider and consumer (trust as a set of expectations as to future behaviour, reliability, service quality *et al.,*) is typically influenced by both rational and affective drivers that in turn serve to influence technology acceptance levels. The well known and heavily cited TAM (Technology Acceptance Model) of a type proposed by [2] exemplifies this vision. It is implicit within such models, that trust is intimately related to risk; that is to say where there is no risk trust is not relevant. Rather, the higher the risk factors the more reliance needs to be placed by stakeholders on intangible trust requirements so as to mitigate perceived risk.

However, it would be a mistake to say that mere security assurance equates to trust *per se*; rather, intangible perceptions of trust create the necessary pre-conditions and set of constraints within which systems are procured, developed, and are ultimately released. Thus, trust is acts as a super-set (universe of discourse) within which tangible security assurance standards are seen to operate. However, it is important to note that *mere* security assurance itself (e.g. secure message protocols, cryptographic techniques) can of themselves *never fully meet* the intangible trust concerns of users as part of their UX. For one thing, wider intangible cultural trust norms differ greatly across and within societies and cultures [12] and thus are highly relevant to "shared meanings" across cross-trans/national software development teams working across borders or with culturally specific B2B partnerships. Although intangible trust perceptions/expectations between business partners or as between developers and clients mediate requirements negotiations from the earliest stage of the software development lifecycle, there are *two* fundamental problems that we seek to address:

a) Firstly, intangible trust requirements are often at least partially *implicit*, hence developers and clients may not themselves fully realise its potential impact upon eventual system acceptance until too late in the development of the system. When (due perhaps to disagreements or ambiguity or un-stated sets of trust expectations, i.e. norms and meta-norms) such issues may become more explicit, resulting in a mismatch as between a norm and an agreed set of functional requirements. Such matters cannot even be realised (hence alterations made) by system developers or their clients before actual release of the system unless intangible trust is fully and richly articulated (hence made fully visible) to all system stakeholders.

Candidate solution?: Deontic logic has previously been used to define norms and meta-norms in the context of enabling MONA (Portugese Acronym for a norm modeler for tailorable user-interfaces) [14]. It may be that in the future the definition of high-level trust specific norms and meta-norms can (since the natural language version of deontic logic is easily interpretable by clients) be potentially useful in framing intangible trust issues - thus potentially impacting on the design of a user's UX. A potential advantage of the use of deontic norms is that they are expressive enough to reflect well known cultural differences of the wider social world within which the system is seen to operate.

b) Secondly, there is a *notation gap* with respect to articulating intangible trust requirements. Within the rich and expressive notational vocabulary of the UML, there is no specific notational support for trust, other than as a natural text narrative to enrich the domain model. Various notations and formalisms such as state-charts have been adapted to reflect some aspects of intangible HCI trust requirements. However these extended or otherwise specially adapted charting methods are not widely used outside academia. Formal and mathematical notations claim to have been used for trust, yet they often only actually reflect tangible security paradigms. There has been some emergent work on the development of trust specific notations and methodologies such as *The Shared Meanings Design Framework* (SMDF) to capture trust requirements across stakeholder groups. Few if any of these notations are used outside academia.

Candidate solutions?: Perhaps suitable extensions to the well known UML notation can be provided to support the explicit articulation of intangible trust issues (for example the domain model). As yet, this potential has only been tentatively explored in relation to intangible trust [15], though recently approaches such as UMLTrust seek to offer support for intangible as well as tangible security aspects: trust policies, scenarios as well as trust certification [16]. One alternative path going forward is perhaps that one of the many trust specific notations to emerge out of academia will be adopted or otherwise influence industrial practitioners, such as the SULTAN (Simple Universal Logic-oriented Trust Analysis Notation) [17] and associated tool-kit previously developed at UCL.

In our earlier discussions it was apparent that there is no one methodology that is universally adopted. Rather, methods are selected by client-developer partnerships according to the "best fit" to whatever type of software system is proposed. For this reason we are very hesitant to supply any definitive answer to the trust "gap" across every method; but the above suggestions may perhaps at least prove useful as potential candidate solutions. In any event it is our contention that card-sorts have been shown to add value to the probing of intangible trust perceptions using B2C sites. Either in relation to early designs or in relation to pre- or post release UX intangible trust evaluation. So whilst various possibilities exist with respect to enriching existing methodological practice, UX trust perceptions can at least be probed empirically, once an artifact has emerged from the development team. It would of course be more desirable if as part of whatever methodology is used to develop artifacts, that intangible trust notational support could be agreed upon and more widely adopted. Thus far, whilst various tentative suggestions have been made by academia, industrial practice has tended only to support tangible security requirements at the expense of intangible trust concerns.

CONCLUSION

Intangible trust forms an important yet somewhat elusive part of both UX, and wider technology acceptance. Without trust building (explicit or implicit or both) systems will simply be not adopted or "work around's" will be employed. Despite the importance of intangible trust building, there appears to be a methodological and notation 'gap'. The framing of system design within explicit trust norms may prove to be useful since any methodology could be "front-ended" by a set of norms that are method and notational independent of any existing method. The alternative or complementary approach is to leverage an existing notation (e.g. the UML) for intangible trust or perhaps (even more speculatively) to seek to influence industry standards and methods such that they fully encompass intangible trust requirements. Others have tried to develop their own methods yet these are surely doomed to failure unless industrial developers and their clients feel that the trust gap is worth filling (adds real "bottom line" value?) Perhaps at present there is a certain cynicism that leads to rapid system release followed by numerous "patches" that seek to paper-over gaps in requirements. This is both the fault of clients (too demanding time-scales) and developers. But it is also because of a "gap" in the industry methods and notations currently deployed.

As wider notions of UX grow it is to be hoped that this "gap" will be filled as all stakeholders come to realize the importance but also acknowledge the intractability of trust; including the fact that we lack models of trust that take into account "obvious" cultural differences. Thus, there will be no quick "fix", rather the trust gap reflects deep seated cultural divide as between system stakeholders, organisational needs and current paradigms.

Many challenges remain not the least of which is: how to define intangible trust in the first place. Deciding how and what to "measure" becomes a central question - particularly perhaps with respect to the impact of cross-cultural trust norms. The extant literature in B2C e-trust has perhaps tended to place over-relied on methods such as questionnaires and under developed ways of probing user cognition such as card-sorts. Perhaps in the future, hybrid approaches that seek to triangulate as between physiological metrics and cognitive metrics by incorporating neuroscience may add value to evaluating trust as part of UX [1]. This may in turn lead to the definition of objective physiological trust metrics as well as subjective metrics in relation to the UX.

If industry is willing to embrace new "blue-sky" techniques and sees added value in funding studies in Usability Labs., as part of their requirements gathering /interface design validation studies then intangible trust requirements gathering activities could form a normative part of *every* software project that has end users irrespective of the actual choice of methodology. As yet though, too often failure to address intangible trust perceptions result in lack of adoption or expensive "fixes" and software re-releases.

To cite one well known UK Public Sector instance: the lengthy adoption of the NHS (National Health Service) GP-to-hospital "*Choose and Book*" specialist referral e-booking system has been frequently ascribed as being due to an initial failure to address stakeholder trust and mistrust issues at an early stage in the system's initial specification. This initial failure led to not only to an initial lack of adoption by GP's, but was the prime cause of numerous subsequent system upgrades over a lengthy eight year time scale [18].

REFERENCES

[1] French, T. Towards an E-Trust Framework: trust as a semiotic phenomena. PhD Thesis, (2009). Reading University, School of Systems Engineering, UK.

[2] Kim, D.J., Ferrin, D.L., Rao, H.R. A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems, 44* (2008) 544–564.

[3] Deutsch, M. Trust and suspicion. *Journal of Conflict Resolution, 2(3)* (1958), 265-279.

[4] Corritore, C., Krasher, B., & Wiedenbeck. On-line trust: concepts, evolving themes, a model. International *Journal of Human-Computer Studies, 58,* (2003) 737-758.

[5] Gambetta, D. (Ed.) *Trust: Making and Breaking Cooperative Relations.* (2000) Oxford: Basil Blackwell Publications.

[6] Marsh, S.P. Formalising Trust as a Computational Concept. PhD Thesis, (1994). Stirling University, UK.

[7] Fogg, B.J., Marshall, J., Laraki, O., Osipovitch, A., Varma, C., Fang, M., Paul, J.,Rangnekar, A., Shon, J., Swani, P. and Trienen, M. What Makes a Website Credible? : a Report on a Large Quantitative Study. *Proc. SIGCHI conference on Human factors in Computing Systems*, (2001) 61-68.

[8] French, T. Collaborative Virtual Organisation Trust Measurement: Leveraging Corporate Governance Metrics, *Procs. IEEE, i-Society*, (2010) 502-509.

[9] Hasnain, E, and Hall, T. Investigating the Role of Trust in Agile Methods using a Light Weight Systematic Literature Review. *Procs. 9th International Conference on Agile Processes in Software Engineering and Extreme Programming*. Limerick, Ireland, Jun 10-14, 2008. In: *Lecture Notes in Business Information Processing, Volume: 9*, (2008) 204-207.

[10] Poernomo, I., and Tsaramiris, G. Ontology Based UML2 Architecture Generation. *Procs. ICISO*, (2010) 314-321.

[11] Hassenzahl, M, Tractinsky N. (2006). User experience –a research agenda. *Behaviour & Information Technology*, 25 (2), March-April 2006, 91-97.

[12] Wang, Y.D., and Emurian, H.H. An Overview of online Trust: concepts, elements and implications. *Computers in Human Behaviour, 21,* (2005) 105-125.

[13] French, T., and Opatola, K. A Pilot Investigation of E-Banking trust perceptions amongst the Yoruba and Igbo of Nigeria. *Procs. IWIPS 2010,* (2010) 107-116.

[14] Neris, V.P., and Baranauskas. User Interface Design Informed by Affordances and Norms Concepts. *Procs. ICISCO,* (2010) 133-140.

[15] Oussena, S. and French, T. Integrating the Semiotic into UML via Enhancing and Cross-validating Use Case with an Enriched Domain Model. *International Journal of Sociotechnology and Knowledge Development (2010),1,3,*15-31.

[16] Uddin, M., and Zulkernine, M. UMLTRust: Towards Developing Trust-Aware Software, *Procs. of the 2008 Symposium on Applied Computing,831-836,* New-York, ACM.

[17] Grandison, T. Trust management for internet applications, PhD Thesis, (2002). University of London.

[18] *Referral Management: Lessons for Success*. King's Fund, 2010. ISBN 978-1-85717-600-1

Two Gaps instead of One. The Interplay between User Experience Engineering and Interaction Design

Timo Jokela
Helsinki University
timo.jokela@helsinki.fi

ABSTRACT
The paper identifies three main phases of designing user interfaces: user experience (UX) engineering, interaction design and software development. The paper focuses on the interplay between UX engineering and interaction design. It is argued that the gap between these two phases in inherently challenging. To overcome the gap, designers' knowledge on the results of UX activities and general UX guidance is critical. Incentives on good UX can be one means for good results.

Keywords
user experience, interaction design, software design

INTRODUCTION

The workshop theme is the challenge in making the interplay between user experience and software design successful.

This paper argues that there actually are three main phases, instead of two: user experience (UX) engineering[1], *interaction design*, and software design.

Interaction design produces the visual and logical design of the system from the user's point of view (user interface), i.e. the design of:

- individual user interface elements, such as dialogs, menus, tables, etc.
- the structure of the user interface, such as the identification, definition and design of the windows
- visual appearance
- also supporting material such as user manuals

Software development is used to implement interaction design. In other words, interaction design, as such, is not software development.

Thereby, this paper argues that main phases of designing interaction solutions are:

1. UX engineering: provides user driven data to guide interaction design
2. Interaction design: driven by UX engineering activities and UX design guidelines and standards, one creates interaction design solutions (ISO/IEC 2010)
3. Software design: interaction solutions are implemented through software.

Thereby, there are two separate challenges of interplay:

- the interplay between UX and interaction design
- the interplay between interaction design and software design

These different kinds of phases are illustrated in the 'JFunnel' lifecycle model: UX guided interaction design, Figure 1. In the figure, the three activities are marked with different colors:

- orange: UX activities
- light grey: interaction design
- dark grey: software implementation

In this paper, we focus on the interplay challenge between the two first sets of activities: UX engineering (orange) and interaction design (light grey).

UX ENGINEERING AND INTERACTION DESIGN

The distinction between UX activities and interaction design is made because the two kinds of activities have a fundamental difference:

- UX activities produce various kinds of user data, such as user descriptions, UX requirements, and results from evaluations (but does not produce interaction design solutions)
- Interaction design produces the actual user interaction design solutions: GUI designs, user documentation, and so on.

In other words, the role of UX activities is to provide user data to the interaction design activity; the role of the interaction design activity is to transform this data into design solutions that represent good UX. This is illustrated also with the arrows from other activities to the interaction design activity (Figure 1).

This separation is also useful because those who carry out UX activities (UX specialists) are often not the same people as those who produce design solutions (interaction designers). Moreover, the model illustrates that it is not

[1] The term UX engineering is analogous with the term 'usability engineering'. It covers not only UX evaluation but also understanding users, their goals, tasks and environments of use, and UX requirements determination.

enough to introduce UX activities - UX activities have no value unless the results are taken into account in the interaction design activity.

The UX Input for Interaction Design

The UX activities can be categorized into two groups:

- UX *requirement* activities that provide data based on analysis of users without any prototype or model of the system to be developed (activities 1…4)
- UX *evaluation* activities that provide feedback on a prototype or a model of the system under development (activities 6,7)

The UX requirement activities (1…4) provide data such as

- user group descriptions
- user goals, tasks and work descriptions
- UX targets: strategic targets are at business level; operational targets are at user level; UX measures

The UX evalutation activities (6,7) provide

- qualitative findings about which design solutions work and which solutions do not work from the viewpoint of UX
- results to which extent the UX requirements are met

In addition, the generic UX guidelines provide principles and design guidance at generic level. Examples of such guidelines are ones from the ISO 9241 series [1].

