Generating and Using Results from Usability Evaluations

Kasper Hornbæk
University of Copenhagen
www.kasperhornbaek.dk

Background

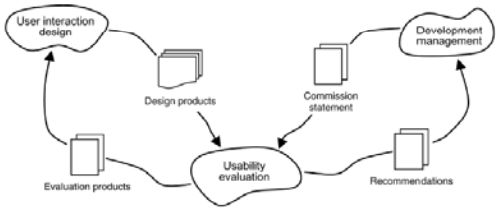
- Associate professor in the Human-computer Interaction group at Copenhagen University
- Interested in usability research and information visualization
- 8 years of usability experience

Challenges in usability research

- Most research studies focus on reports that list usability problems with a certain piece of software (Wixon 2003)
- Only very few studies of usability evaluation are conducted in real industrial settings (Gray & Salzman 1998; Hartson et al. 2001)
- Research is only beginning to address how developers understand and assess usability problems (e.g., Hornbæk & Frøkjær 2004)

Challenges in practice

- Usability work has too little influence on development activities



Hornbæk & Stage (2006)

Today

- Three themes through research studies and exercises:
  - Think aloud tests
  - Describing usability problems
  - Making recommendations from usability tests

What Do Usability Evaluators Do in Practice?

- Think aloud testing is the most widely used evaluation method
- Almost no studies of how professional evaluators actually conduct and report think aloud tests
  - but see Boren & Ramey 2000 and Molich et al. 2004
- Would help researchers understand practice and aid practitioners in improving their work habits
  - Are we supporting the right activities? Are techniques misunderstood or used as prescribed?

YOUR thoughts on think-aloud testing

- Please consider:
  - What would be surprising to outsiders watching YOU conduct and report a think-aloud test?
  - In which ways do you diverge from 'normal' think-aloud tests?
  - What would you like to improve?
  - What may impact the quality of your findings?
- I will now keep quiet for about 10 minutes to give you a chance to generate some answers
- Feel free to discuss with you neighbors!

What do evaluators really do?

- 14 think aloud sessions in seven companies:
  - 2-8500 employees
  - 2-8 of the employees work with evaluation
  - Evaluators' experience between 1 and 8 years
- We followed the setup, carrying out, analysis and discussion of the test
- About 25 hours of recordings

Nørgaard, M. & Hornbæk, K. (2006), "What Do Usability Evaluators Do in Practice? An Explorative Study of Think-Aloud Testing", *ACM Symposium on Designing Interactive Systems (DIS 2006),* 209-218.

Main findings

Area of practice	*Main finding*	*N*	*Example of observations and quotes*
Analysis of results from a test session	Analysis is unstructured	9	Scattered fragments of analysis; no systematic approach used
	Analysis is incomplete	9	Does not identify causes or solutions; restricts discussion to user traits
	Analysis as a summary with the user	3	"Let's sum up"; selecting a few problems for further questioning; listing key findings
Confirmation of known issues as a test focus	Looking for known issues	8	"Now, I am just looking for ammunition"; develops ideas of problems before testing; tasks and questions designed to point out known issues
	Practitioners have foreseen problems	5	"We have a gut feeling", "I told you so"
Practical realities influencing tests	Technical problems	8	System breaking down; long response times in test environment; installation or security messages interrupt workflow
	Unfinished prototypes	6	Parts of prototype missing or inaccessible; "a log-in name should not be Waddik/Fish"; texts and pictures are wrong or out of date
Questions asked during a test	Problems are explained, not experienced	13	"Do you think you would go back to the front page"; "did you notice this column", "what do you expect to see"
	Leading questions	11	Questions address certain parts of GUI or system; evaluator hints the solution; "Can you do this in another way?"
	Unnecessary or obvious questions	10	"You did figure out to press the print button?"; asking user to locate information that clearly appears on the present screen; asking if user would like relevant information
Trying to meet laboratory-style scientific standards	Evaluators want similar conditions for users under test	5	"We have to make sure all users get the same questions"
	Rigid or artificial procedures	5	Laboratory style procedures; Danish evaluators speaking English to a Danish user; measuring subjective satisfaction overly systematic
Uncovering usability problems or utility concerns	User points to utility or lack thereof	10	"I would not do it like this"; user chooses to solve task without help of system
	Evaluator probes utility concerns	7	Asking about normal workflow; asking whether a task is realistic; "What would you typically do?"

