A Systematic Review About User Experience Evaluation

Camila Loiola Brito Maia(✉) and Elizabeth Sucupira Furtado

University of Fortaleza (Unifor), Fortaleza, CE, Brazil
camila.maia@gmail.com, elizabethsfur@gmail.com

Abstract. The user experience related to a product can determine its success or failure. As a result, companies have invested in research to understand what the user feels when he or she uses a product, but it is not simple to get this information, and many approaches arose through the years. This article presents a systematic review on the user experience evaluation field, based on 25 studies. We found (1) that psychophysiological measures are not yet widely applied in the evaluation of the user experience; (2) researchers prefer qualitative approaches; (3) the evaluations are mostly manual (not real time, therefore), and using ready-made products; and (4) most studies used single measurement, not considering the use of the product over time. This scenario shows an opportunity for UX evaluation methods which consider the use of the product over time and provide real time results.

Keywords: User experience · UX · Evaluation · Measurement

1 Introduction

A product or service that gives good experience to user can offer competitive advantage to companies, since the product could be recommended by satisfied customer. According to Pine and Gilmore [1], a good experience occurs when the user engages in a way that creates a memorable event. This experience is personal, and exists just for that user who got emotional, physical, intellectual and, sometimes, spiritual engagement with a particular product. For companies, if a product or service provides good experiences to users, they can gain competitive advantage, once the satisfied client probably will recommend this product. For this reason, UX is vital to commercial success of the companies [2], and many of these companies already have a team of experts whose focus is to ensure a product that provides a good experience for the users [3].

However, to generate good experiences is necessary to understand how the user feels [4]. In usability tests we usually check efficiency and effectiveness of the product, but not how users feel. If we understand what the user feels while using the product, we can measure and continuously improve his experience. Following this idea, several studies related to the evaluation of user experience have been written and disseminated using qualitative and quantitative analysis.

A. Marcus (Ed.): DUXU 2016, Part I, LNCS 9746, pp. 445–455, 2016.
DOI: 10.1007/978-3-319-40409-7_42

In this way, the purpose of this article is to present results of a systematic review about evaluation of user experience, by investigating how this evaluation of user experience is performed, in terms of moment, used techniques and elements of user experience taken into account.

This paper is organized as follows: Sect. 2 contextualizes the research, Sect. 3 details the systematic review, Sect. 4 presents the analysis of results, and Sect. 5 presents the final considerations.

2 Contextualization

In order to understand and evaluate user experience is essential to collect data from the user. The majority of approaches typically rely on observation [5, 6], psychophysiological measures [7, 8], and questionnaires or surveys [9, 10].

However, it is difficult to measure user experience. First, because several factors influence the user experience, like feelings, culture and communicability [11]. Furthermore, the user experience is multidimensional [2], consisting of various elements, and there is still no consensus among authors about these elements of user experience [2, 3, 12, 13]. There are several proposed models, such as Hassenzahl [14] and Mahlke and Thuring [15], but it seems that the only consensus is that the user experience is composed of pragmatic aspects (utility and usability) and hedonic aspects (emotional) [16].

In addition, as a given product may have users with heterogeneous profiles [17] and each experience is individual [16, 18] the final user experience is likely to have a different value for each user.

Finally, it is not easy to define a criterion to measure user experience because it is subjective and holistic [16].

After measuring user experience, it is important to evaluate this experience, because it provides feedback to those who developed the product, and future improvements can be generated for the user based on his feedback [18]. Moreover, evaluation allows comparisons between different versions of the same product, between different user groups, and in different contexts [18].

According to Machapa and Greunen [19], given the complexity of UX, evaluating an experience and understanding how the user classifies this experience is not a simple task. To evaluate the user experience, we have to collect data related to it. The main instruments used to collect these data are questionnaires and surveys, psychophysiological measures and observation.

In addition, for Kujala and colleagues [20], evaluate a momentary user experience is in most cases not very reliable for predicting the user experience in real life or to evaluate the success of a product. We need information about the long-term user experience because it causes people to continue to use a product and recommend it to others [20]. Furthermore, retrospective evaluations for long term experiments are based on the memory of the user and may be vulnerable [20].

3 Systematic Review

The systematic review was conducted in three phases: planning, execution and analysis of the results, shown below.

3.1 Planning

Objectives, research questions and search strategy were defined in planning phase.