These inputs can be categorized as follows:

- UX activities 1-4 provide input for how to design systems to support users' tasks and work
- Generic guidelines provide input for how to design details of interactions
- UX activities 6 and 7 provide feedback to both of these levels of design: how the user interface support users' work; and how well the details are designed

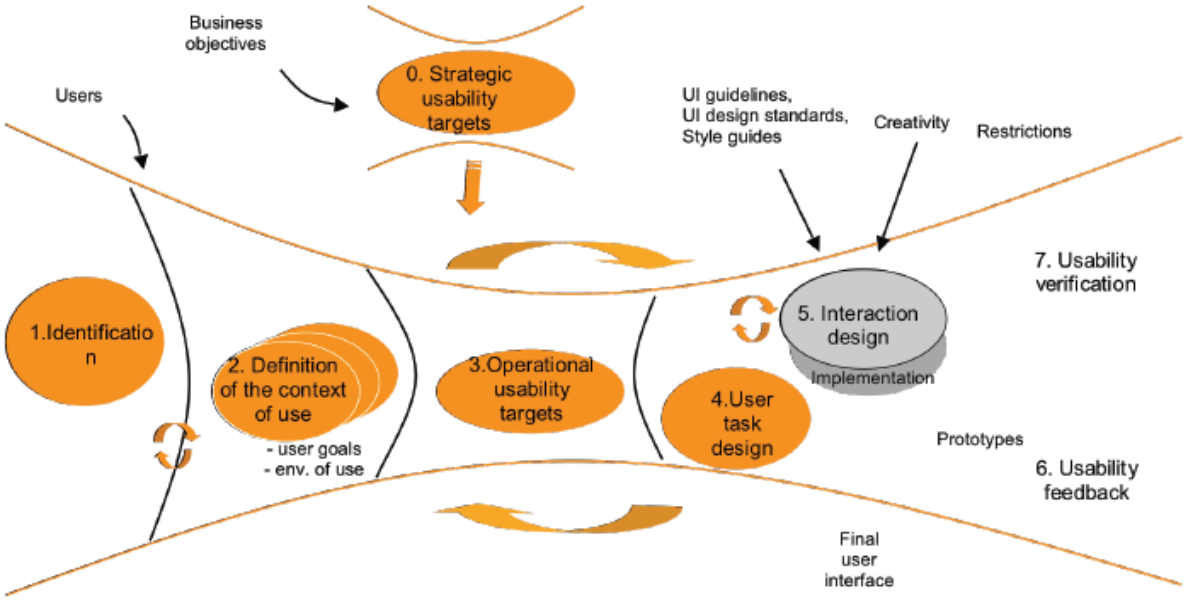

Figure 1. The JFunnel user experience life-cycle model [2]. The different types of activities are marked with different colors: UX engineering activities are orange, interaction design light grey, and software design (implementation) dark grey.

The Challenge between UX and Interaction Design

One can see that there is an obvious challenge in the interplay between UX activities 0…4 and interaction design: user definitions, user goals and task descriptions, and UX targets do not provide concrete guidance for interaction design. They provide only requirements, and it is the task of designers to transform this data into UX designs solutions.

For example, consider a UX requirement "The system should be learnable without user training". This kind of requirement represents a gap between UX and interaction design. It is a requirement for interaction design but does not provide any guidance for how to achieve it. It is a design challenge to produce design solutions that would meet the requirement.

UX evaluation activities 6, 7 provide clear results what works and what does not work. But the results do not as such tell what a better interaction design solution is.

UX guidelines provide more concrete guidance, but still they are more requirements than solutions. For example, "speak user's language" does not say what the

understandable terms of a specific system are. One concrete category of UX guidelines, however, exists: interaction design patterns (e.g. [5]); although focus currently more on usability than UX).

The Interplay between UX and Interaction Design

The conclusion is that the interplay between UX and interaction design is an inherent challenge. The extent to which interaction designs truly represent good UX depends very much on the interaction designers: to which extent they are willing and able to transform the input of the UX activities and generic UX guidelines into interaction design solutions.

Thereby designers' knowledge on factors affecting UX is extremely important: results from the different UX activities and contents of general UX guidelines. One solution that has worked in designing good usability [3] – give incentives to designers – may also have a positive impact for designing good UX.

DISCUSSION

This paper focuses on the gap between UX engineering and interaction design, and argues that there is a gap between these two activities. UX requirements may be tough, and it may be a challenge to produce design solutions that meet such requirements.

Would there be a direct gap between UX and software design?

One should understand that there does not exist a single 'right' interaction design solution: the UX requirements may be achieved with different kinds of design solutions. Thereby one should aim at producing such interaction design solutions that are easy to implement with software. In other words, one should aim for interaction design solutions that (1) meet the UX requirements and (2) are easy to implement with software. Probably this is achievable in most cases. But if not, then we have a gap between UX engineering and software design.

REFERENCES

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

[2] Jokela, T. (2008). A Two-Level Approach for Determining Measurable Usability Targets. 5th COST294-MAUSE Open Workshop on Valid Useful User Experience Measurement (VUUM), Reykjavik, Island.

[3] Jokela, T., J. Koivumaa, et al. (2006). "Quantitative Usability Requirements in the Development of the User Interface of a Mobile Phone. A Case Study." Personal and Ubiquitous Computing 10(6): 345-355.

[4] Microsoft (2010). Windows User Experience Interaction Guidelines.

[5] Tidwell, J. (2006). Designing Interfaces: Patterns for Effective Interaction Design, O'Reilly.

Investigating the Usefulness of Methods for Evaluating User Experience of Social Media Applications

Amela Karahasanović
SINTEF ICT and University of Oslo
Postbox 124, Blindern, 0314 Oslo, Norway
+47 48 10 88 95
amela@sintef.no

Marianna Obrist
ICT&S Center, University of Salzburg
Sigmund-Haffner-Gasse 18
5020 Salzburg, Austria
+43 662 8044 4814
marianna.obrist@sbg.ac.at

ABSTRACT
The usage and importance of social media or Web2.0 applications such as Youtube, Flickr, Facebook is rapidly increasing over the last years. They all build up on user communities, provide networking opportunities for their members, and are strongly related to audio-visual user-generated content (UGC). Providing the user a good experience is a central success factor for such applications. Apart from standard usability principles the much broader concept of user experience (UX), including aspects such as fun, enjoyment, emotion, sociability and other factors have become relevant in the design of interactive systems. However little has been known on the usefulness of different evaluation methods for UX in the context of social media applications. We need to understand what new requirements for applying UX evaluation methods on these applications evolve and how to choose which of the existing methods are suitable for capturing different aspects of UX. This paper reports results and lessons learned on the usefulness of seven UX evaluation methods that were applied for evaluating ten different applications supporting non-professional users in sharing and co-creating user-generated content. The results might be useful for practitioners and researchers developing social media applications when planning UX evaluation studies.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information Systems**]: Evaluation/ Methodology, H.5.2 [**User Interfaces**]: Evaluation/Methodology

General Terms
Human Factors

Keywords
User experience, Evaluation methods, Social media, Web 2.0, Communities, User generated content, Audio-Visual content

INTRODUCTION
For many applications, such as social media and social network sites, applications for sharing and co-creating audio-visual content, and, for instance, games, it is important that people enjoy using them. Thereby, providing people a good experience and evaluating their UX is becoming more and more essential [24].

Hassenzahl [9] states that a "good UX is the consequence of fulfilling the human needs for autonomy, competency, stimulation (self - oriented), relatedness, and popularity (others - oriented) through interacting with the product or service (i.e. hedonic quality)". Pragmatic quality, such as the usability of a system, is also contributing to a positive experience, but only through facilitating the pursuit of meaningful hedonic needs. The most important characteristics of UX are its normative nature (differentiating between a positive, desired experience and a negative, undesired experience that a user can have when interacting with an application) [10] as well as its dynamic nature [15].

Next to reach a common understanding on UX, there is still a lack of research on UX evaluation methods in general (see for instance an overview on UXEM in [30]) and on their usefulness in particular. Research papers and textbooks such as [4] provide surveys of different evaluation methods according to their appropriateness in different evaluation phases, their objectivity, reactivity and needed resources. However, little has been known about the usefulness of these methods for evaluating social media applications. This paper presents results and lessons learned of a case study we conducted to investigate the usefulness of both traditional and new methods for evaluating UX.

This work was carried out in the framework of the European research project CITIZEN MEDIA (http://www.ist-citizenmedia.org/) which aimed to develop social media applications supporting non-professional users in sharing and co-creating user-generated content (UGC). Several applications have been developed and evaluated at three testbeds, namely in Germany, Norway and Austria. The evaluation activities for all three testbeds were guided by a common evaluation framework consisting of pre-selected UX factors (e.g. fun/enjoyment, motivation, emotion, sociability, as well as usability) and a set of

evaluation methods considered as relevant for the context of the project (see [20]). Well known methods were combined or adapted in order to capture UX.

BACKGROUND

Over the years many usability evaluation methods have been proposed and evaluated. The research of Gray and Salzman [7], Hartson et al. [8], and Blandford et al. [1] established a basis for critical evaluation and selection of usability evaluation methods.

Blandford et al. [1] propose a comprehensive list of ten criteria for evaluating UEMs. *Reliability*, also called internal validity, is the extent to which different analyses of the same system, using the same UEM, yield the same insight. *External validity* is the ability to apply the findings in the real world context. *Thoroughnes*s is a proportion of real problems identified by a method. *Effectiveness* is the product of reliability and thoroughness. *Productivity* is the number of problems a UEM identifies. *The practicalities* criterion is concerned with what is needed to integrate a method within design practice. The *analyst activities* criterion describes what analysts do when applying a UEM. *Persuasive power* is concerned with the ability of an analyst working with a UEM to persuade developers to change the system. *Downstream utility* is usefulness of the findings in informing design. *Scope* describes what kind of problems a method is useful and not useful for finding. When comparing usability engineering methods, Holzinger [11] considers the following criteria: applicability in phase, required time, needed users, required evaluators, required equipment, required expertise, and intrusiveness.

Recently there has been growing interest in UX evaluation methods [30]. Several workshops have been organised to focus on the methods, techniques, and tools for evaluating UX such as CHI 2008 [28], CHI 2009 [19], INTERACT 2009 [25] and COST294-MAUSE workshops ([16][29]) and special issues of HCI journals (e.g. [10][17]). Väänänen-Vainio-Mattila et al. [29] identified a set of requirements for practical UX evaluation methods. Requirements for UX evaluation in an industrial context have been identified by Ketola et al. [12]. Roto and colleagues investigated 30 UX evaluation methods during a SIG session at the CHI'09 conference ([25][26]). They found differences in requirements on UX evaluation methods in academia and industry. Industry needs methods that are lightweight, fast, and relatively simple to use. Academia emphasizes the importance of scientific rigor in the methods. Common requirements for industry and academia are: including experimental aspects and allowing repeatable and comparative studies in an iterative manner.

Although a majority of UX evaluation methods originates from usability [26], knowledge on UEMs is not completely transferable on UX. A clear understanding of the differences between usability and UX evaluation methods and measurement models is still missing. There is a need for systematic knowledge on UX methods. Furthermore, there is a need for UX evaluation methods targeting community oriented applications [3].

By investigating usefulness of seven methods used for evaluation of user experience in the context of applications for sharing and co-creating user-generated content, this paper aims to increase our knowledge on UX methods.

UX EVALUATION FRAMEWORK APPLIED WITHIN THE PROJECT

We have developed a common framework for evaluating and addressing users' experiences. Based on the previous work and the needs of the project we have identified eight central factors considered as relevant for investigating users' experiences with audio-visual networked applications. UX is investigated from an individual perspective, and is further influenced by the social context [15] of the evaluated applications [22]. Thus, we included co-experience (UX6) and sociability (UX7) as relevant factors addressing these social influences on the individual experience in our UX evaluation framework [20]. The co-experience approach [1] was considered as relevant for the testbeds – urban and rural communities – within the CITIZEN MEDIA project, as it focuses on the sharing of an experience and provides the basis for building relationships. From a methodological point of view we tried to investigate UX as social by applying group-based evaluation methods, which still need to be extended in the future [3]. Table 1 lists these factors together with the main questions (further sub-questions were defined) they address. These UX factors were applied to collect user feedback from all three testbeds and to detect common UX problems or demands (see resulted UX patterns in [18]).

Table 1. UX Factors used as a Starting Point within the CITIZEN MEDIA Project

UX Factor	Question
UX1 Fun/enjoyment	To what extent do users enjoy the applications in real usage?
UX2 Emotion	Which emotions arise from the usage of the developed applications?
UX3 Motivation	Why are users motivated to participate, contribute and co-create networked audiovisual content?
UX4 User engagement	Who are the users with the highest interest in user-generated media?
UX5 User involvement	How does user involvement increase over time?
UX6 Co-experience	How do the developed applications support co-experience?
UX7 Sociability	How do the developed applications support human-human interaction?
UX8 Usability	To what extent are the users satisfied with the usability of the developed applications?

Table 2 briefly describes the usage of the evaluation methods within the project. For each method we give a brief description of the type and maturity of the application that was evaluated. A detailed description of the UX factors and evaluation methods we used can be found in [20].

Table 2. Used Evaluation Methods within the CITIZEN MEDIA Project: Well known methods were combined or adapted in order to investigate UX in Germany, Norway, and Austria.

(UX) Evaluation Methods	Method Description
Lab based user study	(1) User study with think-aloud and eye tracking [13]; IPTV (Internet Protocol TV); early prototype (2) User study with bio-physiological measurements; IPTV; working product (after 3 months of usage)
Focus group	Group interviews with a facilitator [14]: (1) Less structured than usual focus groups; combined with a short questionnaire; IPTV; early in the process [22]. (2) Focus group with free exploration session integrated into a workshop; two web-based applications; during the design phase
Experience sampling (ESM)	ESM implemented as a part of the application [21]; answering by clicking on smiley-faces; web-based application for collaborative story telling; non-public alpha version of the application
Online survey	Web-based survey with closed questions [13]. Web based application for sharing User Generated Content (video); shortly after the application went online; use case (content based communication); both early and later in the evaluation
Group-based expert walkthrough	(1) Scenario based usability inspection method [5]; web based application for sharing User Generated Content (video); after the application went online (2) A variation of the method combining elements from focus groups and usability evaluation [6]; also used in combination with focus group elements; web-based application for sharing music; prior to redesign; in combination with focus group elements used for evaluation of beta-version (3) Group-based expert walkthrough in combination with focus groups elements; hands-on sessions also included; web-based application for collaborative story telling; non-public alpha version of the application (4) In combination with discussion and free exploration; unified Electronic Program Guide [23].
Extended heuristic evaluation	Extended heuristic evaluation was a variation of the standard heuristic evaluation where the test leader moderated the evaluation and provided additional explanations; web-based application for sharing UGC (photos and texts) on a city map; evaluation of the paper prototype
Interview	Interviews with application domain experts preceded by hands-on sessions [5]; web based application for sharing User Generated Content (video); ready product

OUR APPROACH

As a step towards building and consolidating knowledge on UX methods, we wanted to explore the usefulness of the UX methods in a real-life context from the perspective of the researchers and developers working in the project. We thus focused our research on the scope and downstream utility. In the context of this research, scope describes what kind of user experience factors a method is good and not good in finding. Both downstream utility and scope are subjectively evaluated by the researchers in the project. To collect the data we developed two open-ended questionnaires. The questionnaire evolved through several iterations for optimal clarity and accuracy. We sent the survey to eight researchers involved in the evaluation activities. Six of the researchers were experts in HCI and usability, and two were master students focusing their studies on HCI and user experience research. All of them had relevant methodological expertise and were provided training if needed.

The first questionnaire collected the following information about the evaluation method: description of the method, the resources used on data collection and analysis, description of the amount and the type of the collected data, and the rationale for using this method. The second questionnaire

collected background information about a researcher, the researcher's general opinion on the method, the researcher's experience with the method in this project, including usefulness and drawbacks of the method and lessons learned. The analysis was done by one researcher. To reduce the threat to validity that might introduced by this, and to facilitate analysis of the qualitative data we used the coding process described by Seaman [27]. The collected answers were categorized according to the above described criteria. .

USEFULNESS OF THE EVALUATION METHODS AND RECOMMANDATIONS

This section both reports our findings on the usefulness of the used evaluation methods for capturing UX, and provides relevant recommendations. Table 3 gives an overview of the methods we used for capturing UX factors. When describing the usefulness of a method for capturing different UX factors, the researchers also reflected on the cost/benefits of a method.