How are think-aloud sessions analyzed?

- No sessions were immediately followed by an attempt at a structured analysis of the session
- Analysis seems a weak part of the think aloud setup
  - Challenge for researchers
  - Practitioners should more systematically collect and discuss their observations

Questions during the sessions

- Some questions are not used to understand usability problems that are experienced by the user, but ask the user to predict problems
- For these problematic questions:
  - Researchers must note that practitioners want to get a variety of information: can this be done in a valid way?
  - Practitioners should be cautious with respect to how they put questions to users

Utility versus usability

- We may distinguish usability and utility
- Utility concerns are less frequently reported compared to usability problems
- Utility problems seem very important:
  - Researchers should develop techniques to investigate utility
  - Practitioners could ask more about the utility of a system and about how users normally would solve a task

Main findings

Area of attention	*Main finding*	*N*	*Example of observations and quotes*
Analysis of results from a test session	Analysis is unstructured	9	Scattered fragments of analysis; no systematic approach used
	Analysis is incomplete	9	Does not identify causes or solutions; restricts discussion to user traits
	Analysis as a summary with the user	5	"Let's sum up"; selecting a few problems for further questioning; listing key findings
Confirmation of known issues as a test focus	Looking for known issues	6	"Now, I am just looking for ammunition"; develops ideas of problems before testing; tasks and questions designed to point out known issues
	Practitioners have foreseen problems	5	"We have a gut feeling", "I told you so"
Practical realities influencing tests	Technical problems	8	System breaking down; long response times in test environment; installation or security messages interrupt workflow
	Unfinished prototypes	6	Parts of prototype missing or inaccessible; "a log-in name should not be WaddleFish"; texts and pictures are wrong or out of date
Questions asked during a test	Problems are explained, not experienced	13	"Do you think you would go back to the front page"; "did you notice this column", "what do you expect to see"
	Leading questions	13	Questions address certain parts of GUI or system; evaluator hints the solution: "Can you do this in another way?"
	Unnecessary or obvious questions	10	"You did figure out to press the print button?"; asking user to locate information that clearly appears on the present screen; asking if user would like relevant information
Trying to meet laboratory-style scientific standards	Evaluators want similar conditions for users under test	6	"We have to make sure all users get the same questions"
	Rigid or artificial procedures	5	Laboratory style procedures; Danish evaluators speaking English to a Danish user; measuring subjective satisfaction overly systematic
Uncovering usability problems or utility concerns	User points to utility or lack thereof	10	"I would not do it like this"; user chooses to solve task without help of system
	Evaluator probes utility concerns	7	Asking about normal workflow; asking whether a task is realistic; "What would you typically do?"

Summary of real think-aloud tests

- Critical issues found simply by looking at how think aloud tests were conducted
- We believe these issues will diminish the utility and validity of the results generated

What makes for useful problems?

- Aims at characterizing which ways of describing usability problems that developers find useful
- Mainly about form of description but also about the kinds of problem described
- May help improve evaluations (short term), but also help align usability and design work

Hornbæk, K. & Frøkjær, E. (2005), "Comparing usability problems and redesign proposals as input to practical systems development", *ACM Conference on Human Factors in Computing Systems* (CHI 2005), 391-400

Hornbæk, K. & Frøkjær, E. (2006), "What kind of usability-problem description are useful for developers?", *Proceedings of the Annual Meeting of Human Factors and Ergonomics Society*, 2523-2527

Aim and motivation

- Motivation is threefold:
  - Few recommendations on how to report findings from a usability evaluation
  - Those recommendations are based on the opinions of usability specialists (e.g., Dumas, Molich & Jeffries 2004; Capra & Smith-Jackson 2005)
  - Improve usability evaluations' impact through studying how they are used

