

Direct Integration: Training Software Developers and Designers to Conduct Usability Evaluations

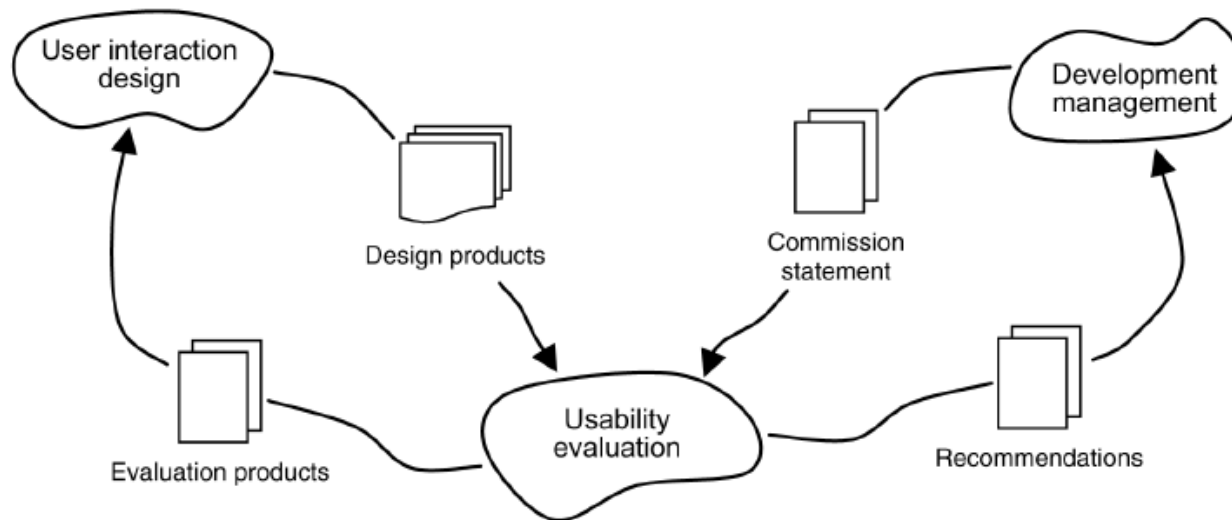
Mikael B. Skov
Jan Stage

Presented by:

Jan Stage
Aalborg University
Department of Computer Science
jans@cs.aau.dk

Usability Evaluation in Software Development

- **Usability evaluation and user interaction design are two separate activities in software development**



- **This separation is often carried through as a complete separation of work between evaluators and developers**

Improving the Interplay

- **Many problems in software development originate from this division of work**
- **It is often suggested that we need to bridge the gap or improve the interplay and many of the contributions here intend to do that**
- **Two general approaches:**
 - Better methods for usability evaluation to improve relevance
 - Better forms of feedback to improve impact
- **But ...**
- **... wouldn't it be much easier to avoid the gap in the first place**

Web Site Development

- **A large proportion of design and development projects**
- **A very diverse group of users with very little patience**
- **Projects with limited budgets and a very short time span**
- **Cannot afford a full-scale usability evaluation and cannot wait**
- **This is challenging usability engineering methods**

Do-It-Yourself

- **If you can't afford to have a garage built – what do you do?**


Hillhout Carport
Hillhout Carport 18x12 (3.53mx5.46m).

Great for to keeping your car "frost-free" in winter and nice and cool in the summer months.

Hillhout Carport features:

- The "passage" width is approx 9'8" (3.0m)
- The posts are 9 x 9 x 290cm
- The distance between the posts on each side is 180cm
- The roof is made of synthetic material. Inclusive of barge board
- Guttering and drain-pipe included
- Galvanised brackets
- Instruction manual included for easy self assembly

Panels displayed are not included.



Price incl. VAT £ 999.00 [Add to Basket](#)

- **Web site usability engineering: train the developers and designers to evaluate and improve their own products**

Training Course

- **Teach software developers and designers to conduct usability evaluations**
- **Provide participants with skills in formative usability evaluation**
- **No prerequisites**
- **It is done in a week**
- **Result: a usability report**

#	Lecture	Exercises
1	Introduction to the course and basic website technology	<p>Pilot test: Each team conducts simple pilot usability tests of websites to train their practical skills in usability evaluation. The teams choose the website themselves. Experience with conducting tests and the results achieved are discussed afterwards.</p>
2	Basic introduction to usability issues and guidelines for interaction design	
3	The think-aloud protocol and how to set up a test scenario. User groups and their different needs	
4	Application of questionnaires for collecting data and how to use different kinds of questions	
5	Computer architecture and website technology	<p>Usability evaluation: The teams conduct a usability evaluation of the Hotmail website according to a specification provided by the course instructors. The usability evaluations are conducted at the university in assigned rooms for each team. After the usability test sessions, the teams analyze the empirical data and make a usability report that describes the identified usability problems.</p>
6	Describing the usability testing method and how to collect and analyze empirical data	
7	Other usability evaluation methods and how to conduct a full-scale usability test session	
8	Website structures, information search and web surfing	
9	Guidelines for website design and principles for orientation and navigation	
10	Principles for visual design and different interaction styles	

