what & why usability specifications?

◆ a set of precisely stated usability objectives
  – usability attributes & measures that team agrees will reflect degree of success in usability engineering
◆ a management tool for usability engineering
  – sign-off up front about proposed metrics & levels
  – accompanied by agreed-upon process for achieving
◆ what they are not: a guarantee
  – easy to set usability goals that anyone could achieve!
  – ultimate usability depends on design/redesign; these just help you to see if you are converging as hoped

examples of usability attributes...
(think about your project and what you might want to measure)
from H & H’s calendar management example...

<table>
<thead>
<tr>
<th>Usability Attribute</th>
<th>Measuring Instrument</th>
<th>Value to be Measured</th>
<th>Current</th>
<th>Worst</th>
<th>Planned</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial performance</td>
<td>‘Add appointment’</td>
<td>Seconds to successfully add</td>
<td>15 sec</td>
<td>30 sec</td>
<td>20 sec</td>
<td>10 sec</td>
</tr>
<tr>
<td>(no online calendar use)</td>
<td>task per Benchmark 1</td>
<td>appointment on first trial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial performance</td>
<td>‘Delete appointment’</td>
<td>Number of errors on first</td>
<td>0 errors</td>
<td>2 errors</td>
<td>1 errors</td>
<td>0 errors</td>
</tr>
<tr>
<td>(regular use of competing</td>
<td>task per Benchmark 2</td>
<td>trial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>calendar)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First impression</td>
<td>Questionnaire</td>
<td>Average score</td>
<td>??</td>
<td>0</td>
<td>0.5</td>
<td>1.25</td>
</tr>
<tr>
<td>(adapted from QUIS)</td>
<td></td>
<td>(range -2 to 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

components of the spec

- **usability attribute**
  - perhaps qualified by user assumptions

- **measuring instrument**
  - benchmark tasks, questionnaire responses

- **value to be measured**

- **levels to be specified:**
  - current level
  - worst level
  - planned (target) level
  - best level
benchmark tasks

- typically emphasize core, representative functions
  - representative for attribute, eg., first-time use
  - perhaps also less frequent but highly critical tasks
  - usually single (or small set of) UI features; more complex scenarios should be broken down into subtasks

- try to cover a range of (relevant) usability attributes
  - initial performance, learnability, retainability, advanced feature usage, long-term use

- straightforward to set up; can be repeated reliably
  - specific about what to do, but not on how it is done

- yield unambiguous quantitative data (time/errors)

in phase 3: develop the benchmarks you’ll test in phase 4

text examples of benchmark tasks

(for Eudora mail package, initial performance)
In your In Mailbox, open the oldest message that has not yet been read.

(for Eudora mail package, learnability)
??

(for Eudora mail package, advanced features)
??

(for Eudora mail package, retainability)
??
questionnaires as measuring instruments

- need is for quantitative data regarding satisfaction
- classic example is Shneiderman’s QUIS:
  - Categories: screen layout, terminology and system information, learning, system capabilities
  - scalar judgment on specified dimension (not too large)
  - usually average (can also treat as ordinal)
- begin with demographic and user background info
  - documents assumptions about user classes
- often also includes other questions of interest
  - scales not part of specs, open-ended comments or suggestions at end of questionnaire, etc.

---

**Multi-point rating scale**

*The terms on the menus that I used to select objects and actions were:*

<table>
<thead>
<tr>
<th>very easy to understand</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>very hard to understand</th>
</tr>
</thead>
</table>

**Likert scale**

*The fields I used to enter information were arranged in an efficient manner.*

<table>
<thead>
<tr>
<th>strongly agree</th>
<th>agree</th>
<th>somewhat agree</th>
<th>neutral</th>
<th>somewhat disagree</th>
<th>disagree</th>
<th>strongly disagree</th>
</tr>
</thead>
</table>

**Semantic differential**

*Rate the system’s support for finding a specific item in the online catalog.*

<table>
<thead>
<tr>
<th>easy</th>
<th>clear</th>
<th>fun</th>
<th>difficult</th>
<th>confusing</th>
<th>boring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>extremely</td>
<td>quite</td>
<td>quite</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>slightly</td>
<td>neutral</td>
<td>extremely</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>slightly</td>
<td></td>
</tr>
</tbody>
</table>
Phase 3 documentation: format for usability specifications (5):

| Attribute: | <the characteristic of usability being engineered> |
| User assumptions: | <demographic or background characteristics> |
| Measuring instrument: | <the benchmark task> or <the rating(s) that will be used to assess subjective reactions>** |
| Value to be measured: | <the actual metric(s) that will be tracked> |
| Current level: | <best guess based on existing technology & tasks> |
| Worst acceptable level: | <minimum, clearly achievable: adequate> |
| Planned target level: | <ambitious but feasible: successful> |
| Best possible level: | <anticipated error-free expert value> |

*Include a range of attributes, but remember you need to test!!*

---

**final words—be careful...**

- validity of the specifications regarding usability
- measurability of the attributes
- specification of user classes
- reasonable levels for values measured
- scope (feasibility) of testing implied
- understanding and buy-in of project members
  - preferably all stakeholders, certainly management