YOUR views on forms of description

- Usability problems in the handouts are taken from www.jobindex.dk
- Please try to come up with at least three kinds of description that you expect to be most useful to a developer?
- I will now keep quiet for about 10 minutes to while you discuss the description of the usability problems: feel free to discuss the problem descriptions with your neighbors

YOUR views on forms of description

Procedure

- 43 students evaluated a large web site with
  - Think aloud user testing (Molich 2003)
  - Metaphors of human thinking (Hornbæk & Frøkjær 2002)
- Four developers individually assessed the 619 problems identified on
  - "How useful is the problem in the further development of Jobindex?"

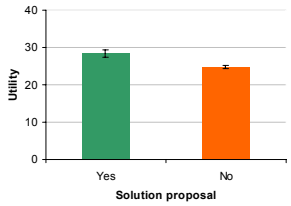
Not useful ⊢——————⊣ Very useful

Procedure, cont.

- Each problem was judged on five aspects:
  - Solution proposal
  - Persistence
  - Justified
  - Observable user action
  - Clarity
- Interrater agreement on these judgments was substantial

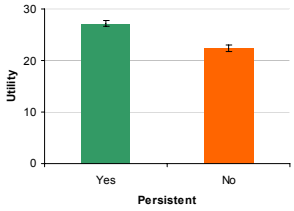
Results

- Solution proposals improve utility
- Unclear if solutions are used (cf. Hornbæk & Frøkjær 2005)



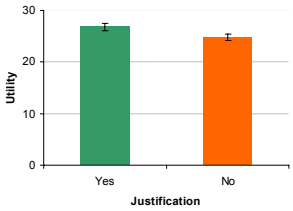
Results II

- Problems judged as persistent receive higher utility assessments by developers



Results III

- Problems that include justifications are assessed as being of significantly higher utility



Other differences

- Difference for clarity of problems
- No difference for observable user actions
  - In contrast to Hornbæk & Frøkjær (2005)
- Some differences between usability evaluation methods:
  - Think aloud tests did not produce more problems with observable user actions
  - The inspection technique generates more problems with a justification

Discussion and conclusion

- Mostly the effects are small
- Using developers' perceptions was useful
- Implications for how to describe problems
  - Include solution proposals
  - Justify *why* something is a problem
  - Make descriptions long enough to be understandable
  - Look for complex and persistent problems

How to give recommendations?

- Based on an interactive session from CHI 2007 made with:
  - Rolf Molich, DialogDesign (DK)
  - Jeff Johnson, UI Wizards Inc. (CA, USA)
  - Josephine Scott, TechSmith (MI, USA)
- with contributions by
  - Nina Vaught, Vaught Usability Group, Inc. (OR, USA)
  - Steve Krug, Advanced Common Sense (MI, USA)

http://doi.acm.org/10.1145/1240866.1240924

Pre-Conference Activities

- Six usability problems sent out to several discussion lists
- Request for recommendations
- Prize for best recommendations:
  A seat on the panel at CHI
- Panel based on 9 sets of recommendations from experienced professionals plus extensive discussions of a "Master Solution"

Useful and Usable Recommendations

- It seems reasonable to distinguish two dimensions of a professional usability recommendation:
  Useful (correct)
  Usable (comprehensible and constructive)

Useful Recommendation

- Solves the problem
- Implementable
- Appropriate

Usable Recommendation

- Clearly spelled out
- Unambiguous
- Easy to understand for target group

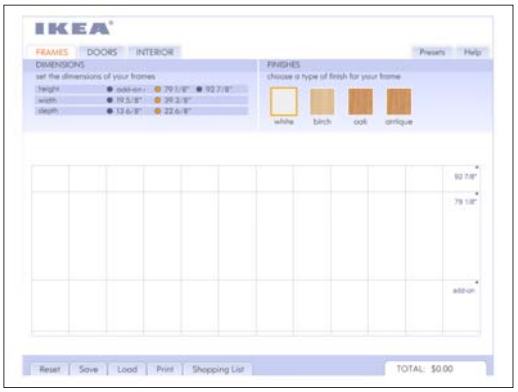
Useful and Usable Recommendations

- We rated all the recommendations (9*6 = 54) on these dimensions
- When we evaluate the usability of a recommendation, we first consider the recommendation fully useful and then judge how usable the recommendation is
- A recommendation that is not considered useful at all may still be fully usable, and vice versa
- We also take into consideration that a usable recommendation should be short

