Usability Testing

Reading: *Don't Make Me Think! A Common Sense Approach to Web Usability*  
by S. Krug, Ch. 9-11

*Handbook of Usability Testing, 2ed*  
by J. Rubin / D. Chisnell, Ch. 2-5

slides created by Marty Stepp  
http://www.cs.washington.edu/403/
Usability testing

- **usability testing**: Evaluating a product by testing it on users.
  - usability has become a distinguishing factor for products (Apple)
  - focuses on individual usage, using the product to do something specific

- **focus group**: A group of people are asked about their perceptions, opinions, beliefs and attitudes towards a product, service, concept, advertisement, idea, or packaging.
  - focus group is a GROUP process; reactions; abstract; done early
Lack of usability testing

• Many companies don't usability test, or do it very little.
  – if done at all, often done with ~2 weeks left in development!

• Reasons given *not* to usability test:
  – not enough time
  – not enough money
  – no expertise in doing it
  – no lab or location in which to perform it
  – don't know how to interpret the results

• How do the authors respond to these criticisms?
Countering criticisms

• Why bother usability testing if we don't have time, money, or expertise to do very much of it?

• Why can't the developers just test the product by using it themselves and seeing what does / doesn't work for them?
  – developer knows product too well; can't see it like a newbie
  – even limited testing is better than none
  – few early tests are better than many late tests
  – ideally, usability testing is iterative; done over and over
## Inexpensive testing

<table>
<thead>
<tr>
<th></th>
<th>Traditional Testing</th>
<th>Lost-Our-Lease Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Users Per Test</strong></td>
<td>Usually eight or more to justify the set-up costs</td>
<td>Three or four</td>
</tr>
<tr>
<td><strong>Recruiting Effort</strong></td>
<td>Select carefully to match target audience</td>
<td>Grab some people. Almost anybody who uses the Web will do.</td>
</tr>
<tr>
<td><strong>Where to Test</strong></td>
<td>A usability lab, with an observation room and a one-way mirror</td>
<td>Any office or conference room</td>
</tr>
<tr>
<td><strong>Who Does the Testing</strong></td>
<td>An experienced usability professional</td>
<td>Any reasonably patient human being</td>
</tr>
<tr>
<td><strong>Advance Planning</strong></td>
<td>Tests have to be scheduled weeks in advance to reserve a usability lab and allow time for recruiting</td>
<td>Tests can be done almost any time, with little advance scheduling</td>
</tr>
<tr>
<td><strong>Preparation</strong></td>
<td>Draft, discuss, and revise a test protocol</td>
<td>Decide what you’re going to show</td>
</tr>
<tr>
<td><strong>What/When Do You Test?</strong></td>
<td>Unless you have a huge budget, put all your eggs in one basket and test once when the site is nearly complete</td>
<td>Run small tests continually throughout the development process</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>$5,000 to $15,000 (or more)</td>
<td>About $300 (a $50 to $100 stipend for each user and $20 for three hours of videotape)</td>
</tr>
<tr>
<td><strong>What Happens Afterwards</strong></td>
<td>A 20-page written report appears a week later, then the development team meets to decide what changes to make</td>
<td>Each observer writes one page of notes the day of the test. The development team can debrief the same day</td>
</tr>
</tbody>
</table>
When should a usability test be performed?

- best done **early** in the software lifecycle
- best done **often** / repeatedly

- type of test may vary depending on how far along the process you are
  - **early**: paper prototype
  - **middle**: compare UI designs
  - **later**: verify UI's usability

- can keep a historical record of usability results for each test
Benefits of multiple tests

**ONE TEST WITH 8 USERS**

Eight users may find more problems in a single test.

But the worst problems will usually keep them from getting far enough to encounter some others.

**TOTAL PROBLEMS FOUND: 5**

---

**TWO TESTS WITH 3 USERS**

First test

Three users may not find as many problems in a single test.

Second test

But in the second test, with the first set of problems fixed, they’ll find problems they couldn’t have seen in the first test.

