

Input from usability evaluation in the form of problems and redesigns: results from interviews with developers

Erik Frøkjær

Department of Computing
Universitetsparken 1
DK-2100 Copenhagen
+45 35321456

erikf@diku.dk

Kasper Hornbæk

Department of Computing
Universitetsparken 1
DK-2100 Copenhagen
+45 35321400

kash@diku.dk

ABSTRACT

Usability problems predicted by evaluation techniques are useful input to systems development; it is uncertain whether redesign proposals aimed at alleviating those problems are likewise useful. We compare problems and redesign proposals as input from usability evaluation into industrial software development, as discussed in the literature. We do so by presenting comments from interviews with system developers on what aspects of problems and redesigns they find to be of utility. Our study suggests that redesigns should be given more attention, both in comparisons of usability techniques and in practical usability evaluation.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces—Evaluation/Methodology; D.2.2 [Software Engineering]: Design Tools and Techniques—User Interfaces.

General Terms

Measurement, Design, Experimentation, Human Factors.

Keywords

Usability evaluation, redesign, think aloud, metaphors of human thinking, empirical study, usability inspection.

1. INTRODUCTION

Most of the research on usability evaluation methods assumes that good usability evaluation techniques are those that best support an evaluator in generating problem descriptions while using the techniques; Hartson et al. [4], for example, suggests treating usability evaluation techniques as functions that produce problem lists, ignoring issues of how to treat problem descriptions and redesigns. This assumption has several limitations:

- Problem descriptions are sometimes very brief. The 46 usability problems described in [7, appendix 1], for example, is on the average about 28 words long. Therefore, problem descriptions may appear unclear or incomprehensible to readers other than the evaluator.
- When analyzing the effectiveness of usability evaluation techniques, problems are often compared in order to match similar problems. This matching process, however, turns out to be difficult and precarious [9].

- Sometimes no design exists that alleviate the usability problems described, e.g. because the changes needed conflict with other requirements of the design or dictate extremely complex functionality. Designers may waste resources in trying to cope with such problems.
- Generation of lists of usability problems may not matter much in practical systems development. Wixon [13] comments on a recurring discussion regarding comparison of evaluation techniques that '[i]t is short sighted in that it ignores that problems should be fixed and not just found'.

Taken together, these limitations suggest that it is feasible to examine alternatives or supplements to problem identification and description as the goal underlying the creation and comparison of usability evaluation techniques.

This paper explores if and how redesign proposals may supplement problem descriptions as valuable input from usability evaluation to practical systems development.

2. PROBLEMS AND REDESIGNS

Only few studies have investigated redesign proposals as an outcome of usability evaluation [2,8,11,12]. For example, Dutt et al. [2] considers the ability of heuristic evaluation and cognitive walkthrough to produce requirements for redesigns. While requirements are related to a specific technique, the study doesn't describe the format or nature of those requirements. The study by Sawyer et al. [12] on the impact of inspections on software development suggests that '[p]roviding specific recommendations to fix specific problems has a tremendous positive effect: The development group need not spend time thinking of a solution, plus we gain a psychological advantage in offering constructive suggestions rather than just criticism' (p. 379). This study, however, does not compare usability problems and redesigns, nor point out particularly useful aspects of redesign proposals.

In practical usability work, redesign proposals are often made in the form of quick fixes. Dumas et al. [1] mentions how usability reports from teams of expert evaluators often include proposals for how to fix problems. Usually, however, the quick fixes are only as brief as problem descriptions. They suffer from some of the same limitations that were attributed to usability problems in the introduction. Further, proposals are sometimes quite vague, leading the authors to question 'would the developer who created this site be able to make better choices from these suggestions?' (p. 29). This suggests that some more developed form of redesign proposals could be feasible.

In summary, related work provide some arguments for redesign proposals as (part of) the result of usability evaluation. None of the studies, however, have moved beyond quick fixes integrated with or quite similar to usability problems. Thus, little is known about the utility of redesign proposals, especially of their relative merits compared to problem descriptions.

3. INTERVIEWS WITH DEVELOPERS

As part of a study that compared evaluation techniques we interviewed developers about their perception of usability problems and redesign proposals. Details of the study will be reported elsewhere; here we focus just on interviews with developers.

