

User Experience Evaluation Methods in Academic and Industrial Contexts

Virpi Roto^{1,2}, Marianna Obrist³, Kaisa Väänänen-Vainio-Mattila^{1,2}

¹ Tampere University of Technology, Human-Centered Technology, Korkeakoulunkatu 6, 33720 Tampere, Finland. [virpi.roto, kaisa.vaananen-vainio-mattila]@tut.fi

² Nokia Research Center, P.O.Box 407, 00045 Nokia Group, Finland. virpi.roto@nokia.com

³ ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria. marianna.obrist@sbg.ac.at

ABSTRACT

In this paper, we investigate 30 user experience (UX) evaluation methods that were collected during a special interest group session at the CHI2009 Conference. We present a categorization of the collected UX evaluation methods and discuss the range of methods from both academic and industrial perspectives.

Keywords

User experience, Evaluation methods

INTRODUCTION & BACKGROUND

As particular industry sectors mature, usability and technical reliability of products is taken for granted and users start to look for products that provide engaging user experience (UX). Although the term user experience originated from industry and is a widely used term also in academia, the tools for managing UX in product development are still inadequate. UX evaluation methods play a key role in ensuring that product development is going to the right direction.

Many methods exist for doing traditional usability evaluations, but user experience (UX) evaluation differs clearly from usability evaluation. Whereas usability emphasizes effectiveness and efficiency [5], UX includes hedonic characteristics in addition to the pragmatic ones [6], and is thus subjective [3,12]. Therefore, UX cannot be evaluated with stopwatches or logging. The objective measures such as task execution time and the number of clicks or errors are not valid measures for UX, but we need to understand how the user feels about the system. User's motivation and expectations affect the experience more than in traditional usability [13].

User experience is also very context-dependent [12], so the experience with the same design in different circumstances is often very different. This means that UX evaluation should not be conducted just by observing user's task completion in a laboratory test, but it needs to take into account a broader set of factors.

In order to gain more insights on globally used UX evaluation methods, we organized a Special Interest Group (SIG) session on "User Experience Evaluation – Do you know which Method to use?" at the CHI'09 conference [16]. The objective of the CHI'09 SIG was to gather evaluation methods that provide information on how users feel about using a designed system, in addition to the efficiency and effectiveness of using the system. This is a common requirement for all UX evaluation methods, but various kinds of methods need to be designed for different cases. Specific, various methods are often needed for academic and industrial contexts and a toolkit of methods would help in finding the proper method for each case. In this paper, we present and discuss the results from the SIG session in particular with the goal to answer a research question "What are the recently used and known UX evaluation methods in industry and academia?"

DATA COLLECTION PROCEDURE

The data on UX evaluation methods were collected during a 1.5 hour SIG session at CHI'09. The audience included about 35 conference participants, equally representing academia and industry. The goal of the session was to collect UX evaluation methods used or known in industry and academia. In the opening plenary, we explained what we mean by UX evaluation, emphasizing the hedonic/emotional nature of UX in addition to the primarily pragmatic nature of usability. We also highlighted the temporal aspects related to UX. We presented three examples of UX evaluation methods with our pre-defined method template: psycho-physiological measurements [9], Experience Sampling Method [11], and expert evaluation using UX heuristics (based on [16]). The template that we used for collecting the data from the participants contained the following information:

- Method name
- Description
- Advantages
- Disadvantages
- Participants (who is evaluating the system?): UX experts, Invited users in lab, Invited users on the field, Groups of users, Random sample (on street), Other

- Product development phase: Early concepting, Non-functional prototypes, Functional prototypes, Ready products
- What kind of experience is studied: Momentary (e.g. emotion), Use case (task, episode), Long-term (relationship with the product)
- Collected data type: Quantitative, Qualitative
- Domain-specific or general: General, For Web services, For PC software, For mobile software, For hardware designs, Other
- Resources in evaluation: Needs trained moderator, Not much training needed, Needs special equipment, Can be done remotely, Typical effort in man days (min-max)

The participants could fill in more than one method template. They could check multiple choices and we asked them also to mark whether the reported method comes from industry or academia.

The participants formed groups of 4-6 people to enable discussion on the methods that the group members knew or had been using. For sharing the collection of methods in the plenary we first had a quick presentation round of the most important UX method per group and then put all collected method templates on an interactive wall. The SIG participants interacted with the other participants via post-it notes attached to the method descriptions. After the session, we transcribed the forms and distributed them to the participants for enabling further clarifications.

