Expert Reviews, Usability Testing, Surveys, Continuing Assessments

Some principles:

• “Focus on the users and their tasks, not the technology”
• “Conform to the users’ view of the task”
• “Don’t complicate the users’ task”
• “Think “outside-in”, not “inside-out””
• “Deliver information, not just data”
• “Try it out on users, then fix it!”

- Jeff Johnson, *GUI Bloopers: Don’ts and Do’s for Software Developers and Web Designers*, Morgan Kaufmann, 2001, 3rd edition (from Table of Contents)
Potential pitfalls

Designers
• entranced
• not objective
• not representative of users

Experienced designers
• recognize their limitations
• know “testing” necessary
Evaluation plan

Influenced by

• design stage
• novelty of project
• expected number of users
• criticality of H-C interface
• product cost
• available time
• experience of designers
• experience of evaluation team

Evaluate, assess (“test”)

• before
• during
• after
Testing

HCI testing ≠ Software testing

<table>
<thead>
<tr>
<th>HCI testing</th>
<th>Software testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>assesses general suitability of HCI</td>
<td>meets specification?</td>
</tr>
<tr>
<td>for human use</td>
<td>for system use</td>
</tr>
<tr>
<td>for “real world” task</td>
<td>for technical task</td>
</tr>
<tr>
<td>in organizational working environment</td>
<td>inside computer system</td>
</tr>
<tr>
<td>tests the design</td>
<td>tests the implementation</td>
</tr>
</tbody>
</table>
## Testing criteria

**HCI testing ≠ software testing**

<table>
<thead>
<tr>
<th>HCI</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>partially objective,</td>
<td>objective, specification</td>
</tr>
<tr>
<td>partially subjective</td>
<td></td>
</tr>
<tr>
<td>only sometimes yes-no</td>
<td>yes-no answer</td>
</tr>
<tr>
<td>many different types of</td>
<td>a few specialized, standardized types of</td>
</tr>
<tr>
<td>criteria</td>
<td>criteria</td>
</tr>
<tr>
<td>inherently partially</td>
<td>unambiguous, unique interpretation</td>
</tr>
<tr>
<td>ambiguous, involve taste,</td>
<td></td>
</tr>
<tr>
<td>opinion</td>
<td></td>
</tr>
<tr>
<td>not well defined</td>
<td>precisely defined</td>
</tr>
<tr>
<td>perfection in principle</td>
<td>perfection in principle possible</td>
</tr>
<tr>
<td>not possible</td>
<td>(sometimes difficult)</td>
</tr>
</tbody>
</table>
Testing

HCI testing validates against requirements, foreseen and unforeseen

Software testing verifies against concrete specification (as opposed to software validation)

HCI testing and Software testing test

• different things
• in different ways
• against different kinds of criteria
• with different results and different kinds of results
Stages of evaluation

- Informal experiments with self, colleagues
- Demonstrations to customers
- Expert reviews
- Design and test the usability test (pilot test)
- Usability tests (experiments)
- Surveys
- Experiments on customers ("beta" tests)
- Acceptance tests
- Reevaluation during active use (repeating some of the above)

All planned, results analyzed in detail
Informal experiments with self, colleagues

• simple
• inexpensive
• easy to do
• useful
• not very extensive
• not sufficient
Demonstrations to customers

- ad hoc
- usually limited number of customers
- representative?
- can be helpful
- cost-effectiveness?
Expert reviews

• experts simulate users
• formal and planned inspection, assessment
• one expert not enough
  different experts find different problems
• more expensive
• more effective
• reasonably but not completely
  representative of users
• cost-effectiveness good
• but still often not enough
Design and test the usability test (pilot test)

• usability test often required
• it must be designed and tested
• by experienced experts
• comparable to market tests
  user needs, reaction, perception
• goals: find flaws, statistically test differences (e.g. in solution alternatives)
• pilot test: test the test with fewer users than full scale usability test
• mock-ups (less expensive, quicker)
• modify plans for the usability test based on results of pilot test
Usability tests (experiments)

• full scale user tests
• with real potential users
• representative of distribution of user characteristics
• real usage environment
• experiences logged
• logs analyzed
• limitations:
  emphasizes first-time usage
  coverage of interface features limited
Surveys

• complement other evaluation methods
• clear goals important
• survey form must be carefully designed and prepared reviewed tested on small sample otherwise ineffective and costly
Experiments on customers ("beta" tests)

• basically: usability testing with actual, often paying customers
• large number of users/testers
• longer period of time
• systematic test?
• systematic analysis?
• typical user cross-section?
Acceptance tests

• interface meets specification?
• specification:
• measurable criteria, e.g.
  identifiable functionality
  user learning time
  speed of task performance
  user error rate
  user retention
  subjective user satisfaction
• specific, detailed criteria
• generalities inappropriate, useless
Reevaluation during active use

• user feedback, e.g.
  online
  telephone
  problem reports
  user groups, newsletters
• interviews
• user group discussions, panels
• identify desirable improvements
• logging usage and user experience (privacy?)