

The Interplay Between Usability Evaluation and User Interaction Design

Kasper Hornbæk

Department of Computer Science, University of Denmark, Denmark

Jan Stage

Department of Computer Science, Aalborg University, Denmark

Usability evaluations inform user interaction design in a relevant manner, and successful user interaction design can be attained through usability evaluation. These are obvious conjectures about a mature usability engineering discipline. Unfortunately, research and practice suggest that, in reality, the interplay between usability evaluation and user interaction design is significantly more complex and too often far from optimal. This article provides a simple model of the interplay between usability evaluation and user interaction design that captures their main relationships. From the model, what is seen as the key challenges in improving the interplay between evaluation and design is outlined. The intention is to create a background against which the remainder of this special issue, containing 5 research articles presenting empirical data on the interplay between design and evaluation and a commentary, can be contrasted.

1. INTRODUCTION

Research into building usable systems consists of two, largely independent strands. On one hand, more than 20 years of research in human-computer interaction has created and compared techniques for usability evaluation (e.g., Lewis, 1982; Nielsen & Molich, 1990). On the other hand, methods for design of user interaction have significantly advanced in the last decades through, for example, the widespread use of contextual design (Beyer & Holtzblatt, 1998), agile development methods (Cockburn, 2001), and participatory design (Greenbaum & Kyng, 1991). However, we have seen little substantial exchange of results between the strands, and sparse efforts to combine their methods in practice. Larry Constantine, a prominent software development researcher, and his colleagues observed that "integrating usability into the software development process is not easy or obvious" (Juristo, Windl, & Constantine, 2001, p. 21).

In real-life software development, one of the successes of human–computer interaction is having contributed to the widespread interest in and use of usability evaluation. Nevertheless, usability evaluation and user interaction design are frequently carried out in surprising independence of each other. For example, it has proven difficult to integrate usability evaluation at relevant points in user interaction design with successful and to-the-point results. Most usability practitioners can retell experiences of usability reports that spend most of their lives archived in desk drawers with little or no influence on the development of the software they concern. The increased pressure to lower time-to-market does not facilitate increased interplay between evaluation and design; outsourcing of usability evaluation or development activities may also change and possibly worsen the integration of design and evaluation.

In this article, we discuss issues that affect the interplay between usability evaluation and user interaction design. We describe some of the research that have enumerated difficulties of making the interplay work and offer a model of the main activities and challenges in bringing usability evaluation and design closer together. The model serves to give a background for the empirical articles in the reminder of this special issue, but it also raises research questions still in need of answering.

2. A MODEL OF THE INTERPLAY

We posit that the model presented in Figure 1 captures the main relationships between usability evaluation and user interaction design.

A *usability evaluation* is an activity in software development with the purpose of determining the usability of the software system that is being developed. A usability evaluation is based on *design products* shaped in the user interaction design activity. The products from user interaction design are typically running applications, early implementations, or quick-and-dirty prototypes. Yet they could be design specifications, paper prototypes, mock-ups, or throwaway prototypes.

A usability evaluation is conducted to fulfil a specific purpose. In the figure, this is represented generally as a *commission statement* that results from a management

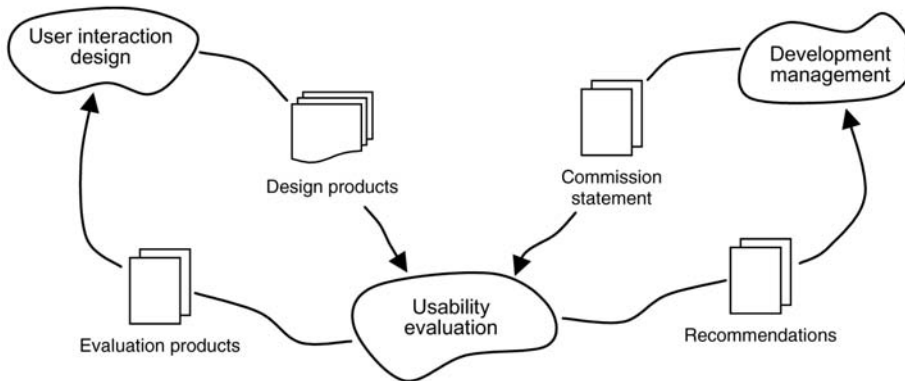


FIGURE 1 The interplay between user interaction design and usability evaluation.