Research Objectives. Based on the scenario described in the contextualization, the general objective of this research is to identify how the user experience is evaluated: the moment of evaluation, what techniques are used, which elements of experience are taken into consideration and the objectives of evaluation.

Research Questions. To meet the objectives of this systematic review, four research questions were formulated, as follows:

Q1. Do the authors consider a single measurement to assess UX or measurement over time?
Q2. How is done the UX evaluation (when, how, where and with what resources)?
Q3. What UX elements are considered when evaluating UX?
Q4. What are the objectives when evaluating UX?

Search Strategy. The search focused three bases: IEEE Xplore, ACM and Science Direct. The inclusion criteria considered in the systematic review were: articles published from 2008 (analysis of the last eight years), publication in conference or journal, articles available to read (permission to access) and written in English. The search string used was:

("user experience") AND ("evaluate" OR "evaluating")

3.2 Execution

The search yielded the following results: At ACM, 210 studies were found, while 418 studies were found in IEEE Xplore, and 104 in Science Direct, resulting in 732 studies. After the search, there were two phases of exclusion of studies. The first phase consisted of applying the inclusion criteria, and the analysis of titles and abstracts of the studies, excluding 645 of them, and leaving 87 studies for phase 2. The second phase consisted of reading the introduction, the detail of the proposal and conclusions, excluding 62 studies, and leaving 25 studies to collect data about user experience evaluation. The selected studies are presented in Table 1.

Table 1. Selected studies

ID	Year	Reference
1	2015	[21]
2	2015	[22]
3	2015	[23]
4	2014	[10]
5	2014	[24]
6	2014	[25]
7	2014	[26]
8	2013	[27]
9	2013	[28]
10	2013	[29]
11	2013	[5]
12	2013	[30]
13	2012	[31]
14	2012	[32]
15	2012	[33]
16	2012	[34]
17	2012	[35]
18	2011	[36]
19	2011	[20]
20	2011	[2]
21	2011	[19]
22	2010	[37]
23	2010	[38]
24	2010	[39]
25	2010	[40]

4 Analysis of Results

In the final phase of the review, data were collected from the selected studies.

With regards to the question Q1 (*"Do the authors consider a single measurement to assess UX or measurement over time?"*), Fig. 1 shows that 76 % of the studies consider only one measurement for user experience, i.e., only 24 % of the studies measure user experience over use.

To answer the question Q2 (*"How is done the UX evaluation (when, how, where and with what resources)?"*), several information was collected. The first one refers to the time of data collection: most studies consider the best data collection time at the end of the experience (32 %) or during and after it (36 %), as shown in Fig. 2. In addition, 16 % of the studies perform data collection during the experience and 16 % do it before and after.

Regarding the automation of collection of such data, most studies realize data collection manually (80 %), 12 % do it in mixed form (manual and automated) and 8 % do it automatically, as can be seen in Fig. 3.

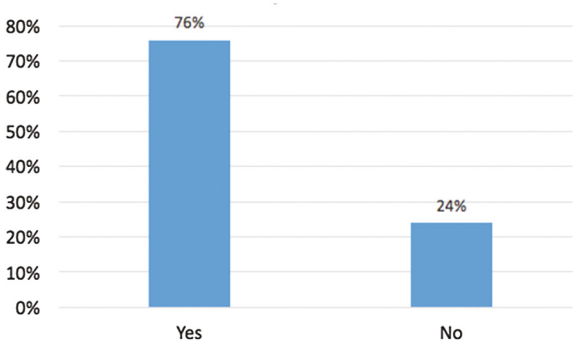

Fig. 1. Single measurement?

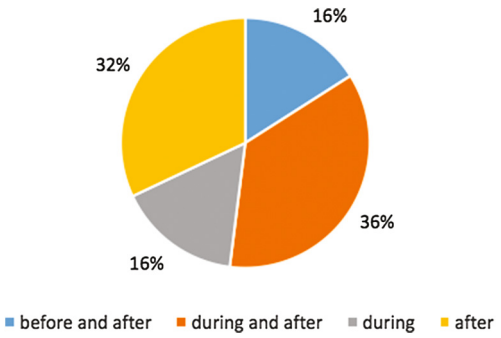

Fig. 2. Moment of data collection

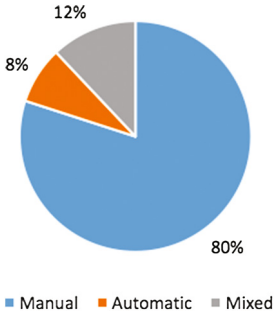

Fig. 3. Automation of user experience's data collection

After collecting user experience data, 84 % of the studies mentioned perform the evaluation of the user experience manually, 8 % do it automatically, 4 % do it in mixed form, and 4 % of the studies did not report how the user experience evaluation is performed. Figure 4 shows this information.