RESULTS FROM THE SIG

In the SIG, 33 UX evaluation methods were identified. Three methods were clearly investigating merely usability or non-experiential aspects of the system, which left us 30 methods for further analysis. Half of the methods were reported to be academic and the other half industrial. 15 methods could provide means for evaluating also the experiential aspects, but the method was either not primarily foreseen for producing experiential data or we could not tell this from the given method description. We found 15 of the collected methods clearly tackling the experiential aspect of the evaluated system, providing insights about emotions, value, social interaction, brand experience, or other experiential aspects. We took the set of 30 methods into further analysis.

There are many possible ways to categorize the collected methods, and the different categorizations help UX evaluators to pick the right method for the purpose. In this paper, we categorize the collected methods along their applicability for lab tests, field studies, online surveys, or expert evaluations without actual users. This high level selection of a method category is typically done before choosing a specific evaluation method, so we hope our categorization helps UX evaluators to pick the right method for the need. The primary source for the data categorization was the “Participants” field in the used template (see the previous section), but during the analysis, we noticed that the provided options on the template were not fully covering all categories of the reported methods, so we

slightly modified the categories. We now have two different categories for field studies: one for short studies where an evaluation session is conducted in the real context, and the second for longitudinal field studies where the participants are using the system under evaluation for a longer period than the evaluation session(s). We also added a category called Mixed methods for methods which emphasized using several different methods for collecting rich user data. We used the information from “Can be done remotely” field (in Resources section) with the “Random sample (on street)” field, and created a new category called Surveys. The seven identified categories and the number of methods in each category are described in Table 1. Note that one method may be applicable for several categories.

Lab studies with individuals	11
Lab studies with groups	1
Field studies (short, e.g. observation)	13
Field studies (longitudinal)	8
Surveys (e.g. online)	2
Expert evaluations	2
Mixed methods	6

Table 1: Categorization of the collected methods by the type of participants in the UX evaluation

As can be seen from Table 1, most of the presented methods investigate UX during evaluation sessions where an invited participant is observed, interviewed, or self-reporting the experience. Each method category is described in more detail in the following sections including also concrete method examples.

Lab Studies

Lab studies have been highly popular for usability evaluations due to their efficiency and applicability for early testing of immature prototypes. In a traditional usability test, invited participants are given a task to carry out with one or several user interface designs, and they think aloud while doing the task. The analyst observes their actions and aims to understand users’ mental models in order to spot and fix usability problems.

Lab studies are very much needed also for evaluating UX in the early phases of product development. Three collected methods aimed to collect experiential insights during a usability evaluation, for example, by paying special attention to experiential aspects of user’s expressions while thinking aloud. This kind of extended usability test is the easiest way to extend the current evaluation routines to the experiential aspects. As noted in some of the templates, the extended usability test sessions may reveal more experiential findings if the sessions are organized in a real context of use. Altogether 6 methods in our collection were marked as applicable for lab and real context tests likewise.

Some of the methods are applicable for lab tests only, such as Tracking Realtime User Experience (TRUE) method [9] presented in our SIG session. Lab-only methods include psycho-physiological measurements [7,14] and other methods that require careful equipment setup and/or a controlled environment.

Only two UX methods were used for investigating a group or participants at once, instead of one participant at a time. Both of these methods were used for investigating social interactions. Focus group method [15] was in our methods collection as well, but the reported focus group method was targeted for usability testing and we omitted it from the analysis. It is still unclear if focus groups could be used for experiential evaluations, since personal experiences may not be revealed in an arranged group session.

Field Studies

Since UX is context-dependent, it is often recommended to examine it in real life situations whenever the circumstances allow it (see a detailed overview on in-situ methods in [1 or 8]). We were happy to see 21 field study evaluation methods, 8 of which investigate UX during an extended period of time, and the rest were either for organizing a test session of the prototype in real context or for techniques to observe and interview participants in real context. Many of the methods were reusing the exploratory user research methods, such as ethnography, for evaluating a system on the field. It is sometimes hard to make a difference between exploratory user research (to understand users' lives) and UX evaluation (to understand how the evaluated system fits into users' lives). We class a method as an evaluation method if the participants are using a certain selected system during the study and the focus of the study is to understand how the participants interact with it. If there is no predefined system to be investigated, the user study is of explorative rather than evaluative nature.