Forty-three undergraduate and graduate students chose to conduct the evaluation and redesign in a class on HCI and systems design. They evaluated one of Denmark's largest job portals, www.jobindex.dk. The evaluators had one week to conduct the evaluation, and performed it individually. They were told to use approximately eight to ten hours on conducting and reporting the evaluation. Twenty-one evaluators received reference [10] as description of think aloud user testing; twenty-two evaluators received reference [5] as description of the usability inspection technique called metaphors of human thinking.

After completing the evaluation, each evaluator produced three redesigns, one for each of the three parts of Jobindex evaluated. Thirty-six evaluators handed in redesigns, for which they had been asked to use around ten hours. Evaluators were told to create redesigns that addressed some of the usability problems they considered to be the most critical for the users of the application. They were told to imagine that they should provide input for a discussion of whether a redesign decision should be worked out into further detail and possibly be implemented. Evaluators were asked to provide (1) a brief summary of the redesign; (2) a brief argument why the proposed redesign is important; (3) an up to one page explanation of interaction and design decisions in the redesign; and (4) up to two pages of illustrations of how the redesign works.

In practical usability work, the development team has a decisive role in choosing which usability problems to correct and which redesign proposals to follow. Therefore, problems and redesign proposals were assessed by four core members of the development team at Jobindex: (a) the founding director who plays a crucial role in the development; (b) two developers each working on and responsible for parts of the application that were evaluated; (c) a web content manager, responsible for a part of the application evaluated. For brevity, we refer to these four persons as developers. The developers individually assessed a selection of problem descriptions and redesign proposals. Problems and redesigns were presented to developers in a randomized order, alternating between 11 problems, a redesign proposal, 11 problems, etc. One of the developers rated all problems and redesign proposals; the other developers rated those problems and redesigns concerning the part of the application that they work on. The results of the assessment is not included in this paper.

Approximately a week after developers had finished assessing the usability problems and redesign, we conducted individual interviews with them. We asked about their background,

experience with rating problems, and impressions of the qualities of redesigns and problems. In addition, we presented them with examples of problems and redesigns that they had assessed as having high or low utility, and asked for their reasons for the assessment. Because the web content manager was working on a part of the application mainly delivering information, we did not interview that developer about redesigns (as this would have regarded changes to content only, not the more complex interaction parts of the user interface). Each interview lasted around an hour.

3.1 Descriptions of usability problems

All developers felt that they already knew most of the problems described by the evaluators. One of the developers said, for example, 'There is not so much new in it' and continues:

the issues that have been identified, they are either issues which we do not judge as very important, or issues we were well aware of already and with which we knew there were problems ... but have not had the time to deal with

While agreeing on the problems, developers appeared to assess severity somewhat differently from evaluators. One of the developers expressed surprise that evaluators had taken such effort to point out a problem he agreed existed but otherwise considered minor. Another said that 'practical experience shows that users can do that', practical experience probably referring to the web logs. Of those usability problems developers said they did not know, actual bugs were given much attention, e.g. 'that [a problem description] is one of our serious problems, it is a bug that we have been chasing without being able to find its cause ... such a bug has a high priority on our list'.

The developers' main uses of the problems seemed more to be for prioritizing what to do something about and for confirming design decisions nearing completion, rather than for getting surprising new information. For example,

usability problems ... what one cares about is the extent of them, how many is saying that some thing is a problem and how many is saying that some other thing is a problem, that help me prioritize what I should focus on

An aspect of usability problems emphasized by one of the developers was the reference to users and their problems, e.g. 'I liked best those [problems] that said that the users ... that the user tests showed something'.

The developers also noted limitations in the problem descriptions which impacted their utility in the systems development. For example, when seeing a problem again during the interview, one of the developers gave the following example:

so if an evaluator's comment is that the password is too short, then my comment is: what do you mean by that, too short for what? Exactly because it is short users may be able to remember it, but if he says that the password is too short because a hacker could log in and steal you personal information, then I could say OK now we are talking about that problem

Thus, the lack of clear reasons why something is a problem was considered a shortcoming. Occasionally, problem descriptions would point out something as a problem, but ignore that alternative designs would lead to similar or worse usability problems. In discussing how to show hits of a search in job advertisements, one developer argued:

ok, so you cannot see where the hit was...on the other hand if we presented the [place in the add] where the hit was instead of the nice form of the add, then that would lead to problems also...so you present a problem, but what is the solution to that problem...sometimes you have, you have some alternatives [to the currently implemented solution], but because there is a problem with one alternative then it is not sure that the other [alternative] is better

Finally, some of the descriptions of usability problems would ignore issues outside of the development team's control. Some problems suggested changing the label of a button for uploading an image to which one of the developers commented that 'we don't have control over the text on it' (because this is done by the operating system) and thus considered that problem to be of low utility.