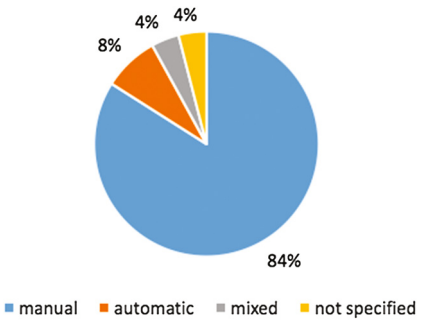

Fig. 4. Automation of user experience evaluation

Most of the studies also provide a controlled environment for experiments with users (44 %), while 36 % run the experiments in the real environment of the user. About 20 % of the studies did not specify what kind of environment carried out the experiments, as shown in Fig. 5.

Figure 6 shows the tools and techniques used to collect the information about the user experience. The majority of the studies (84 %) mentioned questionnaires to assess the user experience. Some of them use only the questionnaire, while others use the questionnaire at the end of the experiments, in conjunction with other tools or techniques. Other tools and techniques mentioned to evaluate user experience are interview, observation, reports, video recording, eye-tracking, among others.

As can be seen in Fig. 7, 72 % of the studies used ready-made products for the experiments, 8 % used prototypes, 16 % used both ready-made products and prototypes, and 4 % of the studies did not specify the product phase.

The graph in Fig. 8 answers the question Q3 (*"What UX elements are considered when evaluating UX?"*). In the chart are listed some facets, according to the model of Peter Morville (2004) plus some elements that were cited and evaluated (but are not cited in this model). The most cited elements were "desirable" and "usable" followed by "attractive (design)" and "valuable". According to that, not all facets are being used on the evaluation of the user experience. The UX element identified and less cited in the studies was "cause engagement" (only once).

The question Q4 addresses the observation made by Law et al. (2014), which states that most of the work does not report whether and how the information collected in the evaluation of the user experience are used in the following development cycles. Figure 9 shows that 56 % of the studies used the evaluation of experience to evaluate a prototype or product (focus on the product, not the user), 28 % categorized or just made records about the methods of evaluation, 8 % evaluated the user experience, 4 % compared user profiles, and 4 % evaluated user satisfaction. Most of the studies, even evaluating the user experience, uses the data to improve the product (in the future), not the user experience.