Surveys

Surveys can help developers to get feedback from real users within a short time frame. Online surveys are the most effective way to collect data from international audience and the number of participants in a survey can easily be much bigger than in any other methods. Online surveys are a natural extension to testing Website experiences, but if the tested system requires specific equipment to be delivered to participants, online surveys are more laborious to conduct.

In our collection of UX evaluation methods, two survey methods were reported. One was a full questionnaire, AttrakDiff™¹, the other was about using Emocards [3] to collect emotional data via a questionnaire. Emocards was the only method that was reported to be applicable for conducting UX evaluation even ad hoc on the street.

Expert Evaluation

Recruiting participants from the right target user group of the evaluated system is often time consuming and expensive. In an early phase of product development when the prototypes are still hard to use, it is quite common to have some usability experts to examine the prototype against usability heuristics [16]. It is beneficial to run expert evaluations always before the actual user study to

avoid basic usability problems to ruin the expensive user study.

It is challenging to establish expert evaluation methodology for UX, since experiences are very dependent on the person and the person's daily life. The field of UX is not yet mature enough to have an agreed set of UX heuristics to help expert evaluation. If each company and each project team could set UX targets that could be used as heuristics in expert evaluations, the heuristics would help to verify that the development is going to the right direction. If UX experts have conducted a lot of user studies, they have more insights on how the target user group will probably experience this kind of a system.

In our collection of UX evaluation methods, the most clearly targeted method for UX experts was about using a heuristics matrix in the evaluation. There was another interesting method, Perspective-Based Inspection, where the participants were asked to pay attention to one specific experiential aspect, such as fun, aesthetics, or comfort. In both cases, one needs to define the attributes that one needs to pay special attention to, and UX experts could be the evaluators in each case.

Mixed Methods

Six reported methods were pointing out the importance of using several different methods to collect rich user data. For example, it is beneficial to combine objective observation data with the system logging data and user's subjective insights from interviews or questionnaires. The combination of user observations followed by interviews was mentioned six times, which is interesting. With usability, we know that user actions can be observed to collect data on usability, but how can we observe how users feel, i.e., observe the user experience? One good example was to video record children playing outdoors and to analyze their physical activity, social interaction, and focus of attention. Children were also interviewed and their opinions of the toys were combined with objective data collected from the video (see e.g. [18]). With children, who may not be able to explain their experiences, observation is a promising method to get data about what they are interested in. It is quite impossible, however, to understand UX and especially the reasons behind UX with plain observations, so mixing several methods is needed especially with observations.

METHODS FOR ACADEMIA AND INDUSTRY

The basic requirements for UX evaluation methods are at least partly different when it comes to applying them in industry versus academic context. The evaluation methods used in industry are hardly ever reported in public, so we were very happy to successfully collect exactly as many industrial methods as academic, 14 both (2 did not reveal the origin). In industry, especially in product development, the main requirements for UX evaluation methods is that they have to be lightweight (not require much resources), fast and relatively simple to use [19]. Qualitative methods are preferred in the early product development phases to provide constructive information about the product design.

¹ <http://www.attrakdiff.de/en/Home/>

For benchmarking and marketing purposes, light, quantitative measurement tools are needed. Examples of UX evaluation methods especially suited for the early phases of the industrial product development were lab study with mind maps, retrospective interview and contextual inquiry. Examples of the quantitative methods applicable for quick evaluations of prototypes were expert evaluation with heuristics matrix and AttrakDiff questionnaire.

On the academic side, the scientific rigor is much more important and thus a central requirement for the UX evaluation method. Often, quantitative results and validity are emphasized in the academic context, at least as an additional viewpoint to qualitative data analysis. Examples of academically valid methods were long-term pilot study, experience sampling triggered by events (e.g. [11]), and sensual evaluation instrument.

Naturally, there are also common characteristics of methods in both industry and academia: In the context of UX evaluation, the methods must include the experiential aspects (as discussed in the Introduction), not just usability or market research data. Also, the methods should preferably allow repeatable and comparative studies in an iterative manner. This is especially important in the hectic product development cycle in industry, but also in design research that needs effective evaluation tools for quick iterations. As the middle ground, industrial research sets requirements that can use a mixture of fast and light, and more long-term, scientifically rigorous methods.

CONCLUSIONS & FUTURE WORK

In this paper, we provide a step towards a clearer picture on what are the recently used and known UX evaluation methods in industry and academia. We investigated 30 UX evaluation methods collected during a 1.5hour SIG session at CHI'09 conference.