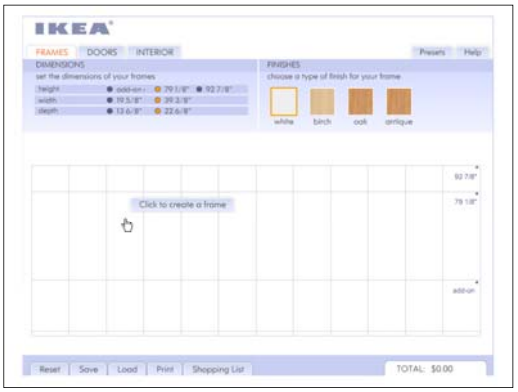
Usability Problems

- Three from www.IKEA-USA.com
  (CUE-5, evaluated by 13 professional teams)
- Three from www.Enterprise.com
  (CUE-6, evaluated by 13 professional teams)

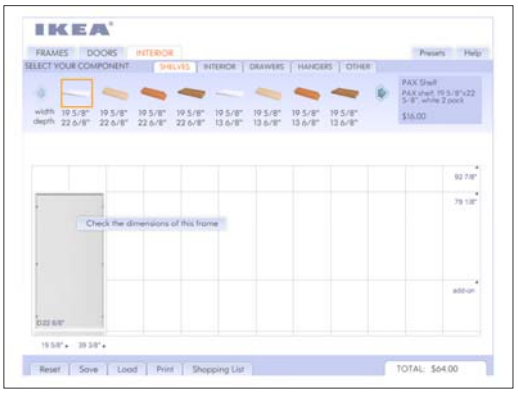
Usability problem 1: The first screen of the PAX Wardrobe planner. Assume that the user clicks *Create Your Own Solution.*



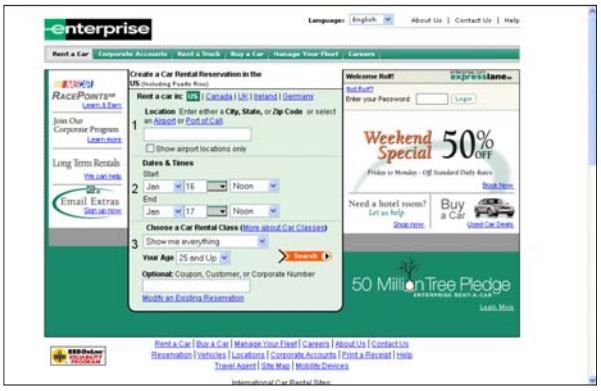
Usability problem 1: Users don't know what they should do first to start building a wardrobe.



Usability problem 1: Users don't know what they should do first to start building a wardrobe. The tooltip wasn't helpful.



Usability problem 2: Users do not realize that interior items, such as shelves, must match the dimensions of their wardrobe exactly. The message, "Check the dimensions of this frame", isn't helpful.



Usability problem 4: When asked to look for the best deal, users overlook specials ("Weekend Special"), thinking it is an ad. Most recall seeing it, but click on it only when prompted.

YOUR Recommendations, Please

- Hypothesis:
  - Recommendations are easy - when they are presented to you
  - They are much more difficult if you need to come up with them yourself

YOUR Recommendations, Please

- I will now keep quiet for about 10 minutes to give you a chance to sketch recommendations for the three usability problems I just presented (or all six, if you have the time)
- The usability problems are in the handouts
- Feel free to discuss your recommendations with your neighbors

Recommendations on Recommendations

- Useful recommendations
  - Analyze the problem(s) to be solved

Analyze the problem(s) to be solved

Usability problem 1.