activity. In practice, there is often no explicit statement of the purpose of a usability evaluation, or it may be informal or vague, and it is definitely not coming from management. However, this only means that in this case the evaluators define the specific purpose of the evaluation, even if this is only done implicitly; by defining this purpose, the evaluators make a management decision. Thus the intention of the figure is to emphasize that such a decision is always made but not to define when or by whom. The implication is that a conscious and systematic usability evaluation should be based on an explicit commission statement.

The figure also illustrates the output from the evaluation activity. The main output is a set of *evaluation products*, that is, an assessment of the usability of the system determined in accordance with the commission statement. These results are fed back into user interaction design. The most typical evaluation result is a conventional usability report with a long and detailed list of usability problems that have been identified during the evaluation. However, there are several other possibilities for providing feedback to the designer, such as more focused reports, multimedia presentations of problems, proposals for redesign, and workshops involving both evaluators and designers. The intention of all these feedback techniques is to overcome the communication gap between evaluators and designers.

The other main output from the usability evaluation is a set of recommendations. When the evaluators conduct a usability evaluation, they gain rich insight into the system that is being evaluated. Some of this insight is useful for making decisions about the further development of the system, which includes suggested prioritizations of usability problems as well as analyses of the reasons for their occurrence. These elements are useful for deciding what to do about the results of the usability evaluation. This is a management decision. In practice, this is much like the commission statement; it is either nonexistent or very informal. Again, this does not mean that management decisions are not made, but often they are made unconsciously or in the wrong forum. If developers decide to ignore a usability report and continue without changes, they have made a significant management decision. The receipt of the evaluation results should therefore be seen as a significant decision point in a development project.

Although this model is less detailed than other available models, for example, that of user centered software development by Mayhew (1999), it pinpoints what we see as the main challenges of creating a successful interplay between usability evaluation and user interaction design. Thus the model is not intended to be complete. The separation of activities is also intended to be only conceptual; in practice, the same people may manage, design, and evaluate software. The following section outlines what in our view are the four most serious challenges to improving the interplay between usability evaluation and user interaction design.

3. CHALLENGES

Figure 1 helps pinpoint at least four challenges in improving the interplay between usability evaluation and design, each corresponding to a product or an activity in the figure.

The first challenge concerns the kinds of products that are fed into usability evaluation from design. Typically these comprise prototypes or running applications. Obviously, this kind of product works well for many evaluations, and considerable interest has gone into describing the relative benefits of low- and high-fidelity prototypes for usability evaluation (e.g., Catani & Biers, 1998; Rudd, Stern, & Isensee, 1996). Some authors, however, have warned that prototypes may misrepresent functionality (Berghel, 1994). A more general concern is that these products capture only part of the activities in the design of user interaction. Little discussion has been made of systematic evaluation based on other design products, such as personas, scenarios, or use cases. This is surprising and, we believe, unfortunate, because these alternative design products may force richer and broader interplay between evaluation and design. Against the argument that evaluation of these design products naturally and mostly implicitly forms part of design practice, we ask if not systematic study of such evaluation, and the development of better techniques for doing it, could be fruitful, similar to the way usability research has increased our knowledge about usability evaluation techniques.

The second challenge concerns commissioning, or focusing, the evaluation. We speculate that a key reason for the suboptimal interplay between design and evaluation is that some evaluations are not focused on the questions that the design team needs answered. Many textbooks provide hints on establishing test objectives (e.g., Rubin, 1994) but nowhere near the amount of details provided for, say, debriefing of test participants. Another example is the preparation of test questions in think-aloud tests. Although most authors agree that this is a crucial step, few explain how to link this activity to the models and current issues in the design activity about typical users' typical tasks. Recently, we have seen some interest in meeting this challenge. Cockton (2005), for instance, proposed a framework for value-centered design, one important idea of which is to derive evaluation goals from high-level expressions of the values aimed for by a particular application. However, we still have to see techniques for doing this, and we still need empirically based reports on the benefits and drawbacks of various evaluation-focus forms.

