



Said Fezazi – [saidfezazi@gmail.com](mailto:saidfezazi@gmail.com) – TU Berlin – SS 2008

# Usability Evaluation Considered Harmful *(Some of the time)*

*Seminar*

*22/04/2008*

# Overview

- Usability evaluation is today considered as a critical part of every design process.
- “considered harmful” -- > critical essay that advocates change
- Why is this a problem ?
- Illustration of the problems. Examples
- General suggestions
- Conclusions
- Other related work

# Motivation

- Why is this a problem ?
  - ▶ A lot of HCI related publications use UE (25% qualitative, 70% empirical)
  - ▶ Papers with qualitative UE methods are more severely judged
  - ▶ Results can be trivial / meaningless / kill innovative but young ideas
- Related work
  - ▶ Tyranny of evaluation (Henry Liebermann 2003), seeded the debate
  - ▶ Shumin Zhai defends evaluation methods, 2003 – Doing it is better than nothing
  - ▶ Cockton, 2007 – not enough methods to fit the different stages (inventor, artist, designer, optimizer)
  - ▶ Others raise questions about the methods themselves

## Description of approach : why and when is UE harmful ? (1/2)

- Blind push for usability evaluation in education, academia and industry
  - ▶ UE encouraged blindly by academia and review committees.
- “Scientific” evaluation methods might not imply science
  - ▶ UE as a weak science: the method forms the research question, and UE acts mostly as an existence proof (when comparing two systems).
  - ▶ Lack of replication.
  - ▶ “Science has one methodology, art and design have another. (...) Could there be a correlation between methodology and results?”

## Description of approach : why and when is UE harmful ? (2/2)

- Premature usability evaluation can eliminate promising ideas and the pursuit of promising ones
  - ▶ Sketches vs. Prototypes
  - ▶ UE favors “Getting the design right” instead of “Getting the right design”
  - ▶ UE tend to focus on the negative
  - ▶ “We converge on that which we can measure, which is what we are already familiar with. Our work degrades into a refinement of the known rather than innovation (...)”
- No input over the cultural adoption aspect
- No input over the *usefulness* (instead of usability)
- Poor usability does not translate into failure (ex: youtube)

# Discussion

- What to do then???
- ▶ Further develop the other aspects of user-centered design:
  - Understanding requirements
  - Consideration of cultural aspects
  - Developing and showing clients design alternatives
  - New interface possibilities through technical innovations
- ▶ As a community, stop insistence on UE
- ▶ Judge before using
- ▶ Recognize UE as a weak science, specially for major innovations
- ▶ Get inspiration from other disciplines on how they judge their designs (ex: architects)

# Contribution

- I can't see a huge contribution aside from ...
  - ▶ A message to the academia from a well known, respected field expert to make things change
  - ▶ Some examples of UE failures
  - ▶ A nicely written essay, but does this problem really deserve a paper ???

# Conclusions

- “Usability evaluation is not a universal panacea.”
- “The choice of evaluation methodology – if any – must arise from and be appropriate for the actual problem or research question under consideration”

## Questions for discussion

- “Science has one methodology, art and design have another. Are we surprised that art and design are remarkable for their creativity and innovation? While we pride our rigorous stance, we also bemoan the lack of design and innovation. Could there be a correlation between methodology and results?”
- Getting the design right vs. Getting the right design
- UE methods considered in this paper: laboratory based user observations, controlled user studies, inspection techniques.

# Usability testing (Wikipedia)

- **Goals of usability testing**
  - ▶ During usability testing, the aim is to observe people using the product to discover errors and areas of improvement. Usability testing generally involves measuring how well test subjects respond in four areas: efficiency, accuracy, recall, and emotional response. The results of the first test can be treated as a baseline or control measurement; all subsequent tests can then be compared to the baseline to indicate improvement.
- **Efficiency -- How long does it take people to complete basic tasks? (For example, find something to buy, create a new account, and order the item.)**
- **Accuracy -- How many mistakes did people make? (And were they fatal or recoverable with the right information?)**
- **Recall -- How much does the person remember afterwards or after periods of non-use?**
- **Emotional response -- How does the person feel about the tasks completed? Is the person confident, stressed? Would the user recommend this system to a friend?**