The recommendation is to initially display a frame that users can modify and to add text hints about how to modify it.

The problem seems to be caused by a lack of clarity that building a wardrobe consists of three steps (selecting frames, doors and interior). Further, the options for dimensions and finish play a double role in describing both what a new frame will look like and what the current frame looks like. Finally, the grid at the lower part of the screen doesn't afford any actions; at first sight it is unclear why it is there.

Recommendations on Recommendations

- Useful recommendations
  - Analyze the problem(s) to be solved
  - Justify the recommendation

Recommendations on Recommendations

Usability Problem 2.

.....

This redesign will simplify navigation by reducing the number of selectable components and will also simplify the display by dispensing with the need for showing width/depth information.

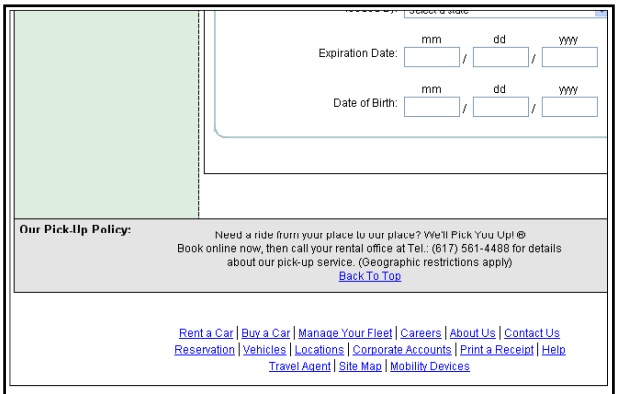
Recommendations on Recommendations

- Useful recommendations
  - Analyze the problem(s) to be solved
  - Justify the recommendation
  - Consider or raise questions about business goals

Recommendations on Recommendations

Given these three problems, the recommended solution depends on Enterprise's intent:

- If Enterprise wants the promotion to publicize the weekend rate, to induce users to rent on weekends when they otherwise would *not* have, then the weekend rate promotion should be moved into a small box in the left column, labeled "Weekend Deals". The size and appearance .....
- Consider whether users' tendency to click on the Weekend Specials promotion is the right measure of success of the promotion. Once a customer sees that Enterprise offers 50% off on weekends, why is there a need for them click on the promotion? Do they not get the weekend .....



Usability problem 5: An exact description of the procedures, benefits and limitations of the "We'll Pick You Up! ®" service is hard to find.

Recommendations on Recommendations

- Usable recommendations
  - Are specific, definite, concrete, actionable

Recommendations on Recommendations

Recommendations:

Place the "We'll Pick You Up" information more prominently.

For more impact, modify the message to read

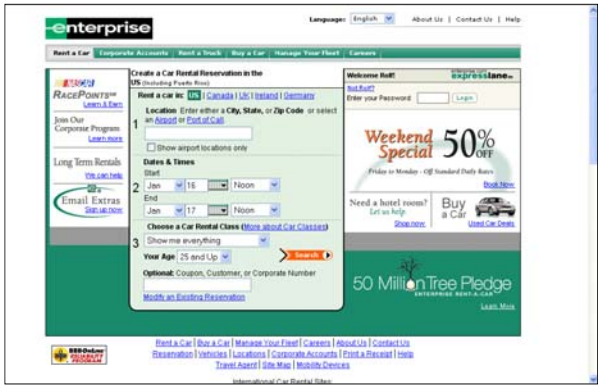
We'll pick you up for FREE *Read our pick up policy*

- Provide an online explanation of the pickup service. Make "We'll Pick You Up! " a link that displays the explanation in a pop-up window.
- To the right of "We'll Pick You Up! ®", add another link, "Learn more", that displays the same information.