Table 3. Scope of the used Evaluation Methods with regard to the UX Factors addressed within the project

UX Factor	(UX) Evaluation Method
UX1 Fun/enjoyment	Lab based user study, focus group, ESM, group-based expert walkthrough
UX2 Emotion	User test, focus group, ESM, online survey, group-based expert walkthrough
UX3 Motivation	Focus group, online survey, group-based expert walkthrough, interview
UX4 User engagement	Focus group, online survey, group-based expert walkthrough
UX5 User involvement	Online survey, group-based expert walkthrough, interview
UX6 Co-experience	Group-based expert walkthrough
UX7 Sociability	Focus group, ESM, online survey, group-based expert walkthrough, interview
UX8 Usability	User test, focus group, ESM, group-based expert walkthrough, extended heuristic evaluation, interview

Lab based user studies with bio-physiological measurements were reported to be useful for capturing fun, emotions, and usability, particularly when reaching the users in a real life environment was difficult. However, the method is complex in terms of data collection and analysis. Hands-on sessions preceded interviews, were integrated in workshops with focus groups, and used in an adapted version of group-based expert walkthrough. The importance of these sessions for capturing UX was emphasised by all researchers. Common experience in exploring applications made it easier for the participants to talk about non-functional aspects of the applications, particularly about enjoyment, emotions, motivation, co-experience and sociability. One could compare enjoyment and emotions of users when using different functions of the old and the new version of an application. When evaluating another application, the participants worked together on a common collaborative task (writing a story together). This common experience made it easier to discuss feelings related to use of these applications such as emotional response when a co-author has deleted a paragraph. A common task has been very useful for initiating discussions on sociability and co-experience.

Recommendation 1: Encourage collaboration. Investigating motivation, user engagement, user involvement, co-experience and sociability at the level of communities and families is essential for applications aiming to support sharing and co-creation of UGC. Both tasks and evaluation methods should reflect this priority. Extending well known methods such as interviews, focus groups, and group-based expert walkthroughs with **hands-on sessions** and usage of **collaborative tasks** has been very useful for capturing these factors.

The researchers also reflected on the importance of different UX factors in the different project phases and in the relation to the availability of other UX evaluation methods. For example, a researcher said: *Especially in this early phase in the evaluation process, issues concerning motivation need to be investigated in detail. The online questionnaire was valuable in doing so.... Since no logging data was available at this point of time in the evaluation process, it was good to receive any information about the usage of the platform.*

Recommendation 2: Start to evaluate UX as early as possible. Early feedback is very valuable to the developers. In particular, feedback on motivation, emotions, and anticipated engagement is valuable. However, one should adapt both the methods and the measurement to the evaluation phase. As the project progresses, one can move towards finer granularity evaluation. For example, one can measure the emotions related to a general idea of a tool for collaborative writing early in a project and emotions related to a particular function of the tool later in the project.

Not surprisingly, usability was easiest to measure, as it is the most standardized factor. When describing the usefulness of a method for capturing usability, our respondents used the term "very useful" without exception. On the other hand, fun, emotions and co-experience were reported as difficult to measure. Furthermore, they pointed out the centrality of usability and its effects on other user experience factors: *In my opinion, a usability test is an essential part of a user experience evaluation, because if the usability of an application is bad, this has further effects on other UX factors like motivation or user engagement among the users.*

Recommendation 3: Evaluate usability and its influence on UX. Evaluating usability together with other UX factors is beneficial particularly early in the project. Other factors often might be affected by usability (e.g. motivation). Capturing several factors together thus makes it easier to understand the results and to organise the studies. One the other hand, one should not explore too many factors in the same study.

A summary of downstream utility (Table 4) is based on self-reported usage of the evaluation results for the further design and the development of the application. All the methods have been reported as useful for the subsequent project phases. When describing the usefulness of the UX feedback collected by a method, the researchers always related usefulness to the complexity of the analysis (simple analysis was appreciated), the phase of the project (early feedback was appreciated), and the necessary effort. For example, feedback from expert interviews was directly used to inform design, but the researcher reported that a lot of interviews were needed in order to capture feedback from different stakeholders.

Table 4. Downstream Utility

(UX) Evaluation Method	Downstream Utility
Lab based user study	Useful; usability problems identified; complex
Focus group	Useful; list of suggestions provided; can explore only limited number of UX factors in a session
Experience sampling	Useful; simple analysis
Online questionnaire	Very useful; UX trends captured early; past behaviour and opinions; simple analysis
Group-based expert walkthrough	Useful; specially for exploring common community experience
Extended heuristic evaluation	Useful; new solutions provided
Expert interview	Useful; demands a lot of effort

Collected feedback influenced design by capturing users' past behaviour and trends, identifying specific problems, identifying solutions, providing better user experience, providing new solutions or ideas for improvements, and providing rationale and ideas for complete redesign. In one case, a negative user experience collected by an expert group walkthrough led to a complete redesign of the application. In particular, feedback on motivation and emotions had a great persuasive effect on the design team. The participants stated clearly that they could not see the purpose of an application for collaborative writing and that writing is something very private for them. For evaluating the next version of the same application, ESM was used together with group-based expert walkthrough for collecting feedback on enjoyment, emotions, and sociability. The feedback was very positive, and only some minor changes of the applications were proposed.

Recommendation 4: Evaluation should be playful and provide added value for the participants. One cannot overemphasize the importance of providing a safe, comfortable and playful evaluation environment, and giving 'something extra' to the study participants. The opportunity to learn and try something completely new and to affect the development of new applications is not only very stimulating and rewarding for the communities of users and experts participating in the evaluation, but also positively affects usefulness of the evaluation methods. When working with communities it is very important to build a trustful relationship for ensuring a successful long term relationship.

Although they are a commonplace in usability evaluation, simple recommendations such as "Conduct evaluation in nice and familiar environment", "Prepare playful tasks", "Use original and playful ways for studies promotion", were repeatedly reported by the researchers as very important for the usefulness of the methods used.

Recommendation 5: Prepare for diversity. In depth knowledge of your communities—the different groups of users and non-users—is essential for successful data collection. Different versions of questionnaires and focus group guidelines should be prepared for different user groups (e.g., professional cabaret artists, amateur artists, and theatres) and evaluators/moderators should be able to speak 'different languages' (e.g., to talk to children, teenagers, and elderly people) at the same time.

When describing the usefulness of the evaluation results, researchers emphasised importance of good knowledge of communities and relationships among them. Questionnaires tailored to different communities have been more useful than general ones. The researchers also reported that good collaboration with designers and developers teams was important for uptake of the evaluation results. Good knowledge of the application including the ideas of the designers that might be not yet implemented or presented at a paper prototype was very useful, as well as the ability to clearly and quickly report the results on user experience. Quotations being typical for users' emotions and motivations were highly appreciated by the designers and developers.

Recommendation 6: Be best friends with the developer. Good knowledge of the application under development is very important for the success of the evaluation. Evaluators/moderators should be able to

explain ideas behind paper prototypes and screenshots. Communicating the results of the evaluation clearly and in formats understandable to the developers is extremely important for uptake of the evaluation results.

CONCLUSIONS AND FUTURE WORK

We conducted a survey among the researchers involved in the evaluation activities of the CITIZEN MEDIA research project that developed a plethora of applications supporting non-professional users in sharing and co-creating user-generated content. Combinations of well known evaluation methods and their home-grown adaptations were used (as there were no clear defined UX evaluation methods available yet, fitting the needs of the project). Our results indicate that group based evaluation methods (group-based expert walkthrough and focus groups) were useful for measuring a broad spectrum of the pre-defined UX factors. Some factors such as emotions, fun, and co-experience were difficult to measure and there is an urgent need for development of such methods. Furthermore methods for sharing individual experience have to be extended to capture shared experience of community of users. Collaborative playful methods and collaborative tasks supported well move from individual user evaluation methods to community evaluation methods (e.g. [3][23]).

Within this paper, we summarized our results and lessons learned from the evaluation activities in several recommendations, which might be useful for practitioners working in the area of UX in general, and UX of social media applications in particular. Furthermore, our experience might be a useful input for the ongoing discussions on UX evaluation methods and measurement within the HCI research community, which special attention on how to support the design and development process of new applications, software, or systems.

As pointed out by Blandford et al. [1], comparison of evaluation methods is very complex and cannot be done by one study. Although our study covers a broad range of evaluation methods, UX factors and social media applications, it does not draw on a large data collection from numerous subjects with different background, experiences, and contexts. Furthermore, usefulness was subjectively evaluated by the researchers in the project while the development process was still in progress. We plan to extend our work by mail-based interviews of the developers investigating down-stream utility in more details and with objective evaluation of usefulness based on the inspection of the project's documentation and tracing of actual design changes. We also encourage other researchers to validate and complement our recommendations by further studies.

ACKNOWLEDGMENTS

This work was supported by the CITIZEN MEDIA research project (funded by FP6-2005-IST-41 038312) and by the R2D2 Networks research project (CELTIC and VERDIKT programs). We thank the partners involved in the development and evaluation activities. We also thank the participants in our studies.

REFERENCES

[1] Battarbee, K. Defining co-experience. In *Proceedings of the 2003 international Conference on Designing Pleasurable Products and interfaces* (Pittsburgh, PA, USA, June 23 - 26, 2003). DPPI '03. ACM, NY, 2003, 109-113.

[2] Blandford, A.E., et al., Scoping Analytical Usability Evaluation Methods: A Case Study Human-Computer Interaction, 2008. 23(3): p. 278-327.

[3] Brandtzæg, P.B., Følstad, A. Obrist, M., Geerts, D. & Berg. R. (2009). Innovation in online communities - Towards community-centric design. In *Proceedings of the First International Conference, UCMedia 2009*, Venice, Italy, December 9-11, 2009. In P. Daras Ibarra and O. Mayora (Eds.), LNICST, Springer, Vol. 40, pp. 50–57.

[4] Dix A., et al., Human-Computer Interaction. 3rd ed. 2003: Prentice Hal.

[5] Følstad, A. Group-based Expert Walkthrough. in R3UEMs: Review, Report and Refine Usability Evaluation Methods, COST924-MAUSE 3rd International Workshop. 2007. Athens, Greece.

[6] Følstad, A., Work-Domain Experts as Evaluators: Usability Inspection of Domain-Specific Work-Support Systems. International Journal of Human-Computer Interaction, 2007. 22(3): p. 217-245.

[7] Gray, W.D. and M.C. Salzman, Damaged merchandise? A review of experiments that compare usability evaluation methods. Human-Computer Interaction, 1998. 13(3): p. 203-261.

[8] Hartson, H.R., T.S. Andre, and R.C. Willings, Criteria for evaluating usability evaluation methods International Journal of Human-Computer Interaction 2001. 13(4): p. 373-410.

[9] Hassenzahl, M. User experience (ux): towards an experiential perspective on product quality. In IHM'08: Proc. of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine, pages 11–15, New York, NY, USA, 2008. ACM.

[10] Hassenzahl, M. and M. Tractinsky, User experience – a research agenda Behaviour & Information Technology, Empirical Studies of the User Experience, 2006. 25(2): p. 91-97.

[11] Holzinger, A., Usability engineering methods for software developers Communications of the ACM, 2005. 48(1): p. 71-75.

[12] Ketola, P. and V. Roto, On User Experience Measurement Needs - Case Nokia. International Journal on Technology and Human Interaction (IJTHI), 2009. 5(3).

[13] Kuniavsky, M., Observing the user experience a practitioner's guide to user research. 2003, San Francisco, USA: Morgan Kaufmann

[14] Krueger, R.A. and M.A. Casey, Focus Groups: A Practical Guide for Applied Research. 2000, Thousand Oaks, Calif: Sage Publications Inc, Thousand Oaks, Calif

[15] Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P. and Kort, J. Understanding, scoping and defining user experience: a survey approach. In CHI'09: Proc. of the 27th international conference on Human factors in computing systems, pages 719–728, New York, NY, USA, 2009. ACM.

[16] Law, E.L.-C., E.T. Hvannberg, and M. Hassenzahl. COST294-MAUSE Workshop on User Experience - Towards a Unified View. Workshop at NordiCHI'06 conference. 2006; Available from: http://nordichi.net.dynamicweb.dk/Workshops/W2-User-Experience_.aspx.

[17] Law, E.L.-C. and v. Schaik, Modelling UX – an agenda for research and practice. Interacting with Computers., 2010. 22(5): p. 313-438.

[18] Obrist, M., Wurhofer, D., Beck, E., Karahasanovic, A., and Tscheligi, M. (2010). User Experience (UX) Patterns for Audio-Visual Networked Applications: Inspirations for Design. Accepted full paper to NordiCHI2010.

[19] Obrist, M., V. Roto, and K. Väänänen-Vainio-Mattila. User experience evaluation: do you know which method to use? CHI Extended Abstracts 2009: 2763-2766. in SIG session at CHI 2009, available in CHI Extended Abstracts. 2009. Boston, USA: ACM.

[20] Obrist, M., Weiss, A., and Tscheligi, M. Evaluating user-generated content creation across contexts and cultures. In IAMCR2007: Proceedings of the 50th Anniversery Conferneces of the International Association for Media and Communication Research (Paris, July 2007).

[21] Obrist, M., Meschtscherjakov, A., and Tscheligi, M. User experience evaluation in the mobile context. In Aaron, Marcus, Cereijo Roibás, Anxo, and Sala, Riccardo (Eds). Mobile TV: Customizing Content and Experience. Human-Computer Interaction Series, Springer London, 2010.

[22] Obrist, M., Miletich, M., Holocher, T., Beck, E., Kepplinger, S., Muzak, P., Bernhaupt, R., and Tscheligi, M. Local communities and IPTV: Lessons learned in an early design and development phase. In Computers in Entertainment 7, 3 (Sep. 2009), 1-21.

[23] Obrist, M., Moser, C., Alliez, D., Holocher, T., and Tscheligi, M. Connecting TV & PC: An In-Situ Field Evaluation of an Unified Electronic Program Guide Concept. In EuroITV2009: Proceedings of 7th European Conference on Interactive Television (New York, NY, USA, 2009), ACM, 91–100.

[24] Preece, J., Y. Rogers, and H. Sharp, Interaction design: beyond human-computer interaction. New York: John Wiley & Sons. 2002, New York: John Wiley & Sons.

[25] Roto, V., et al. User Experience Evaluation Methods in Product Development (UXEM'09), Workshop at INTERACT 2009; Available from: http://wiki.research.nokia.com/index.php/UXEM09.

[26] Roto V., Obrist M., and K. Väänänen-Vainio-Mattila, User experience evaluation methods in academic and industrial contexts, in Workshop on User Expererience Evaluation Methods, in conjunction with Interact'09 conference. 2009: Uppsala, Sweden.

[27] Seaman, C., Qualitative Methods in Empirical Studies of Software Engineering. IEEE Transactions on Software Engineering, 1999. 25, No. 4(July/August 1999): p. 557–572.

[28] Väänänen-Vainio-Mattila, K., V. Roto, and M. Hassenzahl. Now let's do it in practise: User eXperience Evaluation Methods in product development (UXEM). CHI 2008 Workshop; Available from: http://www.cs.tut.fi/ihte/CHI08_workshop/.

[29] Väänänen-Vainio-Mattila, K., V. Roto, and M. Hassenzahl. COST294-MAUSE Workshop on Meaningful Measures: Valid Useful User Experience Measurement (VUUM). in Proceedings of the COST294-MAUSE, 18th June, 2008, Reykjavik, Iceland.