The third challenge concerns the evaluation products, or feedback, from the evaluation process into the design processes. Typically, research has considered feedback in the form of lists of usability problems. For example, such lists have been widely used to compare usability evaluation methods. Among things, this conflates naming of potential problems with identifying real problems, because counts of potential problem will include problems that are not true usability problems (Gray & Salzman, 1998). Different kinds of problems—for example, with respect to generality, type, aspects of the user interface covered, or clarity—are given equal weight when counted. A variety of other arguments have been raised against using problem lists to compare usability evaluation methods (e.g., Hornbæk & Frøkjær, 2005; see several of the articles in this special issue). In practice, usability reports are typically the main form of feedback. These reports appear to be quite varied in content and occasionally fall short of effectively communicating the results of a usability test (Molich, Ede, Kaasgaard, & Karyukin, 2004). Surprisingly, after more than 2 decades of research in usability evaluation, one of the strongest feedback mechanisms in practice remains having designers

and developers observe a think-aloud test. The challenge here appears to be to develop richer and more varied evaluation products that better fit the needs of designers and developers.

The fourth challenge concerns generating recommendations and prioritizations to management of the design. This is conceptually different from evaluation products in that evaluation products may contain no hints on what to do about problems or how to select the most severe ones. In research, there is a surprising lack of studies investigating how to prioritize among usability problems; Hassenzahl (2000) is one notable exception. In practice, the process of analyzing problems to provide recommendations synthesized across a series of usability test is often done in an ad hoc manner with no explicit procedure or method being applied (Nørgaard & Hornbæk, 2006). We see support for analysis, prioritization, and recommendation as particularly important to support usability evaluations in influencing design.

4. TWO APPROACHES TO PRACTICE THE INTERPLAY

This special issue includes five articles that represent two approaches to practicing the interplay between usability evaluation and user interaction design.

The first approach is based on separation. The key characteristic is that usability evaluation and user interaction design are done separately and by different groups of people. With this approach we distinguish between user interaction designers, who are part of the development team that creates and implements the software, and usability evaluators, who conduct the evaluations and produce the results and recommendations that are fed back to design and management. Three articles represent this approach.

Morten Hertzum (this issue) focuses on means for guiding designers in prioritizing their efforts in response to evaluation products. He deals with feedback in the form of lists of usability problems. His article combines two empirical studies, a field study of factors that influence the impact of evaluations and an experimental study of severity assessments made during usability inspections. The results of Hertzum's article provide insight into the impact of such an evaluation product and the extent to which designers' attention is influenced by severity assessments. The article proposes a formula for calculating a severity measure, essentially an attempt at answering our challenge about recommendation and prioritization.

Effie Law (this issue) focuses on the persuasiveness of usability evaluation results and the impact in terms of resolving the problems. She reports from a case study, where she has assessed the utility downstream in the development process of implementing solutions and the effect on the developers. The results emphasize the relation between these factors and suggest implications for the way usability problems are reported, in particular focusing on our challenges about evaluation products.

Rune Høegh, Christian Nielsen, Michael Overgaard, Michael Pedersen, and Jan Stage (this issue) focus on the extent to which developers' opinions can be influ-

enced by different evaluation products. They report a case study and a field experiment that compare providing feedback through observation of user tests and through reading of usability reports. In particular they study developers' understanding of usability data. The results emphasize how observation of user tests and reading of usability reports impacted the developers' understanding of usability problems.

The second approach is based on integration. The key characteristic is that usability evaluation and user interaction design are done in an integrated manner in the same process and by the same group of people. This approach studies designers or developers who conduct evaluation activities as an integrated part of their work. Two articles represent this approach.

Ann Blandford, Suzette Keith, and Bob Fields (this issue) deal directly with the challenge of integrating usability evaluation methods in a design practice. They report from a case study where it was investigated how user concerns could be included within an unstructured, system-focused development process. In this study, they used Claims Analysis as a method for assessing the effects of design decisions on users' experience. The results assess the extent to which this method can be learned, communicated to developers, and applied in practice. They have also identified a tension between user-centered and function-oriented design approaches.