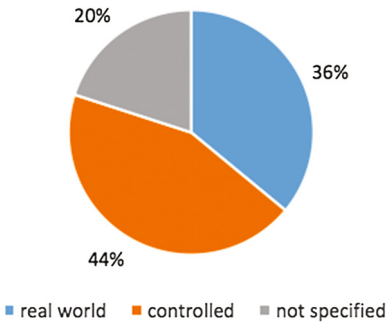

Fig. 5. Type of environment

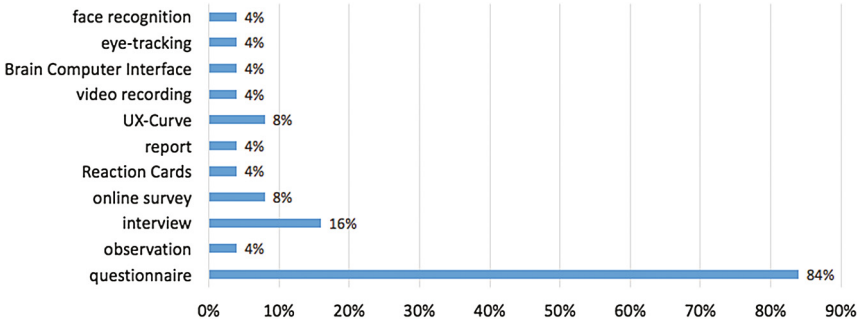

Fig. 6. Tools and techniques used to evaluate user experience

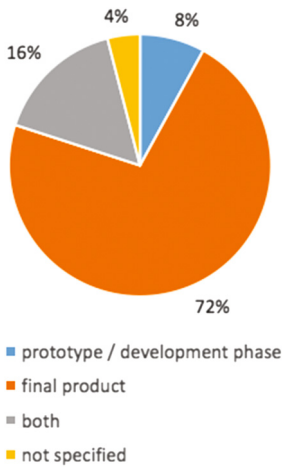

Fig. 7. Development phase

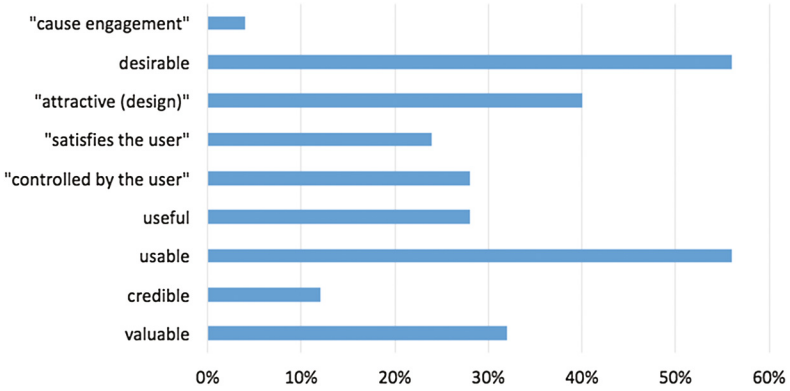

Fig. 8. Elements of user experience cited in the studies

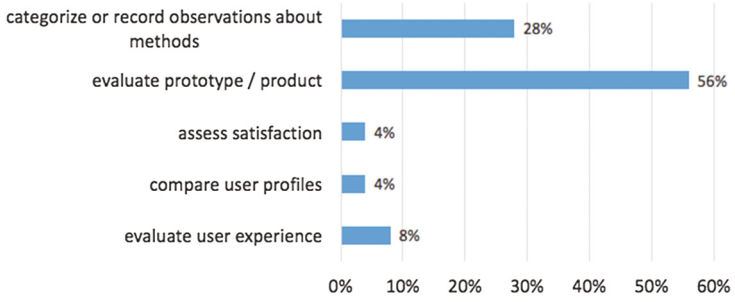

Fig. 9. What was done with the evaluation information?

5 Final Considerations

The main objective of this work was to identify how the user experience is being evaluated so far in the academic community. Data were collected about the moment of measurement and evaluation, techniques and tools used in the evaluation process, and the elements of user experience applied in the evaluation. After searching in the chosen databases and the application of the inclusion criteria, 25 studies were consolidated in this systematic review.

The first conclusion is that researchers prefer to check with participants his/her feeling about the product, even though they used other tools and techniques to evaluate user experience, once questionnaires were administered in 84 % of studies.

Psychophysiological measures, which tells us in real time what the user feels, have been widely used in the HCI area but are not yet being properly applied in evaluation models of the user experience, since 84 % of the studies evaluate the user experience manually.

Most studies cited experiments in controlled environments (44 %), against 36 % in the real user environment. About 20 % of the studies did not report that kind of environment used.

With regards to the development phase when it was used by the user, the product was already complete (developed) in 72 % of the studies. Few evaluation methods of the user experience were cited for product in development.

Despite using an evaluation of user experience, most studies (56 %) used this analysis to evaluate the product, i.e., the focus of the evaluation was the product, not the user.

Moreover, the majority of studies (76 %) utilizes single measurement, not considering the experience over time (during user's day).

Finally, only one of the 25 selected papers performed the evaluation of the user experience in real-time and over time.

Given the analysis above, the conclusion is that more methods for user experience evaluation are needed in a way that the use of the product over time and the information about the user experience in real time are considered. As mentioned earlier, the use of the product over time takes the user to continue using this product and recommend it to others. Moreover, the real time evaluation is important because it avoids relevant information being lost if it is collected after the experiment.

This study has some limitations, like, the use of only three databases for the search, the inclusion of studies from 2008, and the used search string. Although not considered highly relevant, these limitations might be responsible for the exclusion of interesting work and consistent with the theme of this systematic review.

The results of this review will form the basis for future works, since the evaluation of the user experience can provide feedback about the product, enabling improvements in the product and adaptation for the user, among other benefits.