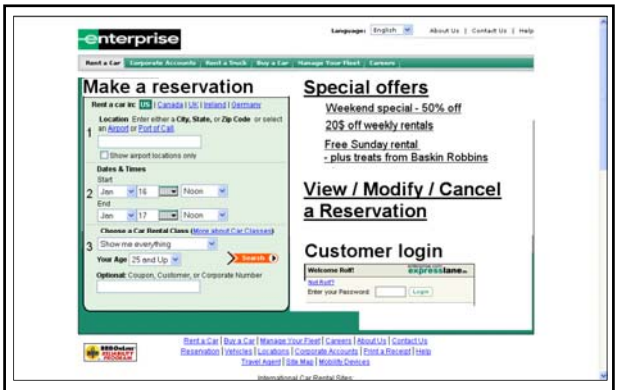
Usability problem 4: When asked to look for the best deal, users overlook specials ("Weekend Special"), thinking it is an ad. Most recall seeing it, but click on it only when prompted.

Insufficient detail:

Team B: “The way specials are promoted and used needs to be redesigned.”



Alternative a: Leave the home page as it is. The fact that users do not click the ad does not prove that they haven’t seen the message. Consider replacing “Book now” with a bigger “Learn more.”



Alternative b: Focus on the key tasks. The marketing messages dilute focus from the main task, which risks loss of business. Make the key user tasks highly visible as outlined above.

Recommendations on Recommendations

- Usable recommendations
  - Are specific, definite, concrete, actionable
  - Provide alternatives

Recommendations on Recommendations

Usability Problem 1:

Two alternative solutions, depending on resources available. Alternative 2 is preferred.

- Common to both alternatives: The word “frame” is too abstract; it is cabinet maker’s jargon. To customers, these are wardrobe compartments. Call them that, not “frames”.

Alternative 1 (small change):

...

Alternative 2 (larger change but preferable):

Summary on recommendations

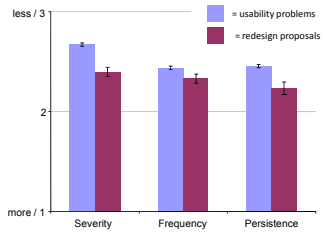
- From professional teams recommendations varied a lot: some were disappointing
- Crafting recommendations is a little like describing usability problems, but also a lot about design
- If you want to participate in follow up study, please e-mail me (kash@diku.dk)

Further data on recommendations

- In the Jobindex study, students also made redesign proposals
- A redesign proposal comprised:
  - (1) a summary,
  - (2) an argument why the redesign is important, including a description of the problem it attempts to solve,
  - (3) a detailed explanation of the proposed interaction design, and
  - (4) illustrations of how the redesign works.

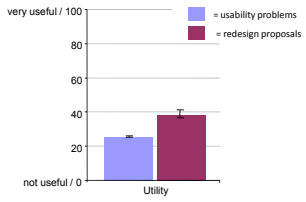
Results

- Redesign proposals assessed as more severe, frequent and persistent than usability problems



Results, cont.

- Redesign proposals seen as more useful
- Difference cannot be explained from the problems underlying the redesigns



Interviews: usability problems

- Developers already knew most of the problems
  - "there is not so much new in it"
- Use problems to support decisions and prioritize
  - "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 helps me prioritize what I should focus on"
- Developers noted shortcomings of problems, including lack of context and clear arguments
  - "so you present a problem, but what is the solution to that problem…sometimes you have, you have some alternatives […], but because there is a problem with one alternative then it is not sure that the other […] is better"

Interviews: redesign proposals

- Redesigns give ideas
  - "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"
- More concrete than usability problems
- More constructive than usability problems
  - "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"

Interviews: redesign proposals, cont.

- Sometimes redesigns were put aside
- but nevertheless found to be of utility
  - "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"
- Developers found both usability problems and redesign proposals useful
  - "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"

Summary

- Two important problems of usability research:
  - The ivory tower difficulty
  - The gulf between design and evaluation
- Adressed issues of
  - How to conduct tests
  - How to describe problems
  - How to give recommendations

- Further info: www.kasperhornbaek.dk or kash@diku.dk