[30] Vermeeren, A.P.O.S. , Law, E.L-C., Roto, V., Obrist, M., Hoonhout, J., and Väänänen-Vainio-Mattila, K. (2010). User Experience Evaluation Methods: Current State and Development Needs. NordiCHI2010, Reykjavik, Iceland

Matching HCI Methods and Developers Values in eXtreme Programming Development Processes

Peter Wolkerstorfer, Manfred Tscheligi, Claudia Oppenauer-Meerskraut, and Arjan Geven
CURE - Center of Usability Research & Engineering
Modecenterstraße 17 / Objekt 2, 1110 Vienna
{wolkerstorfer| tscheligi| oppenauer| geven}@cure.at

ABSTRACT
Here within we describe our practical experience with the orchestration of human computer interaction (HCI) methods and extreme programming (XP) software development processes. We show how we selected the applied methods based on the motivating goals and values of developers by using a means-end approach. We discuss our experiences with the applied methods and conclude with some advice on which HCI methods are optimally supporting extreme programming developers.

Author Keywords
HCI methods, software development, extreme programming, usability.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
A problem when HCI engineers are to collaborate with extreme programming software developers is the difference between cultures: Software engineers on the one hand and HCI experts on the other hand come from different domains with different attitudes, approaches, backgrounds, and even different ways to express themselves. The XP process requires tight cooperation in teams, which reveals differences between engineers and HCI experts very quickly: engineers have a technical approach to software development whereas HCI experts mainly have a psychological background, hence taking a cognitive view on the software development. These differences can lead to problems. Methods to prevent this have to be integrated into the collaboration process. To avoid cultural problems HCI methods must fit the developers´ needs. Our approach to provide this match of methods is to look at the psychology (goals and values) of developers and derive user-experience (UX) requirements (REQ) from them. In our case UX is meant as the programmers experience in relationship to the applied HCI methods. On this basis we select HCI methods to optimally support the development process. The insights in this paper come from a research project where the goal was to orchestrate usability and XP processes.

GOALS AND VALUES OF XP- DEVELOPERS

Foundations
In order to identify the necessary selection of HCI methods we need to consider that HCI methods - in most of the cases - are not the prime focus of XP programmers, not under their constant attention nor necessarily fit for application in XP- processes. Therefore, to achieve a higher user-centeredness and an enhanced usage and acceptance of usability methods in agile teams the following two pillars need to be fostered:

a.) Position usability methods in a way that they fit the agile team structure and process without disturbing the primary task: software development (=adoption towards the organizational and process goals).

b.) Align the usability methods towards the programmer's goals and values in order to achieve acceptance and use of these methods beyond indoctrination (=adoption towards psychological and developer's goals).

Developer's UX Requirements
The methods we used to elicit the requirements are focus groups due to availability of the developers and a means end approach [[2]] as it provides insights into motivating goals and values of developers. This was done to elicit the UX requirements of developers related to HCI methods. The findings suggest that the requirements are, that HCI methods:

- REQ1: have to be easy to apply
- REQ2: are efficient (in terms of time and cost)
- REQ3: are non-intrusive related to the developers workflow
- REQ4: support team orientation and inter-team communication
- REQ5: enable learning and finding new approaches
- REQ6: must make ambition (professional AND personal achievement) achievable

HCI METHOD SELECTION

Pool of Methods

When we started our research project it was not clear to us which HCI methods should be preferred in XP development processes. Hence we started with a method mix containing: user studies, usability laboratory tests, usability expert evaluations, (adopted) personas, and extended unit-tests in the sense of automated usability evaluation (AUE) [[4]]. The question then was: which HCI method to select?

Selection

Personas

Personas are archetypical descriptions of real users, representing the target user group. Personas are often described in a narrative way and are designed to help software developers to get a better understanding of the real end-user they are developing for [[1]]. We have chosen the personas method based on requirements REQ1, REQ2, REQ3 and REQ4.

Extended Unit Tests

In XP unit testing is mandatory. Our approach extends the technical unit tests by adding usability- specific test cases. At the time writing we are experimenting with a graph-based approach [[3]]. We have chosen extended unit tests based on requirements REQ2, REQ3, REQ5 and REQ6.

CONCLUSION

Personas have been a great success and have also been honoured by the developers. An observation worth reporting is that there seems to be a small fraction (in our case: one out of six) of developers who are “resistant” to the personas method – hence reject it completely. We suggest conducting a psychological screening of developers before setting up teams to be able to identify those people in order to cope with the problem.

The experience with the approach to extend unit tests is twofold: on the one hand we succeeded in including automated usability evaluations in the nightly build. On the other hand the actual testing frameworks are not suited for AUE. Hence new AUE-tools have to be developed. Our graph-based approach is promising but – by now – much to abstract (there is a need for graph- and HCI knowledge to interpret the results). For practical implementation such tools have to provide easy to understand and clear usability feedback to the developers.

FUTURE RESEARCH

Future research will have to provide AUE-tools, which enable developers to easily grasp HCI knowledge from the tool in order to implement usability accordingly.

ACKNOWLEDGMENTS

The research herein is partially conducted within the competence network Softnet Austria (www.soft-net.at) and funded by the Austrian Federal Ministry of Economics (bm:wa), the province of Styria, the Steirische Wirtschaftsförderungsgesellschaft mbH. (SFG), and the city of Vienna in terms of the center for innovation and technology (ZIT).

REFERENCES

[1] Pruitt, J. and Adlin, T. 2005 The Persona Lifecycle: Keeping People in Mind Throughout Product Design (The Morgan Kaufmann Series in Interactive Technologies). Morgan Kaufmann Publishers Inc.

[2] Reynolds, Thomas J. and Olson, Jerry C.; Lawrence, Understanding Consumer Decision Making, The means-end approach to marketing and advertising strategy; Erlbaum Associates, 2001

[3] Thimbleby, H. and Gow, J. 2008. Applying Graph Theory to Interaction Design. In Engineering interactive Systems: EIS 2007 Joint Working Conferences, EHCI 2007, DSV-IS 2007, HCSE 2007, Salamanca, Spain, March 22-24, 2007. Selected Papers, J. Gulliksen, M. B. Harning, P. Palanque, G. C. Veer, and J. Wesson, Eds. Lecture Notes In Computer Science, vol. 4940. Springer-Verlag, Berlin, Heidelberg, 501-519. DOI= http://dx.doi.org/10.1007/978-3-540-92698-6_30.

[4] Wolkerstorfer, P., Tscheligi, M., Sefelin, R., Milchrahm, H., Hussain, Z., Lechner, M., and Shahzad, S. 2008. Probing an agile usability process. In CHI '08 Extended Abstracts on Human Factors in Computing Systems (Florence, Italy, April 05 - 10, 2008). CHI '08. ACM, New York, NY, 2151-2158. DOI= http://doi.acm.org/10.1145/1358628.1358648

Analysis of the Persuasiveness of User Experience Feedback on a Virtual Learning Environment

Daniel Müller
IMC
Altenkesseler Str. 17 D3
66115 Saarbruecken,
Germany
+49 681 302 64752
daniel.mueller@im-c.de

Effie L.-C. Law
Dept. of Computer Science
University of Leicester
LE1 7RH
Leicester, UK
+44 116 252 5341
elaw@mcs.le.ac.uk

Stefan Strohmeier
Chair of MIS
Saarland University
66041 Saarbruecken,
Germany
+49 681 302 64751
s.strohmeier@mis.uni-saarland.de

ABSTRACT
The main purpose of this paper is to evaluate a set of theory-grounded User Experience (UX)-related measures which are supposed to persuade course designers of particular UX-related problem areas based on a specific feedback format. Specifically, two online surveys on a university online course were conducted with the former focusing on the quantitative ratings and the latter on qualitative comments. The course designers were asked to assess the persuasiveness of the feedback with respect to eight dimensions. The results show that UX-related problem areas anchored in the DeLone and McLean's Information Systems (IS) Success Model (ISSM) had a consistently higher level of perceived persuasiveness than those anchored in the Technology Acceptance Model (TAM) 3. The discrepancy can be attributed to the nature of items: process- vs. trait-based). Implications for future research on fixing UX-related problems are discussed.

Keywords
Course Designer, Design Characteristic, Feedback Format, IS Success Model, Persuasiveness, TAM3, User Experience.

INTRODUCTION

At the present day rigorous, i.e. theory-grounded, and relevant, i.e. practice-oriented, approaches for the design and evaluation of Virtual Learning Environments (VLE) are needed to improve the understanding and communication of educational needs among all stakeholders, including researchers and practitioners [10]. In this paper VLE are understood as systems for the administrative and didactical support of learning processes in higher education and vocational training settings by means of formal online courses [22]. Hence, it is of great importance to investigate the drivers or determinants of VLE success to assist system and course designers in building, operating and sustaining systems and online courses as integral parts that are useful and accepted by the end-user (here: students). However, a specific theory of successful VLE is currently missing as existing approaches focus on information systems (IS) in general with the DeLone and McLean's ISSM [2, 3] being one of them. Contrary to the TAM, which does not propose concrete system design guidelines [23, 32, 34, 36], the ISSM identifies and provides general qualities which are thought to enhance user satisfaction, the use of, and the net benefit(s) (NB) of using a VLE [2, 3]. However, the main disadvantage of the ISSM used as a general approach is that specific VLE-related success drivers cannot be directly derived from the model itself. Rather, the ISSM offers insights into the process of how general qualities, namely system- and information quality, influence the final success [2, 3]. Hence, the ISSM offers a general and "useful framework for organizing IS success measurements" [27] which can and should be adapted to the VLE context [3, 27]. Though, beside more general recommendations for the selection of success measures [29], there currently lacks a widely accepted set of measures relevant to VLE in particular.

However, some latest research attempts striving for a VLE-specific extension of the ISSM revealed a comprehensive and exhaustively validated set of system- and information-related design characteristics relevant to VLE in particular [22]. As some of these design characteristics, respectively their corresponding items can be adequate measures for UX as well, these UX-related design characteristics may support designers (here: course designers[1]) in their attempts to fix not only usability-related issues [25] but also UX-related problem areas (e.g. *image*, see Table 1). Thereby, UX-related problem areas anchored in the ISSM are compared against the ability of selected UX-related ones anchored in the TAM3 [32] in order to carve out differences in the persuasiveness of the feedback format for course designers. This construct is assumed to be dependent on a) the kind of theory applied (ISSM = product-oriented; TAM3 = state-/trait-oriented) and b) the information richness of the feedback format provided.

According to Nørgaard and Hornbæk [25], the underlying assumption is as follows: The richer the UX problem area-related contextual information contained in the feedback format, the higher persuasiveness of this feedback format is for course designers. Thus, in search for a rigorous and

[1] In addition to research efforts solely focusing on system designers [e.g. 6, 7, 8, 4, 15, 16, 25].

persuasive UX-related feedback format[2], the following research questions (RQ) will be addressed in this paper:

RQ1: To what extent do students as end-users have more problems in specifying UX-related problem areas based on TAM3-related UX items than those based on their ISSM-related counterparts (see Table 1, *students' item rating statements*)?

RQ2: To what extent do course designers perceive UX-related problem areas (see Table 1: based on the *construct label*, *construct definition*, *item wording*, *item-UX-match*, *students' item rating* and *students' item rating statement*) originated in the ISSM to be more persuasive than their TAM3-related counterparts?

RQ3: Which of the UX-related problem areas (TAM3- vs. ISSM-anchored) do course designers perceive to be more persuasive in case *evaluators' suggestions* are provided in addition to the set of UX problem area-related contextual information illustrated in Table 1?

Based on these research questions, the main purpose of this paper is to identify and validate a set of theory-grounded, UX-related measures of which persuasiveness presumably varies with feedback format. In this context, we define *persuasiveness* in terms of convincing course designers about the problematicity of particular UX-related issues, which may entail specific resolutions.

In the paper we first explore the concept of UX as well as feedback formats as a means to persuade designers. Then we present the methodological framework regarding students' specification of UX-related problem areas as well as course designers' assessment of their persuasiveness of the feedback formats generated. Next, we describe the empirical results with regard to the overall persuasiveness of the feedback format (quantitative evaluation) and particular UX-related problem areas (qualitative evaluation) as perceived by the course designers. The aforementioned three research questions will be then discussed. Finally, implications for future research efforts and conclusion are drawn.

BACKGROUND

User Experience

As distinct from usability-centred evaluations which roughly focus on task-related issues such as efficiency and effectiveness [7], "[UX] proposes a more holistic view of the user's experience when using a product than is usually taken in the evaluation of usability" [12]. While pragmatic/do goals are associated with usability, hedonic/be goals address cognitive, socio-cognitive and affective aspects of users' experience in their interaction with artifacts (e.g. users' enjoyment, aesthetic experience, desire to repeat use, positive decision to use a digital artifact and enhanced mental models) [1, 16]. However, a consensual definition of UX does not yet exist, although ISO 9241-210 [13] provides one: "*A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service*" (clause 2.15), which is relatively simpler than, for instance, the comprehensive definition by Hassenzahl and Tractinsky [8], one of the many definitions in the literature [15]. In an attempt to understand the diverse interpretations of UX, the researchers, based on the results of a survey [15], have drawn a conclusion that UX is "*dynamic, context-dependent and subjective, which stems from a broad range of potential benefits users may derive from a product*".

Feedback Formats as a Means to Persuade Designers

Feedback can be understood as "*information about reactions to a product, a person's performance of a task, etc. which is used as a basis for improvement*" [25, 26]. According to Nørgaard and Hornbæk [25], feedback should fulfill the requirement of being persuasive: Firstly, feedback should convince developers that the problem identified does exist and helps them to understand it. Secondly, the persuasiveness of a feedback format is determined by the amount of contextual information about a problem it conveys. Thirdly, the ease the feedback can be used in the developer's everyday work is important. Moreover, given that UX is inherently dynamic, context-dependent and subjective [15, 16], feedback on UX-related problems should essentially be self-reported data to be captured by questionnaire, interview and think-aloud. Nonetheless, data on UX can be gathered with a survey where respondents are first required to rate a set of items with a Likert scale and subsequently interviewed to elaborate their comments on the items.

METHOD

Identifying Students' UX-related Problem Areas

Participants of the current study were eleven students of the online course *Organizational Management*, which was delivered during the summer term 2010 by the Chair of Management Information Systems (MIS) located at Saarland University/Germany. Prior to this study, these participants had been asked to complete a larger-scale online survey with 88 items being originated from ISSM and TAM3. The aim of this first survey (N=30) was to evaluate students' acceptance towards the aforementioned course. Thereby, all items showed high levels of construct validity, evaluated via a satisfactory convergent (Average Variance Explained, Composite Reliability, significant indicator factor loadings exceeding a threshold of 0.70), discriminant, and nomological validity. Some of the first survey items can be mapped to the hedonic attributes of the model of UX proposed by Hassenzahl [6]. As the primary focus lies on more hedonic attributes, more pragmatic/task-related ones such as the perceived usefulness as well as the perceived ease of using a VLE are out of scope of this paper [8, 12, 32]. The mapping was undertaken by the first and second authors of this paper, resulting in 17 items that constitute the second online survey. Consequently, the second survey consists of UX-related items. Specifically, we define a UX-related item as a *problem area* if its mean rating (averaged over all the respondents involved in the first survey) falls between 1.00 and 3.99 and as a *still-to-*

[2] As distinct from research efforts solely focusing on the persuasiveness of usability-related feedback formats [e.g. 25].

be-improved area if it is between 4.00 and 4.50. In this case, a UX-related construct (the column Label in Table 1) could contain both types of area. The corresponding boundary values are defined by the MIS monitoring team responsible for the quality control of the MIS's online courses. Moreover, UX-related problem areas originated from either the TAM3 or the ISSM are randomly put in one sequence to prevent sequence effects. In addition to the item (Table 1) the participants were provided with the corresponding average ratings[3] (1 = strongly disagree; 5 = strongly agree) as well as a hint to their individual ratings of the first online survey (students were provided with a copy of their individual ratings after having completed the first online survey) and were asked to comment on them. The reasons for providing students with the averages as well as a corresponding hint to their individual rating in the first online survey are due to the university's data policy prohibiting the chair's monitoring team to confront students with their individual ratings of a preliminary survey directly. Thus, UX-related problem areas were further specified and thus contextualized based on students' additional qualitative written input (illustrative example per item, see Table 1, column 6).