**TOTAL PROBLEMS FOUND: 9**
Usability study room setup

• preferred: a quiet room with a computer and 2-3 chairs
  – **participant** sits at computer and performs tasks w/ product
  – **moderator / facilitator** guides the user through the process

  – others on dev team observe user, either from the side or from another room (preferred)
    • web cam, one-way mirror, etc.
    • record the user and watch later
Identifying participants

• Can I make my mother test our app?
Does it matter who the users are for a usability test?

– An ideal test has at least 3-4 users who have not been told much about the app beforehand.

– It doesn't matter much who you grab as your user; doesn't have to be just like a real user of the app.

– Everyone's a beginner in a way.
– It's bad to design a site that only experts can use.
– Experts don't mind something simple enough for beginners, so testing with beginners is not bad.

– UNLESS the app requires specific expert knowledge to use.
Facilitating a study

- Who is qualified to be a study facilitator?
- What things should / shouldn't a facilitator do?
  - Anybody with decent people skills can do it.
  - Be friendly.
  - Tell them it's okay to make mistakes; they aren't being tested.
  - Encourage them ask questions and to **think out loud**.
  - Don't lead the user or give them hints about what to do.
  - **Probe**; when they give feedback, ask for more details.
  - Don't appear to be concerned with note-taking or data gathering.
  - Don't be upset if the user fails or gets stuck.
  - Ask user questions when they get stuck.
    - "What are you thinking?" or "What are you trying to do now?"
  - **Tip:** Try taking the test yourself first.
Types of tests

- "get it" testing: Does user understand site's basic purpose?
  - "What do you think this page/site/app is about?"
  - "What do you think the ___ feature is for?"
  - Let them just click around for a while and play with the app.

- "key task" testing: Ask user to do a specific thing, and watch to see how they do.
  - "Your goal is to purchase a book about sailing for under $15."
  - "Change your buddy list preferences to block Amanda."
Another categorization

- **exploratory/formative** - high-level design concepts
  - Can user "walk up and use" it, and see value in the product?

- **assessment/summative** - lower-level operations (later)
  - user performs actual tasks, not vague goals; less moderated

- **comparison test** - match up different prototypes or designs
  - perf data are gathered for each design and compared
  - idea: test a competitor's site, see what they do/don't like
  - "best" design may turn out to be a hybrid of the available choices

- **verification test** - verifies that UI is okay or that a fix works
  - done with an actual product, not just a paper prototype
  - performance expectations are decided and measured
## A usability test plan

<table>
<thead>
<tr>
<th>WHAT TO TEST</th>
<th>PLANNING</th>
<th>ROUGH SKETCHES</th>
<th>PAGE DESIGNS</th>
<th>PROTOTYPE</th>
<th>FIRST USABLE VERSION</th>
<th>“CUBICLE TESTS”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competitors’ sites</td>
<td>Sketch of Home page Names of top level categories and site features</td>
<td>Home page Second-level page template Content page template</td>
<td>As much as you have working</td>
<td>As much as you have working</td>
<td>Each unique page</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FORMAT</th>
<th>LIVE SITE</th>
<th>PAPER</th>
<th>PAPER</th>
<th>HTML prototype</th>
<th>LIVE SITE</th>
<th>HTML page</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>HOW TO TEST</th>
<th>“Get it” Key tasks</th>
<th>“Get it” Names of things</th>
<th>“Get it” Basic navigation</th>
<th>“Get it” Key tasks</th>
<th>“Get it” Key tasks</th>
<th>Key tasks</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>WHAT YOU'RE LOOKING FOR</th>
<th>What do they like/love? How does it fit into their lives? What works well? How hard is it to do key tasks?</th>
<th>Do they get the point of the site? Does it seem like what they need?</th>
<th>Do they get the point of the site? Do they get the navigation? Can they guess where to find things?</th>
<th>Do they still get it? Can they accomplish the key tasks?</th>
<th>Do they still get it? Can they accomplish the key tasks?</th>
<th>Can they accomplish the key tasks?</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>SESSION LENGTH</th>
<th>1 hr.</th>
<th>15-20 min.</th>
<th>15-20 min.</th>
<th>45 min.-1 hr.</th>
<th>1 hr.</th>
<th>5 min. per page</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th># OF TESTS</th>
<th>1</th>
<th>1-3</th>
<th>1-3</th>
<th>1-3</th>
<th>1-3</th>
<th>1 per page</th>
</tr>
</thead>
</table>