3.2 Redesign proposals

Compared to usability problems, the single most frequent comment about redesign proposals is that they give good ideas. For example:

ok, there were some pearls in it ... sometimes things that we had not thought about, especially redesign proposals for saying, ok that way of doing it is also possible

And later on remarks that:

in some situations you may do things one way or the other, and then you can just choose, i.e. whether some list should be alphabetical or just split up...in other situations, like the three level hierarchical selection of job titles, no matter what we do we get into some complicated mess...so if one can find some way of making it more intuitive and usable than other ways, then we accept it eagerly, [because] we haven't quite figured out how to do it ourselves

This input seems especially welcome when developers are tackling a 'particularly hard nut to crack', or when they are just looking for information on 'what is a good idea to get on'.

During all interviews, we asked developers if they could recall usability problems and redesign proposals. Usability problems were mostly remembered by developers as classes of problems, the particular instances was forgotten. One developer said that 'yes, there are several of them [usability problems] that I can still remember' and went on to expand on how redesign proposals on exploring similarities to standard search engines could be incorporated in the design. All developers were, however, able to describe in some detail redesign proposals which they had found interesting:

for example, someone came with a simple solution to a problem that we have had for a long time: we have a selection box where you may choose counties and

cities, which we put into the same selection box ... someone suggest why don't you split it up so that you can either select a county or a city or a country ... make three lists instead of one ... that is one way of doing it which we did not consider previously

A number of attributes of redesigns seem to work well in the developers' opinions. For example, the illustrations (evaluators mostly did these as drawings or mock-ups in HTML) were well liked. For example,

I think it was those [redesign proposals] that I gave a high assessment, they were really interesting ... yes, both of them were characterized by, well they [the evaluators] had grabbed a pencil and made a drawing and said: you could make it in such and such way, thought out of the box so to speak...that is probably the single most positive thing in the entire file [of redesigns and usability problems]

Two developers found the redesigns more concrete than problem descriptions, meaning that they were more clear about what evaluators had in mind when describing the redesign. One of the developers emphasized how, as a form of communication, the redesigns were much more constructive: 'it is almost obvious that it is better to say: if it were this way it was better, rather than just saying: this is wrong... so say this is wrong and here is the alternative'. And finally, all developers stressed how the redesign proposals felt more coherent and complete, i.e. 'there were more meat in them' and 'there is a little more thought in it, a little more completeness'.

As with usability problems, developers pointed out several limitations of the redesigns. For example, some of the redesigns were descriptions of 'more radical proposals for changes, how you can make the things by advanced Java script and stuff like that, that is a new idea but not one that we can use because it is too complicated'. Thus, technical feasibility and coherence with the overall use of technology meant that this proposal did not have much utility for the developer. Similarly, a developer said, reflecting upon a redesign proposal that he recalled: 'then it begins to get confused and complex ... and the problem starts to grow ... but there are no thoughts on which consequences do this have in the rest of the system'.

Still other redesign proposals were put aside because they did not fit with the printing of resumes on paper that the application were also used for.

Even when redesigns were put aside for reasons like above, developers found them to be of utility. For example, one developer noted that he considered the problem a particular redesign tried to solve to be irrelevant, still the solution was interesting: 'this particular one I can remember because it is the right solution, but the wrong rationale'. Another example is when the proposed solution does not feel right to the developer, but the idea behind the solution is fine, e.g. 'I think that the idea that the user can write and add [job descriptions] is not bad at all, but I am not convinced it should be done in this way'.

3.3 General comments on input from usability evaluation

All developers expressed that both usability problem descriptions and redesign proposals were of very high quality,

e.g. 'they are quite good, both the comments and the redesigns, they capture very well what we are trying to do and come up with some good proposals'. We also asked developers if they would want to receive only problems or redesigns, and all expressed that they wanted to receive both.