In summary, the tasks the participants of the second survey had to undertake were: (i) They had to provide their personal statements to each item rating by referring to the online course *Organizational Management*; (ii) based on their statements, they had to explain how they would solve the perceived problem areas. This information could serve as a starting point for the evaluator's suggestions (see Table 1, last column), which were presented to the course designers to evaluate their persuasiveness of the feedback format.

Evaluating the Persuasiveness of the Feedback Format by Course Designers

Three course designers (1 professor, 1 research professional, 1 research assistant) were invited to attend semi-structured interviews to gather their evaluation of the persuasiveness of the UX-related problem areas (the first five columns of Table 1). In addition, the contextual information gathered from the participants in the previous step (the 6th column in Table 1) was further concretized with *evaluators' suggestions* how to solve particular UX-related problem areas (the last column of Table 1). The concretization was undertaken by the MIS monitoring team based on the participants' qualitative written input as described in the previous step. Hence, the final feedback format that the course designers were provided with was a context-rich blending of a) a problem list and b) corresponding redesign proposals [25]. Presumably, this can better illustrate the causes and solutions of the UX-related problem areas, and thus enhance the persuasiveness of such a context-rich feedback format as distinct from pure problem enumerations [14]. In order to survey the *persuasiveness* of a) the overall feedback format in general as well as b) the *UX-related problem areas* in particular (see Table 1), a questionnaire was administered which was mainly based on Norgaard and Hornbaek's [25] measurement instrument to evaluate the persuasiveness of a feedback format as perceived by developers. The corresponding questions are:

Q1: "How useful is the information provided in Table 1 (*construct label*, *construct definition*, *item wording*, *item-UX-match*, *students' item rating* and *students' item rating statement*) to your work on the online course *Organizational Management*? (1=not useful – 5=very useful). Furthermore, please comment on the usefulness of the information provided by referring to the UX-related problem areas *label* (e.g. *user interface appeal*, see Table 1, first column).

Q2: How well does the information provided in Table 1 help you to understand the UX-related problem area(s)? (1= very poor – 5 = very well).

a) Please comment on the level of understandability of the information provided in Table 1 by referring to particular *columns* (i.e. vertical evaluation).

b) Additionally, please differ between the understandability of the information provided in Table 1 by referring to particular *item* (i.e. horizontal evaluation).

Q3: How well does the information provided in Table 1 have an impact on assessing the severity of the UX-related problem area(s)? (1=very poor – 5= very well). Please comment on the severity of (a) particular problem area(s).

Q4: How well does the information provided in Table 1 help you solve the UX-related problem area(s)? (1= very poorly – 5= very well). Please comment on the ability of the information provided in Table 1 to solve a particular problem area(s).

Q5: Do you intend to solve the UX-related problem areas illustrated in Table 1? (1 = not at all – 5 = absolutely). If so, which of the problem area(s) and why?

Q6: Are you convinced that the information provided in Table 1 depicts real UX-related problem areas? (1 = not at all – 5 = absolutely). If not, which of the problem areas and why?

Q7: Is the information provided in Table 1 easy to use to solve particular UX-related problem areas? (1 = not at all – 5 = absolutely). If not, which of the problem area(s) and why?

Q8: Does the information provided in Table 1 have an impact on your (re-)design strategy of particular UX-related problem areas of the online course *Organizational Management* (*here: prioritization of particular problem areas*)? (1 = not at all – 5 = absolutely). Please comment on particular problem areas.

[3]The mean values of the corresponding UX-related problem areas are based on students' ratings gathered within the realm of the preliminary large-scaled online survey to evaluate their acceptance with the online course Organizational Management.

Table 1: UX-related Problem Areas as Perceived by Students' of the Online Course "Organizational Management": Overview

A. ISSM-originated						
Label**	Definition**	Item*/**	Item-UX-mapping** (based on the model of UX [6, 16])	Average Item Rating from the 1st survey/ UX-related problem areas item*/**	Additional UX Problem Specification	
					Participants' qualitative comments given in the second survey/ UX Item-related Problem Description**	Evaluators' Suggestion/ UX-related item problem concretization**
User Interface Appeal	VLE are *appealing* if their graphical user interface has a pleasant appearance [11].	The VLE has an attractive graphical user interface [5].	beauty, visual aesthetics	3.36	-	-
Information Credibility	The information provided by VLE is *credible* if they originate from a trustworthy source (e.g. teacher, certified and/or reputable organizations, etc.) [21].	I trust the learning material's originator (e.g. teacher, professional institution/organization) [22].	trust	4.27	„Your learning materials are arranged very neatly, but they are incomplete in parts."	„Please clarify with your students what exactly is perceived to be incomplete, i.e. missing sample solutions and/or missing learning materials? In case the latter does apply, please clarify once again the chair's didactic policy, i.e. the initial set of slides provided constitute a "starting point" which has to be "enriched" by the students' themselves (dilemma: student vs. teacher viewpoint).
		The learning material's originator (e.g. teacher, professional institution/organization) is an (officially) approved source of information [22].	trust	4.18	„I am not able to judge if he is a recognized source of information."	„If applicable, emphasize the chair's competence in this subject domain (e.g. insert a MIS seal of approval on the learning materials/slides provided which is associated with the chair's latest publications (papers, textbooks), awards, etc.)."
		The learning material's originator (e.g. teacher, professional institution/organization) has a good reputation [22].	trust	4.09	„I cannot answer this question likewise."	„If applicable, insert a MIS seal of approval on the learning materials/slides provided which is associated with the chair's reputation in this subject)."

Information Challenge	The information provided by VLE is *challenging* if the learning materials contain difficult but interesting tasks which stimulate learners' curiosity to solve them [21].	The tasks contained (with)in the learning materials arouse my curiosity [19].	Stimulation	3.36	Partly. It may be better to work with open questions that are discussed with the tutor in class.	„If applicable, please consider open questions too which should be discussed mutually in class (i.e. "offline")."
		The tasks contained (with)in the learning materials arouse my ambition [22].	competence	3,55	["Partly."]	If applicable, please consider open questions too which should a) require students' initiative for further research (i.e. stimulate self-directed learning processes) and which should b) discussed mutually in class (i.e. "offline")."
		The tasks contained (with)in the learning materials are appropriately tricky [22].	competence	3,36	„The tasks seem to be far away from the aspiration level of the final examination. Would be better to provide questions which do have the same difficulty level as the ones asked in the final examination."	„Please carify if the sample exercises do have the same difficulty level as the ones provided in the final examination. If not, please upload a mock exam which should be discussed mutually in the final tutorial."

B. TAM3-originated

Label**	Definition**	Item*/**	Item-UX-mapping** (based on the model of UX [6, 16])	Average Item Rating from the 1st survey/ UX-related problem areas item*/**	Additional UX Problem Specification	
					Participants' qualitative comments given in the second survey/ UX Item-related Problem Description**	Evaluators' Suggestion/ UX-related item problem concretization**
Subjective Norm	The degree to which an individual perceives that most people who are important to him think he should or should not use the system [4, 33].	People who influence my behavior (e.g. fellow students, friends, parents, etc) think that I should use the VLE within the scope of my studies [22, 30, 32].	identification	2.00	Those who are not studying at the university may not deal, and thus may not be interested in the VLE."	„If applicable, communicate the benefits of using the VLE (with)in the course to your students (e.g. citing empirical studies which proved the VLE to have a positive impact on students' training success) so that most of them may perceive the use of the VLE/online course to be "obligatory".

Image	The degree to which an individual perceives that use of an innovation will enhance his or her status in his or her social system [20].	Fellow students at my university who use the VLE have more prestige than those who do not [20, 22].	identification	1.18	„Strange question again!“	„If applicable, please point to the innovative kind of teaching method which is applied (with)in the course so that students may get more aware of the novelty of the approach undertaken (as long as this effect may endure, a “demarcation” from other fellow students may be possible).”
		Fellow students at my university who use the VLE have a high profile [20, 22].	identification	1.18	„One does have a high profile when using the VLE?”	-
		Having the VLE is a status symbol at my university [20, 22].	identification	1.27	“What?”	-
Computer Anxiety	The degree of “an individual’s apprehension, or even fear, when she/he is faced with the possibility of using computers” [31].	VLEs do not scare me at all [31].	evocation (negative)	4.00 (inverse)	„Sure…the use of the VLE was explained very well by the tutor!“	“In order to avoid any kind of inhibition in dealing with the VLE/the online course, a compulsory introduction in the use of the VLE/online course should be implemented at the beginning of the semester.”
		Working with a VLE makes me nervous [31].	evocation (negative)	1.18 (inverse)	“Nervous?”	-
		VLEs make me feel uncomfortable [31].	evocation (negative)	1.27 (inverse)	“This question is pretty strange too!”	-
Computer Playfulness	“….the degree of cognitive spontaneity in microcomputer interactions” [35].	I would characterize myself very spontaneous when I use the VLE [31].	stimulation	3.64	„Just do it, don’t think about it!“	“In order to avoid any kind of inhibition in dealing with the VLE/the online course, a compulsory introduction in the use of the VLE/online course should be implemented at the beginning of the semester.”
		I would characterize myself very creative when I use the VLE [31].	stimulation	2.64	“Strange question likewise. What do you mean by creative?”	-
		I would characterize myself very playful when I use the VLE [31].	stimulation	2.45	Playful? Do not know how to interpret and answer this question?”	-

RESULTS

Course Designers' Ratings of the Feedback Format

Table 2 summarizes the quantitative ratings per question (Q1-Q8) by providing the overall persuasiveness of the feedback format (see Table 1) per course designer, the corresponding mean values, standard deviations and perceived persuasiveness of UX-related problem areas.

Course Designers' Comments on the Feedback Format

Supplement to course designers' overall ratings of the feedback format (see Table 2) the following description reveals how course designers perceived the persuasiveness of particular UX-related problem areas.

Perceived Usefulness of the Information Provided (Q1)

All in all, the *item ratings* and *evaluator's suggestions* are considered to be useful to address the following UX-related problem areas, namely *computer playfulness* and *computer anxiety*. Furthermore, students' critique concerning *information challenge* (i.e. provision of sample tests) was assumed to originate from the lack of publicity for the announcement of such tests which were published in the VLE on a regular base. Besides, one of the course designers raised his concerns regarding *image* as a useful UX-related problem area as it was considered to be not related to the course designers' work as a lecturer or author of the course.

Understandability of UX-related Problem Areas (Q2)

With regard to the course designers' *vertical evaluation* of the understandability of individual variables in the feedback format presented in Table 1, the following results were obtained: The *item* as well as the *combination of item and students' comments* per UX-related problem area was found to a) give the first idea that there exist particular UX-related problem areas and to improve the understanding

Table 2: Perceived Persuasiveness of UX-related Problem Areas Feedback Format

Question		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Overall Rating	Designer #1	4	3	2	2,5	5	5	2	4
	Designer #2	4	4	4	5	4	5	5	5
	Designer #3	4	3	3	4	2	4	4	4
Mean Values		**4.00**	**3.33**	**3.00**	**3.83**	**3.67**	**4.67**	**3.67**	**4.33**
Standard Deviation		**0.00**	**0.58**	**1.00**	**1.26**	**1.53**	**0.58**	**1.53**	**0.58**
Perceived Persuasiveness of Particular UX-related Problem Areas									
User Interface Appeal			+	+	**0**	+			
Information Credibility			+	+	**0**	+	+		
Information Challenge		-	+	+	**0**	+	+		++
Subjective Norm				-	**0**	+	-		
Image		-		-	**0**	++	-		
Computer Anxiety		+			**0**	++			+
Computer Playfulness		+	+	+	**0**	++	+		

- = UX problem areas considered not to be persuasive

0 = neutral

+ = UX problem areas considered to be persuasive

of particular UX-related problem areas. However, in order to solve the problem the *evaluator's suggestions* are inevitable. Furthermore, the information provided in Table 1 was considered to give valuable hints, which, however, would need more in-depth information to understand what actually the problem was (e.g. Is the problem of understanding the material caused by the fact that questions do not have examination-level? Or is the VLE misunderstood as a mere examination-preparation-tool?).

Concerning the *horizontal evaluation* of the information provided in Table 1, course designers found that the following UX-related problem areas were easy to understand: *Computer playfulness, user interface appeal, information credibility* and *information challenge*. On the other hand, *image* and *subjective norm* were considered to be very intangible and not reasonable at all.

Assessing the Severity of UX-related Problem Areas (Q3)

According to Hertzum [9], the severity of a UX-related problem area "is an assessment of the amount of trouble and inconvenience users will experience as a result of a specific aspect of a system. Severity assessments are, however, also recommendations about the urgency of fixing problems". In the context of the current study, our concern was whether the feedback could facilitate the prioritization of UX-related problem. In other words, .the evaluation feedback was intended to persuade the course designers to fix problem areas of different levels of severity (here: ranging from 1.00 – 4.50) (cf. [14]).

Regarding the severity assessments of the given UX-related problem areas (Q3 in Table 2), course designers did not evaluate the construct *computer anxiety*. In addition to that, *subjective norm* and *image* were considered not to constitute severe UX-related problem areas. On the other hand, the remaining set of UX-related problem areas (*user interface appeal, information credibility, information challenge* and *computer playfulness*) was considered to constitute severe UX-related problem areas which should be addressed to improve the course for the forthcoming semester. However, in order to improve course designers' severity assessments, they required the feedback format to contain even more explicit students' severity rankings in order to understand how severe a problem was actually perceived by them.

Capability of the Feedback to Solve UX Problems (Q4)

Concerning the capability of the information provided in Table 1 for solving the UX-related problem areas, the course designers appreciated the way the information was presented ("the table does provide the causes and solutions of the problems"), and here especially *evaluators' suggestions* even though they were considered to be "not operative enough". Thus, the *item* was considered to give "an idea that there exist particular UX-related problem areas, in order to solve them the *evaluator's suggestions* are inevitable."

