Two Challenges in Usability Research

Kasper Hornbæk
A problem has been detected and Windows has been shut down to prevent damage to your computer.

**DRIVER_IRQL_NOT_LESS_OR_EQUAL**

If this is the first time you've seen this stop error screen, restart your computer. If this screen appears again, follow these steps:

- Check to make sure any new hardware or software is properly installed.
- If this is a new installation, ask your hardware or software manufacturer for any Windows updates you might need.
- If problems continue, disable or remove any newly installed hardware or software. Disable BIOS memory options such as caching or shadowing. If you need to use Safe Mode to remove or disable components, restart your computer, press F8 to select Advanced Startup Options, and then select Safe Mode.

Technical information:

*** STOP: 0x00000000 (0x0000000c, 0x00000002, 0x00000000, 0x865aba)

*** gvt3.sys - Address F86b5aba base at F86b5000, DateStamp 3dd901eb

Beginning dump of physical memory
Physical memory dump complete.
Contact your system administrator or technical support group for further assistance.
• “the capability to be used by humans easily and effectively” (Shackel, 1991)
• “a set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users” (ISO 9126)
• “the effectiveness, efficiency, and satisfaction with which specified users can achieve goals in particular environments” (ISO 9241)
Challenge #1: How to find or discover usability problems?
Metaphors of human thinking (MOT)

- Human thinking as described by James (1890) & Naur (1995)
- Walkthrough typical tasks using five metaphors of human thinking
- Try to find mismatches between interface and descriptions of how we think


<table>
<thead>
<tr>
<th>Metaphor of human thinking</th>
<th>Implications for user interfaces</th>
<th>Key questions/Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habit formation is like a landscape eroded by water.</td>
<td>Support of existing habits and, when necessary, development of new ones.</td>
<td>Are existing habits supported? Can effective new habits be developed? Is the interface predictable?</td>
</tr>
<tr>
<td>Thinking as a stream of thought.</td>
<td>Users’ thinking should be supported by recognizability, stability and continuity.</td>
<td>Do the system make visible and easily accessible the important task objects and actions? Do the user interface make the system transparent or does non-task related information distract?</td>
</tr>
<tr>
<td>Awareness as a jumping octopus.</td>
<td>Support users’ associations with means of focusing within a stable context.</td>
<td>Do users associate interface elements with the actions and objects they represent? Can words in the interface be expected to create useful associations for the user?</td>
</tr>
<tr>
<td>Utterances as splashes over water.</td>
<td>Support changing and incomplete utterances.</td>
<td>Are alternative ways of expressing the same information available? Are interpretations of user input made clear? Do the system make a wider interpretation of user input than the user intends or is aware of?</td>
</tr>
<tr>
<td>Knowing as a site of buildings.</td>
<td>Users should not have to rely on complete or accurate knowledge, allow incompleteness.</td>
<td>Can the system be used without knowing every detail of it? Do more complex tasks build on the knowledge users have acquired from simpler tasks?</td>
</tr>
</tbody>
</table>
Are usability problems the right focus?

Limitations of focusing on problems

– Defect-identification typical for Westerners (deBono 1994)
  “This is a basic habit of Western thinking. If you can only get rid of faults everything else will be fine. There are two obvious dangers:
  1. We only focus our thinking on what is wrong.
  2. Getting rid of the faults in a poorly designed system does not result in a better-designed system.”

– Low practical significance (Wixon 2003)
– Poor way of communicating?

What about focusing on redesigns?

Evaluators 
(N <= 43)

Evaluation
N = 43

Redesign
N = 36

619 problems
57 redesigns

Developers 
(N = 4)

Assessment of problems & redesigns

Interviews

Utility = usability problems - redesign proposals

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”

Are usability evaluation methods the right focus?
Jeffries et al. (1991) wrote that ‘by definition, all of the problems found by heuristic evaluators were found by heuristic evaluation’

For more examples, see Hornbæk, K. (2010), "Dogmas in the Assessment of Usability Evaluation Methods", Behaviour & Information Technology, 29(1), 97-111

Participants find problems in a variety of ways

– “what if a problem can not be placed under a criterion?”

– “I am finding more usability problems as I fill out the problem lists. These are added to the problem list […]”
Challenge #2:
How to measure usability?
Why measures?