Method (I)

- **Website: hotmail.com**
- **Participants:**
 - First-year university students
 - Four educations:
 - Informatics
 - Architecture and design
 - Planning and environment
 - Chartered surveyor
- **Setting: their offices**

Total number of students	Total number of teams	Team size <i>Average</i>	Team size <i>Min / Max</i>
234	36	6.5	4 / 8
Number of test subjects <i>Average</i>	Number of test subjects <i>Min / Max</i>	Age of test subjects <i>Average</i>	Age of test subjects <i>Min / Max</i>
3.6	2 / 5	21,2	19 / 30

Method (2)

- **Procedure:**

- Use the techniques of the course
- A two page scenario emphasizing areas of hotmail to evaluate
- Select test monitor and loggers (before the second class meeting)
- Prepare the tests
- Carry out tests with subjects (about 38 minutes for each)
- Evaluate and produce report

- **Data collection**

- The usability reports

Method (3)

Data analysis

- **Step 1:**

- 5 randomly selected reports were marked independently by two persons
- The criteria used for marking were negotiated into a set of 17 variables
 - Evaluation: The way the evaluation was conducted (3 variables)
 - Report: the presentation of the evaluation and the results (6 variables)
 - Result: the outcome of the usability evaluation (8 variables)

- **Step 2:**

- Individually marked each report on the 17 variables (a scale of 1 to 5)
- Counted usability problems

- **Step 3:**

- Comparison and negotiation of grades for the 17 variables and the identified usability problems

Results (I)

Teams	Evaluation		
	Conducting the evaluation	Task quality and relevance	Questionnaire/ Interviews
Student (N=36)	3.42 (0.73)	3.22 (1.05)	2.72 (1.00)
Professional (N=8)	4.38 (0.74)	3.13 (1.64)	3.50 (1.69)

- **The students did quite well in conducting the evaluation**
- **The professionals did significantly better**
- **On task quality and relevance the students seem to do better than the professionals (but not significant)**

Results (2)

Teams	Report					
	Test description	Data quality	Clarity of problem list	Executive summary	Clarity of report	Layout of report
Student (N=36)	3.03 (0.94)	3.19 (1.33)	2.53 (1.00)	2.39 (0.80)	2.97 (0.84)	2.94 (0.89)
Professional (N=8)	4.00 (1.31)	2.13 (0.83)	3.50 (0.93)	3.38 (1.06)	4.25 (0.71)	3.25 (0.71)

- **The students did well on describing the test and providing underlying data in appendices**
- **The worst student performance was in clarity of the problem list and executive summary**

Results (3)

Team	Results							
	Number of problems	Problem categorization	Practical relevance	Qualitative results overview	Quantitative results overview	Use of literature	Conclusion	Evaluation of test
Student (N=36)	2.56 (0.84)	2.06 (1.22)	3.03 (1.00)	3.03 (1.00)	2.28 (1.14)	3.08 (0.81)	2.64 (0.90)	2.44 (1.08)
Professional (N=8)	4.13 (1.13)	3.25 (1.75)	4.25 (1.49)	3.75 (1.16)	2.00 (1.51)	3.13 (0.35)	3.88 (0.64)	2.88 (1.13)

- **The students did poorly on problem identification and description**
- **Both groups did poorly in describing the quantitative results (efficiency and effectiveness)**

Limitations

- **The environment in which the evaluations were conducted was not optimal for the usability test sessions**
- **In some cases, slow Internet access**
- **Motivation and stress factors**
 - Is the motivation between students and professionals comparable
 - Would developers be (equally) motivated to do critical evaluations of their own product
- **The demographics of the test subjects are not varied with respect to age and education**

Conclusion

- **The gap between usability evaluation and software design will disappear if we can train software developers and designers to conduct usability evaluations**
- **We did a course that quickly teaches fundamental usability skills and explored it through a large empirical study**
- **The student teams were able to conduct usability evaluations and produce usability reports of a reasonable quality and with relevant results**
- **The student teams performed well in defining good tasks for the test subjects, and the data material in their reports was significantly better than the professionals**
- **They were less successful on most of the other variables**
- **They performed clearly worse on identification of problems and expressing the problems found in a manner relevant to practitioners**

Thank you ...

Guidance on Usability Problem Identification

Problem Severity	Delay in task completion	Irritation	Expectation to system behaviour
Small	Less than 30 seconds delay	Slight irritation	Minor difference in expected action
Medium	More than 30 seconds delay	Average irritation	Significant difference in expected action
Large	Could not complete the task	High irritation	Critical difference in expected action

Table 4. Guidelines used for categorizing the severity of usability problems.