Testing the UI

Part 1 of 2

This material has been developed by Georgia Tech HCI faculty, and continues to evolve. Contributors include Gregory Abowd, Jim Foley, Diane Gromala, Elizabeth Mynatt, Jeff Pierce, Colin Potts, Chris Shaw, John Stasko, and Bruce Walker. Permission is granted to use with acknowledgement for non-profit purposes. Last revision: March 2011 by Valerie Summet, Emory University
Agenda

• Goals for observation
• Usability specifications
• What to observe
  ▪ Think aloud
  ▪ Cooperative evaluation
  ▪ Performing Tasks
• Observation mechanisms
  ▪ Direct
  ▪ Recording by audio or video
  ▪ Computer logging
• Participants, IRB, & ethics
Observation - What, Why

- Watching users as they perform
- Summative or formative
  - Formative - early in design process, to help form the design
  - Summative - late in design process, to sum up
- Qualitative or quantitative
  - Qualitative - subjective
  - Quantitative - objectively measurable
Usability Specifications

“Is it good enough... ...to stop working on it? ...to get paid?”

How do we judge these things?
FIGURE 4.2
Usability.gov Step by Step Usability Guide. This guide shows all the steps from planning a usability test to performing the actual test and reporting the results.
Conducting an Evaluation

- Determine the performance measures
- Determine the tasks
- Develop the plan
- Ethics approval
- Recruit participants
- Collect the data
- Inspect & analyze the data
- Draw conclusions to resolve design problems
- Redesign and implement the revised interface
- Keep the designers quiet in the background!
Usability Specifications

- Quantitative usability goals, used as guide for knowing when interface is “good enough”
- Established early
  - Part of requirements specifications
    - At center of design contract
  - Evaluation is often used to demonstrate that design meets certain requirements
    - So the designer/developer can be paid :)
Formulating Specifications

- Better be more useful than this...
Measurement Process

• “If you can’t measure it, you can’t manage it”

• Need to keep gathering data on each iterative evaluation and refinement
• Compare benchmark task performance to specified levels
• Know when to get it out the door!
What is Included?

- Common usability attributes that are often captured in usability specs:
  - Initial performance
  - Long-term performance
  - Learnability
  - Retainability
  - Advanced feature usage
  - First impression
  - Long-term user satisfaction
# Usability Requirements

How will you judge whether your design meets the criteria?

<table>
<thead>
<tr>
<th>Usability attribute</th>
<th>Measure instrum.</th>
<th>Value to be meas.</th>
<th>Current level</th>
<th>Worst perf. level</th>
<th>Planned target level</th>
<th>Best poss level</th>
<th>Observ results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial perf</strong></td>
<td>Benchmark task</td>
<td>Length of time to successfully add appointment on the first trial</td>
<td>15 secs (manual)</td>
<td>30 secs</td>
<td>20 secs</td>
<td>10 secs</td>
<td></td>
</tr>
<tr>
<td><strong>First Impression</strong></td>
<td>Likert scale</td>
<td></td>
<td></td>
<td>-2</td>
<td>-2</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>-2 -1 0 1 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data

- Information gathered can be: *objective* or *subjective*
- Information also can be: *qualitative* or *quantitative*

Which are tougher to measure?
Data to Collect

- **Quantitative**
  - Time to complete task
  - Number or percentage of errors
    - Also affects time to complete
  - Percent of task completed in given time
  - Ratio of successes to failures
  - Number of commands used
  - Frequency of help usage

- **Typical users, typical training, typical tasks**
Data to Collect

- **Demographic**
  - Info about the participant, used for grouping or for correlation with other measures
    - e.g. handedness; age; first/best language; SAT score
    - Gather only what is relevant!

- **Qualitative**
  - Descriptions, observations that are not quantified
    - e.g. different ways of holding the mouse; approaches to solving problem; trouble understanding the instructions
One View of User Testing...

Evaluation can help your design...

THE PROJECT STATUS IS “YELLOW LIGHT.”

IN USER TESTS WE FOUND THAT THE PRODUCT LOCKS UP EVERY TWELVE SECONDS. THE INTERFACE IS INCOMPREHENSIBLE AND THE MANUAL IS PURE FICTION.

I THINK IT’S CLEAR WHAT WE NEED TO DO...

SHIP IT AND HOPE SOMEBODY WRITES A “DUMMIES” BOOK ABOUT IT?
The Task

- Essential use cases - gather quantitative data
- Other typical tasks - add breadth, can help understand process
- Tell them what to do, not how to do it
- Real people doing real tasks
- Issues:
  - Lab testing vs. field testing
  - Validity - typical users; typical tasks; typical setting?
  - Run pilot versions to shake out the bugs
Essential Use Cases

- **Specific, clearly stated task for users to carry out**
- **Example: Email handler**
  - “Find the message from Mary and reply with a response of ‘Tuesday morning at 11’.”
- **Users perform these under a variety of conditions and you measure performance**
For Project Part 3

- Aim for evaluation session to take no more than 15 minutes.
- Mix quantitative and qualitative assessments