References

1. Pine, B.J., Gilmore, J.H.: Welcome to experience economy. Harvard Bus. Rev. **76**, 97–105 (1998)
2. Kujala, S., Roto, V., Vaananen-Vainio-Mattila, K., Karapanos, E., Sinnela, A.: UX Curve: a method for evaluating long-term user experience. Interact. Comput. **23**(5), 473–483 (2011)
3. Sproll, S., Peissner, M., Sturm, C.: From product concept to user experience- Exploring UX potentials at early product stages. In: Nordic Conference on Human-Computer Interaction, pp. 473–482 (2010)
4. Roto, V., Obrist, M., Vaananen-Vainio-Mattila, K.: User experience evaluation methods in academic and industrial contexts. In: International Conference on Human-Computer Interaction (2009)
5. Makela, S., Bednarik, R., Tukiainen, M.: Evaluating user experience of autistic children through video observation. In: ACM Conference on Human Factors in Computing Systems, Extended Abstracts on Human Factors in Computing Systems, pp. 463–468 (2013)
6. Keskinen, T., Hakulinen, J., Heimoen, T., Turunen, M., Sharma, S., Miettinen, T., Luhtala, M.: Evaluating the experiential user experience of public display applications in the wild. In: International Conference on Mobile and Ubiquitous Multimedia (2013)

7. Foglia, P., Zanda, M.: Towards relating physiological signals to usability metrics: a case study with a web avatar. WSEAS Trans. Comput. **13**, 624–634 (2014)
8. Yao, L., Liu, Y., Li, W., Zhou, L., Ge, Y., Chai, J., Sun, X.: Using physiological measures to evaluate user experience of mobile applications. In: Harris, D. (ed.) EPCE 2014. LNCS, vol. 8532, pp. 301–310. Springer, Heidelberg (2014)
9. Wiebe, E.N., Lamb, A., Hardy, M., Sharek, D.: Measuring engagement in video game-based environments: investigation of the user engagement scale. Comput. Hum. Behav. **32**, 123–132 (2014)
10. Read, J. C.: Evaluating artefacts with children: age and technology effects in the reporting of expected and experienced fun. In: International Conference on Multimodal Interaction, pp. 241–248 (2012)
11. Masip, L., Oliva, M., Granollers, T.: OPEN-HEREDEUX: OPEN HEuristic REsource for Designing and Evaluating User eXperience. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 418–421. Springer, Heidelberg (2011)
12. Mercun, T.: Evaluation of information visualization techniques – analysing user experience with reaction cards. In: Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization, pp. 103–109 (2014)
13. Law, E.L., Roto, V., Hassenzahl, M., Vermeeren, A., Kort, J.: Understanding, scoping and defining user experience: a survey approach. In: ACM Conference on Human Factors in Computing Systems, pp. 719–728 (2009)
14. Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. Hum.-Comput. Interac. **19**(4), 319–349 (2004)
15. Mahlke, S., Thuring, M.: Studying antecedents of emotional experiences in interactive contexts. In: ACM Conference on Human Factors in Computing Systems, pp. 915–918 (2007)
16. Roto, V., Rantavuo, H., Vaananen-Vainio-Mattila, K.: Evaluating user experience of early product concepts. In: International Conference on Designing Pleasurable Products and Interfaces, pp. 199–208 (2009)
17. Law, E., Schaik, P., Roto, V.: Attitudes towards user experience (UX) measurement. Int. J. Hum.-Comput. Stud. **72**(6), 526–541 (2014)
18. Robert, J., Lesage, A.: Designing and evaluating user experience. In: Handbook of Human-Machine Interaction (2011)
19. Mashapa, J., van Greunen, D.: User experience evaluation metrics for usable accounting tools. In: Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, pp. 170–181 (2010)
20. Kujala, S., Roto, V., Vaananen-Vainio-Mattila, K., Sinnela, A.: Identifying hedonic factors in long-term user experience. In: Conference on Designing Pleasurable Products and Interfaces (2011)
21. Nawaz, A., Halbostad, J.L., Chiari, L., Chesani, F., Cattelani, L.: User Experience (UX) of the Fall Risk Assessment Tool (FRAT-up). In: International Symposium on Computer-Based Medical Systems, pp. 19–22 (2015)
22. Carofiglio, V., Ricci, G., Abbattista, F.: User brain-driven evaluation of an educational 3D virtual environment. In: Iberian Conference on Information Systems and Technologies, pp. 1–7 (2015)
23. Lasa, G., Justel, D., Retegi, A.: Eyeface: A new multimethod tool to evaluate the perception of conceptual user experiences. Comput. Hum. Behav. **52**, 359–363 (2015)
24. Walsh, T., Varsaluoma, J., Kujala, S., Nurkka, P., Petrie, H., Power, C.: Axe UX: Exploring long-term user experience with iScale and AttrakDiff. In: International Academic Mindtrek Conference: Media Business, Management, Content & Services, pp. 32–39 (2014)