• What we mean by the term usability is to a large extent determined by how we measure it
• Usability cannot be directly measured
• Used for iteration (Gould & Lewis, 1985; Nielsen 1993)
Many, many, many measures

The system is …

I feel …
accessible, adequate, annoying, anxiety, appealing, boring, clear, cluttered, comfortable, competent, comprehensible, conclusive, confident, conflict, confusing, connected, convenient, desirable, difficult, disliked, dissatisfied, distracting, easy, effective, efficient, embarrassed, emotional, engaging, enjoyable, entertaining, enthusiasm, excellent, exciting, familiar, favorable, flexible, flustered, friendly, frustrating, fun, good, hate, helpfulness, immediate, important, improving, inefficient, intelligent, interested, intuitive, involved, irritation, learnable, likable, lively, loved, motivating, natural, nice, personal, plain, pleasant, preference, presence, productive, quality, quick, relevant, reliable, respect, responsive, satisfied, sensate, sense of being together; sense of control, sense of success, simple, smooth, sociable, social presence, stimulating, successful, sufficient, surprising, time consuming, timely, tiring, trust, uncomfortable, understand, useful, user friendly, vexed, vivid, warm, well-organized
<table>
<thead>
<tr>
<th>Usability aspects</th>
<th>Objective measures</th>
<th>Subjective measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes (effectiveness)</td>
<td>Expert assessment, comprehension</td>
<td>Users' perception of outcome</td>
</tr>
<tr>
<td>Interaction process (efficiency)</td>
<td>Time, usage patterns, learnability</td>
<td>Subjectively experienced duration, mental workload, perception of task difficulty</td>
</tr>
<tr>
<td>Users' attitudes and experiences (satisfaction)</td>
<td>Physiological usability, reflex responses</td>
<td>Validated questionnaires</td>
</tr>
</tbody>
</table>


---

Relations among measures?
Are measures correlated?

- Effectiveness vs. efficiency: \( r = .247 \pm .059 \)
- May be interpreted as:
  - 6% variance explained
  - small (~.1) to medium (~.3) effect (Cohen 1969)

<table>
<thead>
<tr>
<th>Source</th>
<th>Full papers</th>
<th>Candidates</th>
<th>Included with raw data</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM Conference on Human Factors in Computing Systems</td>
<td>261</td>
<td>94</td>
<td>59</td>
</tr>
<tr>
<td>ACM Transactions on Computer-Human Interaction</td>
<td>49</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>Annual Meeting of the Human Factors and Ergonomics Society</td>
<td>1165</td>
<td>99</td>
<td>14</td>
</tr>
<tr>
<td>Behaviour &amp; Information Technology</td>
<td>99</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>Human-Computer Interaction</td>
<td>36</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>IFIP TC13 International Conference on Human-Computer Interaction</td>
<td>152</td>
<td>49</td>
<td>10</td>
</tr>
<tr>
<td>International Journal on Human-Computer Studies</td>
<td>201</td>
<td>64</td>
<td>7</td>
</tr>
<tr>
<td>Interacting with Computers</td>
<td>127</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>2060</td>
<td>586</td>
<td>73</td>
</tr>
</tbody>
</table>

Table 1. Studies included in the meta-analysis distributed over sources (years 2003-2005).

Are measures correlated, cont.?

- Effectiveness vs. satisfaction: $r = 0.164 \pm 0.062$
- Preference is related to fewer errors
  - Prefer (13% errors) vs. do not prefer (18% errors)
- Six studies measure both effectiveness (objective) and participants’ assessment of their effectiveness (subjective)
  - “were your answers to tasks: very good – very poor” vs. errors in task answers
  - Correlations of these measures are not significantly different from zero

Are measures correlated, cont.?

- Task complexity does not influence the correlations

- More complex measures (e.g., quality of outcome) attenuate correlations
Toward a Theory of Usability?

- Describe and model experiences of use and summary attitudes
- Account for the relation between experiences and summary attitudes
- Integrate and account for both positive and negative experiences and attitudes
- Explain relations among measures; allow predictions of relations; help select measures
Problem counts in comparisons

• 44 novice evaluators were asked to conduct think aloud tests (Molich 2007)
• All evaluators received material describing business goals
• Evaluated an e-commerce site
Including Business Goals in evaluation

Results, company’s assessment

- Problems from Business Goal condition assessed as being of higher downstream utility