Giorgio Venturi, Jimmy Troost, and Timo Jokela (this issue) focus on user-centered design as a method to support integration of usability evaluation and user interface design. They have investigated adoption of user-centered design through a Web survey. The results deal with the extent to which user-centered design practitioners are involved in the early stages of the product life cycle compared to 10 years ago and the shifts that have occurred in the methods and techniques employed. On this basis, it is suggested which organizational factors user-centered design practitioners and their management should consider.

These five articles represent not only two different approaches to handle the interplay between evaluation and design but also a broad variety of research approaches. This includes field studies, case studies, experiments, and surveys. In the final article of this special issue, Gilbert Cockton (this issue) discusses the research behind the first five articles and relates them to what he sees as open problems in our research discipline.

5. CONCLUSION

A lack of interplay between usability evaluation and user interaction design is evident both from the research literature and from studies of and experiences in practical software development. We have outlined a simple model that captures main activities and products that connect evaluation and design. This model raises four challenges for research in usability evaluation: (a) what forms of design products give the best evaluations; (b) how do we most effectively focus an evaluation so as to give pertinent results; (c) what kinds of feedback are most useful in design; and (d) how do we support prioritizations, analysis, and recommendation about what to do with the results of usability evaluations? We believe these challenges are cru-

cial to improve the development of software, and we find that the five articles in this special issue provides much needed empirical data that attempt to address them.

REFERENCES

- Berghel, H. (1994). New wave prototyping: Use and abuse of vacuous prototypes. *ACM Interactions*, 1(2), 49–54.
- Beyer, H., & Holtzblatt, K. (1998). *Contextual design*. San Francisco: Morgan Kaufmann.
- Catani, M. B., & Biers, D. W. (1998). Usability evaluation and prototype fidelity: Users and usability professionals. *Proceedings of Human Factors and Ergonomics Society 42nd Annual Meeting*, 1331–1336.
- Cockburn, A. (2001). *Agile software development*. New York: Addison-Wesley.
- Cockton, G. (2005). A development framework for value-centred design. *Proceedings of CHI '05 extended abstracts on Human Factors in Computing Systems*, 1292–1295.
- Gray, W. D., & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, 13, 203–261.
- Greenbaum, J., & Kyng, M. (1991). *Design at work: Cooperative design of computer systems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Hassenzahl, M. (2000). Prioritising usability problems: Data-driven and judgement-driven severity estimates. *Behaviour & Information Technology*, 19, 29–42.
- Hornbæk, K., & Frøkjær, E. (2005). Comparing usability problems and redesign proposals as input to practical systems development. *Proceedings of ACM Conference on Human Factors in Computing Systems*, 391–400.
- Juristo, N., Windl, H., & Constantine, L. (2001). Introducing usability. *IEEE Software*, 18(1), 20–21.
- Lewis, C. (1982). *Using the "thinking-aloud" method in cognitive interface design* (Research Rep. No. RC9265). Yorktown Heights, NY: IBM T. J. Watson Research Center.
- Mayhew, D. J. (1999). *The usability engineering lifecycle*. San Francisco: Morgan Kaufmann.
- Molich, R., Ede, M. R., Kaasgaard, K., & Karyukin, B. (2004). Comparative usability evaluation. *Behaviour & Information Technology*, 23, 65–74.
- Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of ACM Conference on Human Factors in Computing*, 249–256.
- Nørgaard, M., & Hornbæk, K. (2006). What do usability evaluators do in practice? An explorative study of think-aloud testing. *Proceedings of ACM Conference on Designing Interactive Systems (DIS 2006)*, 209–218.
- Rubin, J. (1994). *Handbook of usability testing*. New York: Wiley.
- Rudd, J., Stern, K., & Isensee, S. (1996). Low vs. high-fidelity prototyping debate. *ACM Interactions*, 3, 76–85.

Copyright of International Journal of Human-Computer Interaction is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.