25. Colombo, L., Landoni, M.: A diary study of children's user experience with EBooks using flow theory as framework. In: Conference on Interaction Design and Children, pp. 135–144 (2014)
26. Santoso, H.B., Isal, R., Yugo K., Basaruddin, T., Sadira, L., Schrepp, M.: Research-in-progress: user experience evaluation of Student Centered E-Learning Environment for computer science program. In: International Conference on User Science and Engineering, pp. 52–55 (2014)
27. Keskinen, T., Hakulinen, J., Heimoen, T., Turunen, M., Sharma, S., Miettinen, T., Luhtala, M.: Evaluating the experiential user experience of public display applications in the wild. In: International Conference on Mobile and Ubiquitous Multimedia (2013)
28. Saarinen, P., Partala, T., Vaananen-Vainio-Mattila, K: Little backpackers: studying children's psychological needs in an interactive exhibition context. In: International Conference on Interaction Design and Children, pp. 415–418 (2013)
29. Sim, G., Cassidy, B.: Investigating the fidelity effect when evaluating game prototypes with children. In: International BCS Human Computer Interaction Conference (2013)
30. Changyuan, G., Shiying, W., Chongran, Z.: Research on user experience evaluation system of information platform based on web environment. In: International Conference on Measurement, Information and Control, vol. 1, pp. 558–562 (2013)
31. Sim, G., Horton, M., Danino, N.: Evaluating game preference using the fun toolkit across cultures. In: Interaction Specialist Group Conference on People and Computers, pp. 386–391 (2012)
32. Sim, G., Horton, M.: Investigating children's opinions of games: Fun Toolkit vs. This or That. In: International Conference on Interaction Design and Children, pp. 70–77 (2012)
33. Staiano, J., Menéndez, M., Battocchi, A., de Angeli, A., Sebe, N.: UX_Mate: from facial expressions to UX evaluation. In: Designing Interactive Systems Conference, pp. 741–750 (2012)
34. Kronbauer, A. H., Santos, C., Vieira, V.: Um Estudo Experimental de Avaliação da Experiência dos Usuários de Aplicativos Móveis a partir da Captura Automática dos Dados Contextuais e de Interação. In: Brazilian Symposium on Human Factors in Computing Systems, pp. 305–314 (2012)
35. Rauschenberger, M., Olachner, S., Cota, M.P., Schrepp, M., Thomasschewski, J.: Measurement of user experience: a Spanish language version of the user experience questionnaire (UEQ). In: Iberian Conference on Information Systems and Technologies (2012)
36. Bach, C., Gauducheau, N., Salembier, P.: Combining interviews and scales in the multidimensional evaluation of user experience: a case study in 3D games. In: Annual European Conference on Cognitive Ergonomics, pp. 157–160 (2011)
37. Schulze, K., Kromker, H.: A framework to measure user experience of interactive online products. In: International Conference on Methods and Techniques in Behavioral Research (2010)
38. Meschtscherjakov, A., Reitberger, W., Tscheligi, M.: MAESTRO: orchestrating user behavior driven and context triggered experience sampling. In: International Conference on Methods and Techniques in Behavioral Research, p. 29 (2010)
39. Korhonen, H., Arrasvuori, J., Vaananen-Vainio-Mattila, K.: Let users tell the story: evaluating user experience with experience reports. In: ACM Conference on Human Factors in Computing Systems, Extended Abstracts, pp. 4051–4056 (2010)
40. Agarwal, A., Meyer, A.: Beyond usability: evaluating emotional response as an integral part of the user experience. In: ACM Conference on Human Factors in Computing Systems, Extended Abstracts, pp. 2919–2930 (2009)