Across usability problems and redesign proposals, developers expressed that problems of utility to them were problems that could be fixed easily and quickly. One developer explained:

typically if something can be easily and quickly fixed ... that is a suggestion which requires four months of development is not as useful as some small suggestion, which corrects a small problem in 10 minutes, then I can correct it immediately

In fact, developers and the web content manager all had corrected one or more problems when we interviewed them, approximately one week after having worked through the problems and redesigns.

4. CONCLUSION

The study shows that developers value redesign proposals as input to their development work. The interviews suggest that (a) redesign proposals help developers understand usability problems, i.e. redesigns contribute to characterizing and making more concrete the problems found, and illustrate why problems are important; and (b) redesign proposals are useful for inspiration and for seeking alternative solutions for problems that the development team has been struggling with. These comments do not mean, however, that developers did not appreciate usability problems, especially when they are well argued, clearly described, documented, and easy to fix. On the contrary, all developers wanted both problems and redesign proposals as input from usability evaluation to systems development.

These results suggest that usability evaluations should place more focus on developing and reporting such proposals than is typically done.

The results stand in contrast to the scientific literature on usability evaluation techniques, which largely ignore proposals for redesigns as input to systems development. Redesign proposals may help move beyond Wixon's [13] complaint that most comparisons of usability evaluation techniques focus exclusively on the techniques' ability to generate problems, ignoring what is needed in practical systems development. Moreover, focusing on redesign proposals may help improve the validity of comparisons of usability evaluation techniques, the limitations of which have been pointed out by several authors [3,6]. This could be expected because redesign proposals, according to the developers interviewed, are more concrete,

more relevant to their work, and better able to give a clear understanding of what an evaluator intended.

5. REFERENCES

1. Dumas, J., Molich, R., & Jefferies, R. Describing Usability Problems: Are We Sending the Right Message?, *interactions*, 4 (2004), 24-29.
2. Dutt, A., Johnson, H., & Johnson, P. Evaluating Evaluation Methods, *Proc. HCI 1994*, Cambridge University Press (1994), 109-121.
3. Gray, W. D. & Salzman, M. C. Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods, *Human-Computer Interaction*, 13, 3 (1998), 203-261.
4. Hartson, H. R., Andre, T. S., & Williges, R. C. Criteria for Evaluating Usability Evaluation Methods, *International Journal of Human-Computer Interaction*, 13, 4 (2001), 373-410.
5. Hornbæk, K. & Frøkjær, E. Evaluating User Interfaces With Metaphors of Human Thinking, *Lecture Notes in Computer Science 2615*, Springer (2002), 486-507.
6. Hornbæk, K. & Frøkjær, E. Two Psychology-Based Usability Inspection Techniques Studied in a Diary Experiment, *Proc. NordiCHI 2004*, ACM Press (2004).
7. John, B. & Mashyna, M. M. Evaluating a Multimedia Authoring Tool With Cognitive Walkthrough and Think-Aloud User Studies, *CMU-HCI-95-105 / CMU-CS-95-189* (1995).
8. Johnson H., Generating User Requirements From Discount Usability Evaluations, in Harris, D. *Engineering Psychology and Cognitive Ergonomics, Vol. 2*, Ashgate Publishing, 1997, 339-357.
9. Lavery, D., Cockton, G., & Atkinson, M. P. Comparison of Evaluation Methods Using Structured Usability Problem Reports, *Behaviour and Information Technology*, 16, 4/5 (1997), 246-266.
10. Molich, Rolf, User testing, Discount user testing, 2003, www.dialogdesign.dk.
11. Muller, M. J. & McClard, A. Validating an Extension to Participatory Heuristic Evaluation: Quality of Work and Quality of Work Life, *Proc. CHI'95*, ACM Press (1995), 115-116.
12. Sawyer, P., Flanders, A., & Wixon, D. Making a Difference - The Impact of Inspections, *Proc. CHI'96*, ACM Press (1996), 376-382.
13. Wixon, D. Evaluating Usability Methods: Why the Current Literature Fails the Practitioner, *interactions*, 10, 4 (2003), 29-34.