Course Designers' Intention to Solve UX Problems (Q5)

Concerning the course designers' intention to solve the UX-related problem areas illustrated in Table 1, the variation was relatively large (i.e. the column Q5 in Table 2, Mean = 3.67, SD = 1.53). Specifically, we computed the so-called impact ratios [28] per course designers:

$$\frac{\text{Number of problems committed to be fixed} * 100}{\text{Total number of problems found}}$$

The results range from one course designer showing an impact ratio of 14% ("The UX-related problem areas presented do not constitute real problems […] So why should I solve them?") to the other two course designers showing an impact ratio of 100% ("I'll try to tackle all problems so as to improve the course and contributing to good learner relations"; "Depending on the resources available one could tackle each of the UX-related problem areas illustrated, especially to foster students' *computer playfulness* and their perceived *image* of using the online course"). However, given that the re-design of the online course takes place within the upcoming semester break, the completed-to-date impact ratio is out of scope of this paper [28]:

$$\frac{\text{Number of problems committed receiving a fix} * 100}{\text{Total number of problems found}}$$

Persuasiveness of the UX Problems (Q6)

Two of the three course designers considered the following UX-related problem areas 1 to constitute no "fake problems" (see Table 1): *Information challenge*, *information credibility* and *computer playfulness*. On the contrary, *image* and *subjective norm* were considered not to constitute real UX-related problem areas. This was mainly due to the fact that course designer could not relate them to the course.

Ease of Use of the Feedback to Solve UX Problems (Q7)

In total, the feedback was considered to be not operative enough. One of the course designers remarked that "for instance, knowing that students perceive materials as incomplete does not help me which concrete information lacks, in which part and why?"

Impact on the Prioritization of UX Problems (Q8)

In a nutshell, the feedback format helped course designers to classify the subsequent UX-related problem areas as critical, namely *computer anxiety* and *information challenge*. The corresponding fixing plan was to formulate exercises more precise and understandable.

DISCUSSION

The present study provides a systematic evaluation of how course designers perceived the persuasiveness of a theoretically-grounded feedback format concerning students' UX with a university's online course.

In total, course designers' mean ratings across the eight questions in the questionnaire were all above 3.00 with standard deviations ranging between 0.00 (*usefulness of the feedback*) and 1.53 (*intention to solve UX-related problem areas*; *ease of use of the feedback to solve particular UX-related problem areas*). The large variations are due to the fact that one course designer did not perceive the UX-related problem areas listed in Table 1 as "real" problems (Q5), and they were found to be not operative enough (Q7).

In particular, the findings revealed that *computer anxiety* and *computer playfulness* were perceived as very useful (Q1). Furthermore, the feedback regarding *user interface appeal*, *information credibility*, *information challenge* and *computer playfulness* helped course designers to understand the corresponding UX-related problem areas (Q2), had an impact on the course designers' severity assessments of the corresponding problem areas (Q3), their intention to solve the UX-related problem areas revealed (Q5, + *subjective norm*, *image* and *computer anxiety*) as well as the persuasiveness of the UX-related problem areas (Q6) illustrated in Table 1. Moreover, *information challenge* and *computer anxiety* were considered to be the two most critical UX-related problem areas to be addressed within the forthcoming re-design of the online course (Q8). Finally, in order to solve particular UX-related problem areas as revealed by use of students' item ratings (see Table 1, column 5), course designers required the feedback format to contain *evaluator's suggestions* (Q4).

Regarding the research questions addressed in Introduction, we revisit them here with reference to the empirical findings gathered:

RQ1: Students had more problems in specifying UX-related problem areas based on TAM3-related UX items than those based on their ISSM-related counterparts (see the column Participants' qualitative comments in Table 1). In particular, students struggled in specifying TAM3-related UX problem areas such as *image* ("One does have a high profile when using the VLE?"), *computer anxiety* ("This question is pretty strange too!") and *computer playfulness* ("Playful? Do not know how to interpret and answer this question?"), whereas this was only the case for *information credibility* („I am not able to judge if he is a recognized source of information") regarding ISSM-anchored UX problem areas.

RQ2: As depicted in Table 2, course designers considered UX-related problem areas originated in the ISSM to be more persuasive than their TAM3-related counterparts. This may be mainly due to the fact that course designers did not perceive the "intangible" TAM3-anchored UX-related problem areas as relevant to their particular course (e.g. *subjective norm* and *image*).

RQ3: Course designers perceived UX-related problem areas anchored in TAM3 or the ISSM to be most persuasive in case evaluators' suggestions (see Table 1, column 7) were provided in addition to the other UX problem area-related contextual information provided in Table 1 (see Q1/4: "in order to solve the problem the evaluator's suggestions are inevitable"). No significant differences between TAM3- and ISSM-anchored UX-related problem areas were reported by the course designers.

IMPLICATIONS

The above mentioned results should generally provide a starting point for future research. In particular, future research efforts should focus on elaborating ways how to

formulate items related to TAM3 constructs such as *subjective norm* and *image* so that they may better relate to (course) designers' concrete work. Furthermore, course designers' severity assessments as well as their evaluation of the persuasiveness of the realness of such problem areas may be improved and facilitated by the use of more explicit students' severity ratings (i.e. underline the meaning of students' item ratings). The main benefit of further refining UX-related items anchored in theory-grounded constructs may be the improvement of the transparency and comparability of the corresponding research outcomes.

In addition, as the capability of the feedback to solve UX-related problem areas was considered to be limited due to its lack of information richness, future research work should investigate which potentially persuasive elements need to be included in a re-design proposal for fixing UX-related problems.

CONCLUSION

In this paper a systematic evaluation of how course designers' perceived persuasiveness of a theoretically grounded feedback format was carried out. Specifically, a problem list with corresponding redesign proposals of TAM3- and ISSM-anchored UX-related problem areas was identified and evaluated. Outcomes of this study will presumably stimulate future research on resolving UX problems. In particular, the combination of quantitative and qualitative data can not only gain better insights into issues but also support future (course) design and evaluation efforts that may contribute to students' positive UX while interacting with a VLE and online courses.

ACKNOWLEDGMENTS

We thank Anke Diederichsen for the review and pretest of the course designers' questionnaire, Christian Gasper and his research assistant for providing their persuasiveness with the feedback format.

REFERENCES

[1] Bevan, N. What is the difference between the purpose of usability and user experience evaluation methods, in *Proceedings of the Workshop UXEM'09* (Uppsala Sweden, 2009).

[2] DeLone, W.H., and McLean, E.R. Information systems success: The quest for the dependent variable. *Information Systems Research 3, 1* (1992), 60-95.

[3] DeLone, W.H., McLean, E.R. The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems 19, 4* (2003), 9-30.

[4] Fishbein, M., and Ajzen, I. *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA, 1975.

[5] Fu, F.-L., Chou, H.-G., Yu, S.-C. Activate interaction relationships between students' acceptance behavior and e-learning, in Dong, G., Lin, X. Wang, W., Yang, Y., Xu Yu, J. (eds.), Joint 9th Asia-Pacific Web Conference (APWeb 2007) and 8th International Conference on Web-Age Information Management (WAIM 2007), LNCS 4505, 670-677.

[6] Hassenzahl, M. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction, 19* (2004), 319-349.

[7] Hassenzahl, M., and Sandweg, N. From mental effort to perceived usability: Transforming experiences into summary assessments, in *Proceedings of CHI '04* (New York, NY, 2004), ACM Press, 1238-1286.

[8] Hassenzahl, M., and Tractinsky, N. User Experience: A research agenda. *Behavior & Information Technology 25, 2* (2006), 91-97.

[9] Hertzum, M. Problem prioritization in usability evaluation: from severity assessments toward impact on design. *International Journal of Human-Computer Interaction* 21, 2 (2006), 125-146.

[10] Hevner, A.R., March, S.T., Park, J., and Ram, S. Design science in information systems research, *MIS Quarterly 28, 1* (2004), 75-105.

[11] Hong, W., Thong, J.Y.L., Wong, W.-M., and TAM, K.-Y. Determinants of user acceptance of digital libraries: An empirical examination of individual differences and system characteristics. *Journal of Management Information Systems, 18, 3* (2001-2002), 97-124.

[12] Isleifsdottir, J., Larusdottir, M. Measuring the user experience of a task oriented software, in Effie L-C. Law, Nigel Bevan, Georgios Christou, Mark Springett, and Marta Lárusdóttir (eds.), *Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement '08* (Reykjavik Iceland, June 2008), 97-101.

[13] ISO 9241-110. Ergonomics of human-system interaction - Part 110: Dialogue principles (2006), ISO Copyright Office, Geneva, Switzerland.

[14] Law, E. L-C. Evaluating the downstream utility of user tests and examining the developer effect: a case study. *International Journal of Human-Computer Interaction 21, 2* (2006), 147-172.

[15] Law, E. L-C., Roto, V., Hassenzahl, M., Vermeeren, A., and Kort, J. Understanding, scoping and defining user experience, in *Proceedings of CHI '09* (Boston, USA, April 2009), 1-10.

[16] Law, E. L-C, and van Schaik, P. Modelling user experience: an agenda for research and practice. *Interacting with Computers Interaction* in press (2010), 1-10.

[17] Lee, G.T., Dahlan, N., Ramayah, T., Karia, N., and Hasmi Abu Hassan Asaari, M. Impact of interface characteristics on digital libraries usage. *Malaysian Online Journal of Instructional Technology 2, 1* (2005), 1-9.

[18] Lindgaard, G. *Usability Testing and System Evaluation: A Guide for Designing Useful Computer*

Systems. Chapman & Hall, London and New York, 1994.

[19] Martínez-Torres, M.R., Toral Marín, S.L., Barrero Garciá, F., Gallardo Váquez, S., Arias Oliva, M, Torres, T. A technological acceptance of e-learning tools used in practical and laboratory teaching, according to the European Higher Education Area. Behaviour & Information Technology 27, 6 (2008), 495-505.

[20] Moore, G.C., and Benbasat, I. Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research 2* (1991), 192-222.

[21] Mueller, D., and Strohmeier. Design characteristics of virtual learning environments: An expert study, International Journal of Training and Development (2010), in press, 1-14.

[22] Mueller, D., and Strohmeier, S. Developing and validating a rigorous and relevant model of VLE success: A learner perspective, *Proceedings of ECTEL '10* (Barcelona Spain, September/October 2010), in press, 1-16.

[23] Mueller, D., and Zimmermann, V. A learner-centred design, implementation, and evaluation approach of learning environments to foster acceptance. *International Journal of Advanced Corporate Learning 2* (2009), 50-57.

[24] Nielsen, J. Heuristic evaluation, in J. Nielsen, and R.L. Mack (eds.), *Usability Inspection Methods*. Wiley, New York, 1994, 25-62.

[25] Norgaard, M., and Hornbaek, K. Exploring the value of usability *feedback formats. International Journal of Human-Computer Interaction 25, 1* (2009), 49-74.

[26] Oxford Dictionaries. Available at: http://oxforddictionaries.com/view/entry/m_en_gb0290410#m_en_gb0290410.

[27] Petter, S., DeLone W.H., and McLean, E.R. Measuring information systems success: Models, dimensions, measures, and interrelationships. *European Journal of Information Systems 17* (2008), 236-263.

[28] Sawyer, P., Flanders, A., and Wixon, D. Making a difference: The impact of inspections, in *Proceedings of CHI '96* (Vancouver BC, April 1996), ACM Press, 376-382.

[29] Seddon, P.B., Staples, S., Patnayakuni, R., Bowtell, M. Dimensions of information success. *Communication of the Association for Information Success 2* (1999), 2-39.

[30] Taylor, S., and Todd, P.A. Understanding information technology usage: A test of competing models. *Information Systems Research 6* (1995), 144-176.

[31] Venkatesh, V. Determinants of perceived ease of use: Integrating perceived behavioral control, computer anxiety and enjoyment into the technology acceptance model. *Information Systems Research 11* (2000), 342-365.

[32] Venkatesh, V., and Bala, H. Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences 39, 2* (2008), 273-315.

[33] Venkatesh, V., and Davis, F.D. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science 46* (2000), 186–204.

[34] Venkatesh, V., Morris, M., Davis, G., and Davis, F. User acceptance of information technology: Toward a unified view, MIS Quarterly 24, 3 (2003), 425-478.

[35] Webster, J., and Martocchio, J.J. Microcomputer playfulness: Development of a measure with workplace implications. *MIS Quarterly 16* (1992), 201-226.

[36] Wixom, B.H, and Todd, P.A. A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research 16, 1* (2005), 85-102.

A New Perspective on UX: The Indirect User Experience

Ole Andreas Alsos
Norwegian University of Science
and Technology
Trondheim, Norway
+47 915 44 825
ole.andreas.alsos@idi.ntnu.no

ABSTRACT
Computer systems are first and foremost designed for single users in traditional desktop situations. The needs of the indirect users, those who are not using the system directly but are yet affected by it, are often overlooked. Based on empirical findings from two usability evaluations in a hospital simulator with physicians and patient actors, this paper discusses how changes in user interface can affect both the user experience of the primary user and that of the indirect user. Finally, it discusses implications for designing for the indirect user

Categories and Subject Descriptors
H.5.2 [**User interfaces**]: User-centered design,

General Terms
Human Factors

Keywords
User Experience, UX, indirect user experience, CSCW, mobile

INTRODUCTION

Most ICT systems are primarily designed for *primary users*, users who are working independently in traditional desktop situations. User experience, as defined in ISO 9421-210 [1], is first and foremost associated with the primary user: "[It is] *a person's perceptions and responses that result from the use or anticipated use of a product, system or service*".

The needs of *indirect users*, users who are not using the system directly but yet affected by it, are rarely in the minds of the system designers and developers. This is normally unproblematic for traditional systems designed for single users. However, as collaborative and mobile systems are becoming more common, the use of such systems will have effects on people outside the sphere of the primary user. While most systems have clear roots in the needs of the primary users, there is often little or no focus on the indirect users.

In this paper we seek to define the *indirect user experience.* Further, drawing on usability evaluations of mobile devices used in a realistic hospital setting, we identify in what ways the design of the user interface affect the indirect user in that setting. We also discuss implications on how the user interface can be designed to accommodate the needs of the indirect user.

BACKGROUND

Traditionally, an end-user is considered as the person who directly interacts with an information system. However, end-users, as defined by Faulkner [2], can be (1) *direct users*, who use the system themselves, (2) *indirect users*, who ask other people to use the system on their behalf, (3) *remote users*, who do not use the system, but depend on the output, or (4) *support users*, who ensure that the system works for others, such as direct users.

We choose to use a simpler and more straightforward end-user categorization. We divide end-users into (1) *direct users* and (2) *indirect users*, where the first category includes primary users and all other stakeholders who directly interact with an information system. The latter includes end-users who are not directly interacting with the system and corresponds to Faulkner's [2] indirect and remote users.

A number of HCI publications refer to a 1997 draft of ISO 9241-11:1998 [3]. This version contained reference to indirect users: "[Satisfaction is] the comfort and acceptability of the work system to its users and *other people affected by its use*". However, this reference to indirect users was omitted in the final version (see [4] for an example).

METHODS

The empirical grounding for this position paper come from two simulation-based usability evaluations of mobile systems for hospitals [5,6]. Both evaluations were conducted in a simulated hospital environment with multiple users; real physicians and patient actors in the hospital beds.

In the first evaluation, we explored several ways of letting doctors use handheld devices together with bedside mounted patient terminals for viewing x-ray images together with the patient [5].

In the second evaluation we explored interaction techniques for a handheld medication system, one paper based and three mobile patient record systems [6].

In the two evaluations, both the physicians and the patients were interviewed about aspects concerning the user experience of the mobile devices. In this case, the patient is the indirect user.