We presented a picture on used and known UX evaluation methods, but there is still a lack of a clear understanding of what characterizes a UX evaluation method compared to a usability method. An important step further will be reached when a common definition of UX will be available. Meantime, it is important to collect and extend the current set of UX evaluation methods on a global basis.

Future work includes further broadening the collection of UX evaluation methods in workshops in Interact'09 and DPPI'09 conferences, and by further investigating UX evaluation literature. We will also extend our analysis to include alternative method classifications. We will also deepen the analysis by investigating the needs for further development of UX evaluation methods.

ACKNOWLEDGMENTS

We thank the participants of the SIG CHI2009 for their active contribution to the session, the discussion and the feedback provided afterwards for improving and extending the list on user experience evaluation methods. Part of this work was supported by a grant from the Academy of Finland.

REFERENCES

1. Amaldi, P., Gill, S., G., Fields, B., Wong, W., (Eds.), (2005). Proceedings of In-Use, In-Situ: Extending Field Research Methods workshop. Available online: http://www.ics.heacademy.ac.uk/events/misc/inuse_proceedings.doc (accessed: 26.06. 2009).
2. Beck, E., Obrist, M., Bernhaupt, R., and Tscheligi, M. (2008). Instant Card Technique: How and Why to apply in User-Centered Design. In Proceedings PDC2008.
3. Desmet, P.M.A. (2002). Designing Emotions. Doctoral dissertation, Technical University of Delft.
4. ISO DIS 9241-210:2008. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (formerly known as 13407). International Standardization Organization (ISO). Switzerland.
5. ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability. International Standardization Organization (ISO). Switzerland.
6. Hassenzahl, M. 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction* 19, 4 (Dec. 2004), 319-349.
7. Hazlett, R., L., (2006). Measuring emotional valence during interactive experiences: boys at video game play, In: Proceedings of the SIGCHI CHI '06, ACM, New York, USA, pp. 1023-1026.
8. Fields, B., Amaldi, P., Wong, W., Gill, S., (2007). Editorial: In-use, In-situ: Extending Field Research Methods, In: *International Journal of Human Computer Interaction*, Vol. 22, No. (1) pp. 1-6.
9. Ganglbauer, E., Schrammel, J., Deutsch, S., Tscheligi, M. (2009). Applying Psychophysiological Methods for Measuring User Experience: Possibilities, Challenges, and Feasibility. Proc. User Experience Evaluation Methods in Product Development workshop. Uppsala, Sweden, August 25, 2009.
10. Kim, J. H., Gunn, D. V., Schuh, E., Phillips, B., Pagulayan, R. J., and Wixon, D. 2008. Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. In Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems (Florence, Italy, April 05 - 10, 2008). CHI '08. ACM, New York, NY, 443-452.
11. Larson, R. and Csikszentmihalyi, M. (1983). The experience sampling method. In H. T. Reiss (Ed.), *Naturalistic approaches to studying social interaction. New directions for methodology of social and behavioral sciences* (pp. 41-56). San Francisco: Jossey-Bass.
12. Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., Kort, J. (2009). Understanding, Scoping and Defining

- User eXperience: A Survey Approach. In Proc. CHI'09, ACM SIGCHI conference on Human Factors in Computing Systems.
13. Mäkelä, A., Fulton Suri, J. (2001) Supporting Users' Creativity: Design to Induce Pleasurable Experiences. Proceedings of the International Conference on Affective Human Factors Design, pp. 387-394.
 14. Mandryk, R., L., Inkpen, K., M., Calvert, T., W., (2006). Using psychophysiological techniques to measure user experience with entertainment technologies, In: Behaviour and Information Technology, Vol. 25, No. 2, (Special Issue on User Experience), pp. 141-158.
 15. Morgan, D. L., Krueger, R. A. (1993). When to use Focus Groups and Why? In: Morgan, D. L. (Ed.). *Successful Focus Groups*. Newbury Park.
 16. Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J., and Mack, R.L. (Eds.), *Usability Inspection Methods* (John Wiley & Sons, New York, NY).
 17. Obrist, M., Roto, V., and Väänänen-Vainio-Mattila, K. (2009). User Experience Evaluation – Do You Know Which Method to Use? Special Interest Group in CHI2009 Conference, Boston, USA, 5-9 April, 2009.
 18. Read, J. C. and MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In Proc. IDC '06, 81-88.
 19. Roto, V., Ketola, P., Huotari, S. (2008). User Experience Evaluation in Nokia. Now Let's Do It in Practice - User Experience Evaluation Methods in Product Development workshop in CHI'08, Florence, Italy.