Another usability test plan

1. type of test
2. purpose/goals/objectives
3. participant characteristics
4. task list
   (possibly have users try same tasks in different orders)
5. test environment / equipment
6. moderator's role
7. evaluation metrics and data to be collected
8. report
Observing a test

- Should be out of view/room if possible.
- What should observers look for?
  - Does the user "get it"?
  - Can they find their way around the site?
  - How long (time, number of clicks) does it take them?
  - Do they do any "head-slapping" or shocking things?
  - What do the users like and dislike most about the experience?
  - Watch what users do when they get stuck.
    - Do they look for help?
    - Do they re-read the page carefully?
    - Do they go back?
    - Do they just stop and look at the moderator?
    - Do they start wandering the app looking for the feature?
Gather and interpret data

• Observers write down usability test notes.
  – Team looks over notes and decides what to change.
  – *Tip:* Value the user's actions and explanations over opinions.
  – *Tip:* Don't always listen to user suggestions for new features.
  – It can be hard to figure out HOW to fix problems.

• Small changes (tweaks) are often better than huge changes.
  – Often removing or simplifying is better than adding.
  – First fix things that are easy or get the most bang for the buck.
Possible data to collect

- Number/percentage of:
  - tasks completed correctly with/without prompts or assistance
  - tasks completed incorrectly

- Count of:
  - incorrect selections (errors)
  - errors of omission
  - incorrect menu choices
  - incorrect icons selected
  - visits to the help file, index, table of contents, etc.
  - negative comments or mannerisms

- Time required to:
  - access information in online help
  - recover from error(s)
  - complete each task
Performance goals

• Some tests have specific performance goals.
  – Decide **specific goals** you want users to achieve.
    • "At least 2/3 of users will be able to find and change their Network Settings in < 5 minutes."
    • "Every user will be able to find/use the navigation bar."

  – In these tests, moderator offers less guidance and fewer hints
  – Less emphasis on "thinking out loud"
    • Thinking out loud slows users down; they may not finish in time.

  – Alternatives to thinking out loud:
    • Replay steps with user after the test is over.
    • Get them to talk about what they did and why they did it.
    • Hook up electrodes to their brain and measure the current.  (no)
Goals and objectives

• bad goals:
  – Is the product usable?          (too vague / nebulous)
  – Is it ready for release?       (users can't determine that)
  – Is this a good product?        (too subjective)

• better goals:
  – Can users tell what is / isn't clickable?
  – Where do users go to search the site?
  – How easily do users find new products on the site to purchase?
  – Do users use the toolbar at the top of the screen?
  – What do users click to find help when they are stuck?
Users fail

• Typical ways users fail in usability tests:
  – Don't understand the point of the site.
  – They use different vocabulary than you, so they can't find a word for the action to do.
  – Their notion of how to categorize is different.
  – Site is too busy / cluttered.
  – Not clear what the options are on the screen.

• If user momentarily gets stuck or goes astray, that CAN be okay.
  – A "kayak" problem; the boat can right itself.
  – Give them a chance to temporary fail and then recover.
Limitations of usability tests

- Somewhat artificial
- Test results don't prove that a product/design/UI "works"
- Testing may not be the best use of your time.
  - Maybe have a usability expert look at it, to find gross violations.
- It's possible that a UI has an initial learning curve, but is then very powerful/usable. A usability test doesn't measure that.
- Doesn't tell you if the market wants/needs a product like yours.
  - A focus group or survey would be better for that.