RESULTS

Below some of the observations from the experiments related to aspects of the user experience for primary and indirect users are presented.

The new technology increased UX for the primary users: Although the physicians in general were confident and comfortable with the paper chart, they preferred using the mobile device. A number of functions and attributes, such as pocket size, error prevention and undo mechanisms, contributed positive to the UX.

Action transparency: When moving patient records from paper-based media to mobile technology we observed that the physicians' actions were less visible for patients, i.e. the indirect user. While it was easy for the patient to see whether the physician was adding, searching or obtaining information with the paper chart, all actions appeared similar with the mobile device. This was considered negative by the patients.

Nonverbal communication: The physical form factor of the paper chart allowed the physician to use it as a channel for nonverbal communication (i.e. signal that the consultation was ending by closing the chart). This was harder with the PDA, and was considered negative by the patients.

Doctor-patient dialogue: The user interface of the mobile device increased legibility and allowed the physicians to undo and prevent medication errors. On the other hand, the user interface had poor information overview and unfamiliar interaction techniques. This required much of the physicians' attention. According to patients, it affected the doctor-patient dialogue and decreased their satisfaction of the consultation.

Negative patient experience: In some of the design solutions of the first evaluation the doctor controlled the patient terminal through the PDA. While this was seen as a major benefit from the perspective of the physicians who could hide information on the PDA and display public information on the PDA, it was perceived as negative from the perspective of the patients. They perceived the PDA as a mystical thing and did not like that things were hidden for them.

Positive patient experience: In other design solutions the physicians controlled the system through the patient terminal. Unintentionally it allowed the patient to control the terminal. For the patient this was perceived as an improvement. For the physicians, however, it became harder to control the system because they had to bend over the patient to use it.

User interface complexity: In some versions of the first evaluation, the controls for changing information content were present on the patient terminal. The increased complexity of the GUI confused some patients. They rather preferred the versions where these controls were moved onto the PDA.

DISCUSSION

The findings from the evaluations gave new insight related to the user experience.

User experience is relevant for indirect users. The studies demonstrated, not surprisingly, that technology had an impact on the user experience of the physicians, who were the primary users. Further, our observations indicated that the system design also had an impact on the indirect users. The patients had some sort of user experience; they had strong perceptions and responses about the system, although they had not used the systems directly themselves.

Indirect user experience defined

Based on the findings that UX is relevant for indirect users, I attempt to define the indirect user experience based on the ISO 9241-210 definition [1]:

> **Indirect user experience** is defined as *a person's perceptions and responses that result from another user's use of a product, system or service.*

Having defined indirect user experience, we present a further analysis of our findings.

Firstly, the observations showed that even if UX was improved for physicians, it had in some cases negative effects for patients.

> *Improving the user experience for the primary user can have negative consequences for the indirect user.*

Second, we also observed that when the indirect user experience was improved, it sometimes created problems for the physician.

> *Improving the user experience for the indirect user can have negative consequences for the primary user.*

We consequently are faced with tradeoffs between the needs of the primary and indirect users.

Impact of social factors on indirect user experience

As shown, the indirect user experience was influenced by the user interface of the system. However, there may also be other factors affecting the indirect user experience. Firstly, the user experience may differ because they have different roles and interests in the situation. Second, they are in a social context where the indirect user is experiencing (at least) two things; the direct user's interactions with the system, and the direct user's social interactions.

Implications for design

Accommodating the needs of the indirect users is important. In the context of a ward round with a physician

and a number of patients, a positive indirect user experience can have positive effect on the doctor patient dialogue, which is important for the treatment and care of the patients [7]. In the context of business, for example a travel agent serving a traveler, an improved indirect user experience can have positive effect on the customer experience. In business, this often means returning customers and increased revenue [8].

Below we suggest some implications for design based on the findings from the evaluations:

Give system feedback to the indirect user: By increasing the action transparency (i.e. increase visibility of actions) or providing system feedback also to the indirect user, one can increase the indirect user experience.

Support non-verbal communication: Indirect user experience is correlated with the ability of the primary user to communicate with the indirect user. The system can hinder this communication, especially the non-verbal aspects when the system occupies the hands of the primary user. Therefore, the physical form factor of the device needs to afford nonverbal communication.

Use the language of the indirect user: By presenting the information for the primary user in the language of the indirect user, the primary user can be guided to use simpler terms and communicate on the same level as the indirect user (i.e. physicians use terms like "blood sugar level" instead of "glucose").

Provide a tailored GUI for the indirect user: If feasible and necessary, an additional device/GUI with information tailored for the indirect user should be provided. This will give the indirect users a version of the information where unnecessary complexity is trimmed away.

Implications for software development

Indirect user experience also has some consequences for how we develop software:

Design for the indirect user: Address the needs, and include the perspective of the indirect user into requirements. This can ensure positive user experiences for indirect users.

Evaluate with the indirect user: Indirect users should be present when the system is evaluated, and their opinions should be collected.

CONCLUSION

When designing information systems that have effects on people beside the primary user, the designer and requirements engineer must address the need of all types of end-users. This includes the needs of the *indirect user*, and implies that one has to design for the *indirect user experience*. Sometimes this implies that the designers deal with conflicting needs between the direct and indirect users.

REFERENCES

[1] ISO 9241-210:2009. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Organization for Standardization

[2] Faulkner, X. Usability Engineering. Palgrave Macmillan, 2000.

[3] ISO 9241-11:1998. Ergonomics of human system interaction - Part 11: Guidance on usability specification and measures. International Organization for Standardization

[4] O'Grady, M.J., O'Hare, G.M.P., Sas, C. Mobile agents for mobile tourists: a user evaluation of Gulliver's Genie. Interacting with Computers 17 (2005).

[5] Alsos, O.A., Svanæs, D. Interaction techniques for using handhelds and PCs together in a clinical setting. Proc. NordiCHI '06 (2006) 125-134.

[6] Alsos, O.A., Dabelow, B., Faxvaag, A. Doctors' Concerns of PDAs in the Ward Round Situation — Lessons from a Formative Simulation Study. Methods of Information in Medicine.

[7] Ong, L. M. L. de Haes, J. C. J. M. Hoos, A. M. Lammes, F. B., Doctor-patient communication: A review of the literature. Social Science \& Medicine, 1995. 40 (7): p. 903-918.

[8] Peppers, D., Rogers, M., Return on customer: creating maximum value from your scarcest resource. 2005. Currency

Learning from User Experience in the Design of a Playful Peer Feedback Tool

Vibeke Vold
Uni Digital, Uni Research
Allégaten 27
5020 Bergen, Norway
+47 90013289
Vibeke.Vold@uni.no

Barbara Wasson
Department of Information Science and Media Studies
University of Bergen, Norway
+47 55584120
Barbara.Wasson@uib.no

ABSTRACT
Today's youth thrive in informal participatory communities where they not only consume but also act as contributors or producers. In a participatory culture of learning, students' active contributions to their learning are stressed and peer assessment is considered as an important component. In this paper we investigate which user experiences should be supported in a playful peer feedback tool within a participatory culture of learning.

Keywords
User Experience (UX), participatory culture of learning, active learning, peer feedback.

INTRODUCTION

Today's youth participate in a variety of social media where they create and disseminate ideas or news, collaborate and connect with people. Examples of social media software are numerous and enable communication, collaboration, multimedia and entertainment through blogs, and social networking (Facebook, MySpace), wikis, Flickr, YouTube, Second Life, etc. Through their participation in these informal communities today's youth develop new media skills [28]. In their 2006 report on *Confronting the Challenges of a Participatory Culture* Jenkins et al. [15] identify 10 new skills—Play, performance, Simulation, Appropriation, Multitasking, Distributed Cognition, Collective Intelligence, Judgement, Transmedia Navigation, Negotiation, Networking—developed through collaboration and networking.

While there is an increasing view of learning as a participative activity in the learning community [17], schools and institutions have been slow to react to the emergence of this new participatory culture [15]. An important component in the design of learning environments is implementing this contemporary culture of learning [17]. Peer assessment is used as a means to empower students and peers by enabling students to take charge of their learning and become active learners who take responsibility for, and manage, their own learning [1].

In our work on peer assessment we are inspired by the ease and playfulness with which participants interact and give each other feedback in participatory environments . We try to harness this and draw on the new media skills the students are developing in our design of a playful peer assessment tool. A review of current research in the field of peer assessment and feedback, and observation of user experiences during a field trial, are used to inform the design of a playful peer feedback tool in a participatory culture of learning.

LEARNING IN SCY

The EU 7th framework SCY (Science Created by You; www.scynet.eu) project addresses learning in science offering learners a learning experience based on real life, challenging assignments [3]. In SCY-Lab (the SCY learning environment) learners work individually and collaboratively on "Missions" which are guided by socio-scientific questions such as "How can we design a climate-friendly house?" [3]. Learners have to gather and process information, design and conduct experiments, make interpretations and abstractions, communicate their conclusion or, in other words, engage in processes of active learning, based on inquiry, knowledge building, and learning by design [28].

SCY uses a pedagogical approach that centres around products called "emerging learning objects" (ELOs) that are created by learners [3]. The ELOs, such as a CO_2-friendly house design or a concept map, are the vehicles for gaining an understanding of the general science skills, social and presentations skills, and domain concepts the student has developed [28]. Thus assessment in SCY is centred on these ELOs.

PEER ASSESSMENT IN A PARTICI-PATORY CULTURE OF LEARNING

Peer assessment enables students to take charge of their learning, and become active learners who could take responsibility for, and manage, their own learning [1, 2, 4, 26, 30]. For example, it enables students to learn to assess and to develop assessment skills, either when they enact peer assessment themselves or when receiving an assessment from their peers, and at the same time, it enhances students' learning through knowledge diffusion and exchange of ideas, even when they are incorrect [28].

Peer assessment has also been found to motivate students to engage in the learning process [22]. Research on students' views about peer assessment has shown that students are motivated by the fact that they want to impress their peers [11] and by the fact that peer assessment is productive. It makes them think, learn more, be critical, and be structured [2, 6, 23, 24]. In addition, peer assessment introduces the students to the perspective that the focus of instruction is not only on the end product(s) but also on the process, and it highlights the value of collaboration (e.g., social interactions, trust in others; [19]. Peer feedback is a form of peer assessment where peers give opinions, suggestions for improvements, ideas, etc. to each other. It has been found that students are more willing to accept feedback given in "student-speak" and students may be more willing to accept feedback from peers [7]. It has also been emphasized that the accuracy of the peer feedback may not be that crucial [9] and that the consequence of variety of accuracy in peer feedback might just be a benefit [26].

A PLAYFUL PEER FEEDBACK TOOL

In our work in the SCY project we are focused on providing "playful" peer assessment possibilities in a science learning environment and in this manner empower the users to become active learners who take responsibility for, and manage, their own learning [29]. The tool is "playful" because it is lightweight and designed to take advantage of new media skills.

User experience

Over the last decade "user experience" became the buzzword in the field of human-computer interaction (HCI) and interaction design [14]. It has become a catchphrase, calling for a holistic perspective and an enrichment of traditional quality models with non-utilitarian concepts, such as fun [18, 5], joy [10], pleasure [16], hedonic value [12] or ludic value [8].

Good user experience (UX) is the goal of most product development projects today [20]. Hassenzahl [13] argued that future HCI must be concerned about the pragmatic aspects of interactive products (i.e. fit to behavioral goals) as well as about hedonic aspect, such as stimulation (i.e.) personal growth, an increase of knowledge and skills), identification (i.e. self-expression, interaction with relevant others) and evocation (i.e. self-maintenance, memories). Focus on the positive aspects of technology use has also been a trend in psychology [21] and within UX this idea has been adopted outlining one of HCI's main objectives to contribute to our quality of life by designing for pleasure (by creating outstanding quality experiences) rather than for absence of pain (or preventing usability problems) [14].

How to sustain a good user experience?

Many UX researchers argue that good UX comes from the value and meaning of the product concept itself [20]. In order to select the right concept, we need to evaluate the concept ideas, the potential value of the concept idea itself (experiential evaluation) and how the concept idea would fit into participant's own context of living [20]. According to Roto *et al.* [20] the value of the anticipated interaction outcome can be evaluated even thought there is no user interface or interaction design available.

In this paper we address: How can we design a playful peer feedback tool to sustain good user experiences?

FIELD RESEARCH

During a March 2010 field trial of the SCY Mission "Create a CO_2 Friendly House" we observed how peers interact and give each other feedback. The trial was arranged at Sandvika Upper Secondary School in Oslo. It ran for 20 hours, divided over 4 successive Wednesdays, 5 hours each day.

Participants

Three science classes of approximately 30 first year high school students (16-17 years old) were introduced to the SCY project and volunteers for the 4-week field trial were solicited. A selection of 20 students from the volunteers across these classes was chosen to participate. The 3 teachers divided the students in 4 person design teams, each of which chose their own name:

- BioNorway (3 girls and 1 boy)
- New energy (3 boys and 1 girl)
- Power puff (4 girls)
- PikenesJens (2 girls and 2 boys)
- ThumbsUp (2 girls and 2 boys)

Learning Environment

The learning environment comprised SCY-Lab (with its resources and tools) Google search engine, Google SketchUp (for 3D drawings), PowerPoint and Word. No feedback tool was available in SCY-Lab; feedback was given spontaneously and orally within and between groups Figure 1 shows a student working with SCYSimulation in SCY-Lab.

Figure 1. Student working with SCYSimulation in SCY-Lab

The Student Mission and Tasks Given

The Mission challenge given to the students, "Your job is to design a CO_2 friendly house", included 9 tasks:

1. Create one concept map where you explain the importance of reducing global CO_2-levels.
2. Create one concept map where you brainstorm on the design aspects of a CO_2 -friendly house.
3. Make an initial plan on how your design group will proceed with the tasks to ensure a successful project.
4. Become an expert in one of the four fields:
    a. Production of energy,
    b. Laws of energy,
    c. Solar cells and solar thermal collectors, and
    d. Heat pumps.
5. Experts present their work in their original design groups.
6. Revise the initial plan.
7. Design, build and analyze your CO_2 friendly house using different tools that will be provided for you.
8. Write a report for the mayor of your town.
9. Present your group's findings in front of your classmates.

Data Collection

Empirical data, collected during the field trial through observations, videos, and data recordings, included: field notes, video recordings, reports, power point presentations and the collection of ELOs.

Analysis for Assessment Design

During the field trial we were interested in the following questions: Are the students active and take initiative in their own learning process? Do they look at each others ELOs and engage in peer interaction? Do they give feedback? Do they need any support to share and give feedback on each other's ELOs?

Thus the analysis of the empirical data focused on whether the students:
1) shared their ELOs
2) asked questions or presented an argument
3) gave feedback to one another
4) took the feedback into consideration
and the implications of these for the design of a feedback tool.

Episode 1:

Student Jens looked at another team's house design on their screen and asked a question. The other student, Magnus, pointed at their ELO (see Figure 2) on his computer screen.

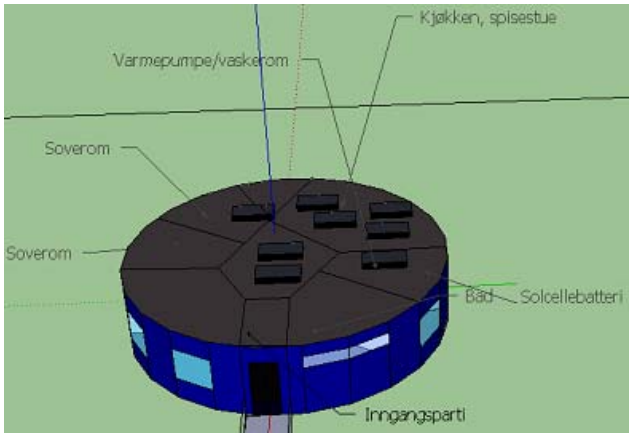

Figure 2. Team NewEnergy House Design ELO

Excerpt 1 (from Field Notes):

Jens (PikenesJens): *Do you have a CO_2 reason for building a round house?*

Magnus (NewEnergy): *We have chosen to design a round house with one floor. We did this to save area and by this also energy. Because the smaller square footage of exterior walls we don't need to insulate as much. We also chose to only use one floor in the house. In this manner we don't have the problem that the heat rises to the 2nd floor and we get an even heat throughout the whole house.*

Excerpt 1 shows the how a student question "Do you have a CO_2 reason for building a round house?" triggered a discussion about why Team New Energy made a circular house.

The relevance of this for the design of SCY assessment is that:
1) This dialogue should be supported by a SCYFeedback tool
2) The content of the dialogue illustrates that a) Jens can ask a question (skill: formulate questions) and b) Magnus can explain and argue for their choice of design (skill: argumentation/reasoning). This shows some of the skills that the teacher will look for in a summative evaluation.

Episode 2:

Student Jens looked at other team's house simulation in SCY-Lab (see Figure 3) on his own computer and got a reply from the teacher.

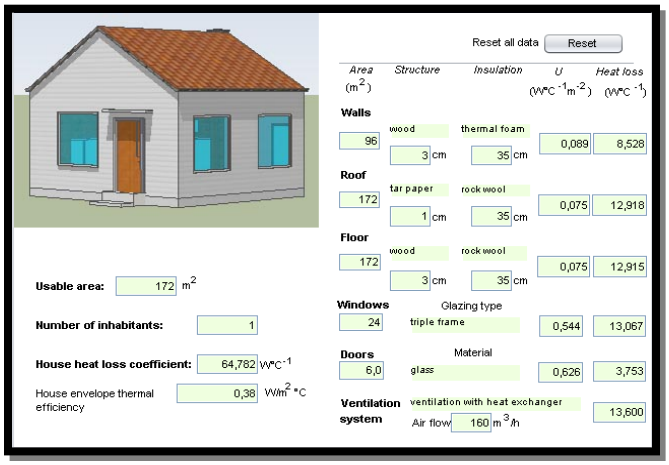

Figure 3. Team New Energy house simulation

Excerpt 2 (from Field Notes):

Jens (PikenesJens): *How can the walls have less surface area (96 m^2) than the floor and roof (both 172 m^2)?*

Teacher: *You have to use the formula for calculating the surface area for circles instead of rectangles.*

Jens (PikenesJens: *What is the forumula?*

Teacher: $A = \pi r^2$

Jens (PikenesJens)*: I have now calculated and I think that their answer is correct. The walls do have less surface area when using a circle than a rectangle!!!*

Teacher: *Laughing. Yes that is correct. You did not expect that did you?*

Jens (PikenesJens)*: No, humm well then I guess that I have understood something new.*

Excerpt 2 shows how a student question "How can the walls have less surface area (96 m^2) than the floor and roof (both 172 m^2)?" triggered a discussion between the student and the teacher.

The relevance of this for the design of SCY peer feedback is that:

1) This dialogue should be supported by the SCYFeedback tool. The peers designing the round house could just as well as the teacher help the students with information about how to calculate the surface area of a circle.
2) The content of the dialogue shows that Jens found that Team New Energy correctly had used the formula and calculated the area and volume of a circle (mathematics domain). It is also plausible that after the communication with the teacher Jens also has gained this skill in geometry of calculating area and volume from complex shapes.

Episode 3

Students' sharing of house simulations generates discussion around the elements in the data simulation of a CO_2 friendly house. Figure 4 shows the house simulation of team New Energy.

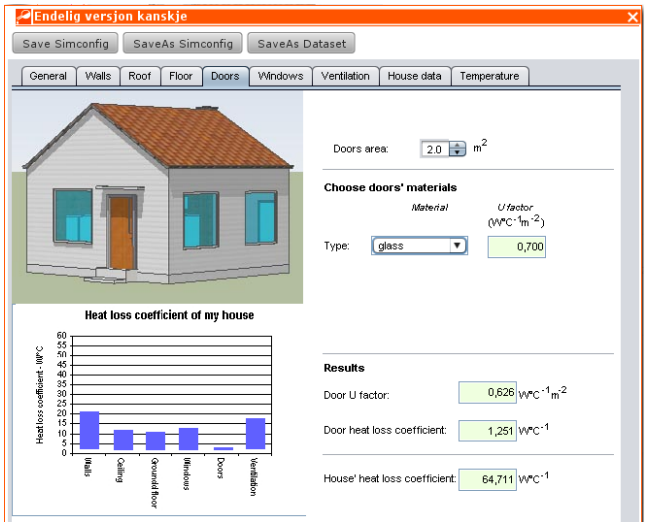

Figure 4. Team New Energy house simulation showing heat loss coefficient of their house

Excerpt 3 (from Field Notes)

Jens (PikenesJens): *Wow! Your graph bar for the door is very small compared to ours! The door area is 2 m^2 and the doors material is glass. How many m^2 does a door need to be? Is 2 m^2 enough? Is glass door better than wood?*

Teacher*: I would think that wood is better isolation material than glass.*

Jens (PikenesJens): *I have checked and you get better values for glass than for door. But the glass is triple!!!*

Teacher: *Ok that might explain it. Tipple glass door might provide better isolation than a single wood door.*

Jens (PikenesJens): *But is 2 m^2 enough for the door?*

Teacher: *How big is the door into the classroom? And how big are you?*

Jens (PikenesJens): *Checking the classroom door and walking through it. I do not think that it is more than 2 m^2. Great then I can reduce the door sixe and get a better heat coefficient. I will also experiment will various door materials.*

Excerpt 3 shows that the student Jens displays general science skills such as being able to visualize, interpret and make judgements about data. By investigating the simulation of another team and comparing this with their own a student gains experience in interpreting data and in investigating how the house simulation variables are related to the overall heat transfer coefficient. The application of the concept of overall heat transfer coefficient with the transfer of heat is a skill within Physics and Thermodynamics. The student discussion and application

of this concept in their house simulation model could demonstrate that they have gained this skill.

The skills of interpreting another team's (Team ThumbsUp) house simulation proved to be useful for Jens (Team PikenesJens) as he got a new perspective on how low the heat loss coefficient for the door could be. Based on the comparison of the two teams' simulation model and feedback from the other student Team ThumbsUp changed their values and managed to reduce the heat loss coefficient for their door.

The relevance of this for the design of the SCY assessment is that:
1) The commenting and questioning of a student made ELO could be supported by the SCYFeedback tool. The students in team New Energy might just as well as the teacher answer questions related to their simulation ELO and the choices behind their selection of values.
2) The ELO sharing led to changes in student ELO and the discussion shows that the student displays skills like for example being able to visualize, interpret and make judgements about data.

Episode 4

Students' presentation of their house design gave a good opportunity for peer feedback in a plenum. Figure 5 shows the students in Team New Energy presenting their use of isolation in their house design.

Figure 5. Team New energy presenting their house design

Excerpt 4 (from Field notes):

Team ThumbsUp*: We see that you have chosen tar paper for roof but is that an environmental friendly material?*

Team New Energy: *It is perhaps not the most environmental friendly, but it is very isolating and thus we do not have to use too much electricity to heat the house.*

Team ThumbsUp*: We think that you should avoid using a material that is not environmental friendly.*

Excerpt 3 shows that Team ThumbsUp is questioning the environmental friendliness of their choice of tarpaper as one of the roof materials. Team ThumbsUp and New Energy discuss if a material that is not environmental friendly can be used. The discussion and peer feedback could be supportive for the students in gaining general science skills such as being able to reflect on one's own knowledge and interpret data.

The relevance of this for the design of SCY assessment is that:
1) This student dialogue should be supported by the SCYFeedback tool.

2) The student questions would then be documented and the teacher could look back at the student dialogue when assessing the student skills.

CONCLUSIONS AND SUMMARY

In this paper we have explored which user experiences should be designed for and supported by a playful peer feedback tool within a participatory culture of learning. Today's youth participate in a variety of social media and develop new skills (e.g., play, simulation, judgement, multitasking). Within learning research the view of learning as a participatory activity where the students themselves participate actively in the learning community has been increasing. Peer assessment has been suggested as a method to be used to empower students to take charge of and manage their own learning.

UX researchers argue that good UX comes from the value and meaning of the product itself. The concept of participatory peer feedback has been further investigated in a school setting with the SCY-Lab learning environment and "Create a CO_2 friendly house" Mission in order to see if the concept idea would fit into participant's own context of learning.

The field study showed that:
- students were looking at each others products (ELOs) and took initiative by asking each other questions
- students naturally engaged in peer feedback dialogues
- students were able to make judgements about other students ELOs and use this to further develop their own skills
- the students seemed to be comfortable with switching between working on own ELOs and investigating other students ELOs
- the students seems to be motivated by playing with other students simulations
- students need support to communicate with each other and give each other feedback on ELOs
- students showed skills in their discussions (e.g. collaboration, formulate questions, argumentation, reasoning, mathematical calculation, judgement, simulation)

The idea of creating a good user experience and also cultivate the students as active learners with a peer based assessment tool seems promising. Findings show that students act, take initiative and they also seem to take pleasure in sharing their products (ELOs) and engaging in

peer discussions. Students do not seem to need instructions and guidance as this playfulness falls naturally for them. However, a need for a means to link peer feedback to ELOs was identified. The goal for the design of the playful SCYFeedback tool for peer assessment should be to facilitate student sharing of ELOs together with opportunities for student feedback on the ELOs. The tool should lay the foundation for a good user experience where student themselves can engage in ELO sharing and take charge of having fun and creating their own pleasure.

Figures 6 and 7 show screenshots of how the ELO display and linking of peer feedback comments to an ELO could be facilitated. Figure 6 shows the ELO Gallery where the students can find published ELOs while Figure 7 shows an ELO and how students could give peer feedback and score the ELO.

Figure 6. ELO Gallery showing students published ELOs

Figure 7. ELO feedback screen with comment and score field

ACKNOWLEDGMENTS

This study was conducted in the context of Science Created by You (SCY), which is funded by the European Community under the Information and Communication Technologies (ICT) theme of the 7th Framework Programme for R&D (Grant agreement 212814). This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. We like to acknowledge the contributions of the large list of people working on SCY to the work presented here.

REFERENCES

[1] Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working Inside the Black Box, assessment for learning in the classroom*. Kings College: London

[2] Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.

[3] de Jong et al. (2009) Learning by creating and exchanging objects: the SCY experience. Special issue on learning objects, *British Journal of Educational Technology*.

[4] Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer-, and co-assessment in higher education: A review. *Studies in Higher Education*, 24, 331–350.

[5] Draper, S.W. (1999). Analysing fun as a candidate software requirement. Personal Technology, 3(1),1-6.

[6] Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International*, 32, 175–187.

[7] Frost, J. & Turner, T. (Eds.) (2005) *Learning to Teach Science in the Secondary School*, Second Edition. Routledge Falmer, London.

[8] Gaver, W.W., & Martin, H. (2000). Alternatives. Exploring Information Appliances through Conceptual Design Proposals. In Proceedings of the CHI 2000 Conference on Human Factors in Computing (pp.209-216). New York: ACM, Addison-Welsley.

[9] Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2009). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, doi:10.1016/j.learninstruc.2009.08.007.

[10] Glass, B. (1997). Swept away in a sea of evolution: new challenges and opportunities for usability professionals. In R. Liskowsky, B.M. Velichkovsky, & W. Wünschmann (Eds.), Software-Ergronomie '97 *Usability Engineering: Intertation von Mensch-Computer-Interaktion und Software Entwicklung* (pp. 17-26). Stuttgart: B.G. Teubner.

[11] Hanrahan, S. J. & Isaacs, G. (2001). Assessing self- and peer-assessment: the students' views. *Higher Education Research and Development*, 20, 53-70

[12] Hassenzahl, M. (2002). The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction, 13, 479-497.

[13] Hassenzahl, M. (2003) The thing and I: understanding the relationship between user and product. In Blythe, C.Overbeeke, A.F. Monk, & P.C. Wright (Eds.), Funology: From Usability to Enjoyment (pp. 31-42). Dordrecht: Kluwer.

[14] Hazzenzahl, M. & Tractinsky, N. (2010). User experience – a research agenda. *Behaviour & Information Technolog*, 25(2), 91-97.

[15] Jenkins, H., Clinton, K., Purushotma, R., Robison, A.J. & and Weigel, M. (2006). Confronting the Challenges of Participatory Culture: Media Education of the 21st Century. Chicago: The MacArthur Foundation.

[16] Jordan, P. (2000). Designing Pleasurable Products: An Introduction To The New Human Factors. London, New York: Taylor & Francis

[17] Kollar, I. & Fischer, F. (2009). Commentary: peer assessment as collaborative learning: a cognitive perspective. *Learning and Instruction*.doi:10.1016/j.learninstruc.2009.08.002

[18] Monk, A.F. & Frohlich, D. (1999) Computers and Fun. *Personal technology*, 3, 91.

[19] Noonan, B. & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research and Evaluation*, 10(17).

[20] Roto, V,, Rantavuo, H. & Väänänen-Vainio-Mattila, K. (2009). Evaluating user experience of early product concepts. International Conference on Designing Pleasurable Products and Interfaces, DPPI09, 13-16 October 2009, Compiegne University of Technology, Compiegne, France.

[21] Seligman, M.E.P. & Csikszentmihalyi, M. (2000), Positive Psychology An Introduction, *American Psychologist*, 55, 5-14.

[22] Sluijsmans, D. (2002). Establishing learning effects with integrated peer assessment tasks. The Higher Education Academy. Retrieved from http://www.heacademy.ac.uk/resources/detail/id437 establishing learning effects

[23] Stefani, L. (1992). Comparison of collaborative, self, peer and tutor assessment in a biochemical practical. *Biochemical Education*, 20(3), 148- 15 1.

[24] Stefani, L. A. J. (1994). Peer, self and tutor assessment: relative reliabilities. *Studies in Higher Education*, 19, 69-75.

[25] Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249–276.

[26] Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimizing new modes of assessment: In search of qualities and standards*. Dordrecht/Boston/London: Kluwer Academic Publishers.

[27] van Zundert, M., Sluijsmans, D. & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270-279.

[28] Vold, V, Wasson, B & de Jong, T. (forthcoming) Assessing Emerging Learning Objects: ePortfolios and Peer Assessment, in *Digital Assessment of 21st Inquery skills* Routledge's new 'Psychology in Education' book.

[29] Wasson, B. & Vold, V. (forthcoming 2010). Leveraging New Media Skills for Peer Feedback in Collaborative Inquiry Learning, *First Nordic Symposium on Technology-enhanced learning (TEL)*, NORDITEL 2010.

[30] Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15, 179-200.

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

COST is supported by the EU RTD Framework Programme