Improving Government Websites and Surveys with Usability Testing
a comparison of methodologies

Jen Romano Bergstrom & Jon Strohl

FCSM | Washington, DC
@forsmarshgroup
About this talk

1. Assessing the user experience is important.
2. There are many ways you can do this.
What is Usability?

“the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.” ISO 9241-11
Usability vs. User Experience (UX)?

Whitney’s 5 Es of Usability

Peter’s User Experience Honeycomb

The 5 Es to Understanding Users (W. Quesenbery): http://www.wqusability.com/articles/getting-started.html
What People do on the Web

What We Design For...

Read

Read

Read

Read

[Pause for reflection]

Finally, click on a carefully chosen link

The Reality...

Look around feverishly for anything that

a) is interesting or vaguely resembles what you're looking for, and

b) is clickable.

As soon as you find a halfway-decent match, click.

If it doesn’t pan out, click the Back button and try again.
Mental models and repeating behaviors
Measuring the UX

“the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.” ISO 9241-11

+ emotions

• How does it work for the end user?
• What does the user expect?
• How does it make the user feel?
Where to test

LABORATORY
• Controlled environment
• All participants have the same experience
• Record and communicate from control room
• Observers watch from control room and provide additional probes (via moderator) in real time
• Incorporate physiological measures (e.g., eye tracking, EDA)
• No travel costs

REMOTE
• Participants in their natural environments (e.g., home, work)
• Use video chat (moderated sessions) or online programs (unmoderated)
• Conduct many sessions quickly
• Recruit participants in many locations (e.g., states, countries)

IN THE FIELD
• Participants tend to be more comfortable in their natural environments
• Recruit hard-to-reach populations (e.g., children, doctors)
• Moderator travels to various locations
• Bring equipment (e.g., eye tracker)
• Natural observations
How to test

ONE-ON-ONE SESSIONS
• In-depth feedback from each participant
• No group think
• Can allow participants to take their own route and explore freely
• No interference
• Remote in participant’s environment
• Flexible scheduling
• Qualitative and Quantitative

FOCUS GROUPS
• Participants may be more comfortable with others
• Interview many people quickly
• Opinions collide
• Peer review
• Qualitative

SURVEYS
• Representative
• Large sample sizes
• Collect a lot of data quickly
• No interviewer bias
• No scheduling sessions
• Quantitative analysis
When to test

- Concept
- Content Strategy
- Conceptual Design
- Clickable Prototypes
- Final Product

Test

@forsmarshgroup
What to measure

EXPLICIT
+ Post-task satisfaction questionnaires
+ In-session difficulty ratings
  + Verbal responses
  + Moderator follow up
  + Real-time +/- dial

IMPLICIT
+ Facial expression analysis
  + Eye tracking
  + Electrodermal activity (EDA)
  + Behavioral analysis
+ Linguistic analysis of verbalizations
  + Implicit associations
  + Pupil dilation

OBSERVATIONAL
+ Ethnography
+ Time to complete task
  + Reaction time
+ Selection/click behavior
+ Ability to complete tasks
  + Accuracy

@forsmarshgroup
Case 1: Department of Veterans Affairs

- **Problems:** What do users want? Does the new design work?
- **Methods:** Focus groups, one-on-one interviews, in-lab usability testing with eye tracking

<table>
<thead>
<tr>
<th>Steps to Complete Task*</th>
<th>Time to Complete Task*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>8</td>
</tr>
<tr>
<td>10%</td>
<td>170 seconds</td>
</tr>
<tr>
<td>21%</td>
<td>32 seconds</td>
</tr>
<tr>
<td>15%</td>
<td>101 seconds</td>
</tr>
</tbody>
</table>

1. Participant repeatedly fixated the upper right hand corner. Participant said that he/she was looking for a search tool on the page. The search tool was in a disappearing banner on the page.
2. Participants had similar fixation counts across bottom links, indicating uncertainty of where to click to get started.
Case 1: Department of Veterans Affairs

• Focus Groups, early in development
  – What people want and expect from this type of site
• Usability testing, throughout development
  – What works well on the new site and what needs work
  – Performance data (time on task, accuracy, steps)
  – What people explore first (first click)
• Eye tracking, final product
  – What attracts attention
  – Where people look for information, repeatedly
Case 2: Internal Revenue Service

- Problem: What parts of the form do people actually read?
- Method: In-lab usability testing with eye tracking
Case 2: Internal Revenue Service

Participants did not read the instructions in their entirety (page 1: left; page 3: right); rather, they skimmed and then moved on to the form where they needed to enter information.

Aggregate fixation count heat map across all participants, Page 1. Participants looked at ‘Purpose of Form’ section the most often.

Length of time spent on each page of the instructions before working on form.
Case 2: Internal Revenue Service

- Usability testing, final product: inform about redesign
  - What people think of the instructions
  - What parts of the form are unclear
  - How they complete the form
- Eye tracking
  - How much of the instructions people read
  - Which sections, in particular
  - Where they drop off
  - When they flip back, what they read
Case 3: Department of Defense

• Problem: Can people complete the survey as intended?
• Method: In-lab usability testing
Case 3: Department of Defense

<table>
<thead>
<tr>
<th>Activity</th>
<th>Scale</th>
<th>Default Selection</th>
<th>Nonresponse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find out more about something or someone</td>
<td>1 Never</td>
<td>within the first box</td>
<td>unclear</td>
</tr>
<tr>
<td>Share ideas/information with peers to advance a topic</td>
<td>1 Never</td>
<td>unclear</td>
<td>nonresponses</td>
</tr>
<tr>
<td>Laugh</td>
<td>1 Never</td>
<td>unclear</td>
<td>confusing; bottom = low value</td>
</tr>
<tr>
<td>Teach others about something you have learned</td>
<td>1 Never</td>
<td>unclear</td>
<td></td>
</tr>
<tr>
<td>Get better at doing something</td>
<td>1 Never</td>
<td>unclear</td>
<td></td>
</tr>
<tr>
<td>Give updates throughout the day</td>
<td>1 Never</td>
<td>unclear</td>
<td></td>
</tr>
<tr>
<td>Have fun</td>
<td>1 Never</td>
<td>unclear</td>
<td></td>
</tr>
<tr>
<td>See/hear something entertaining</td>
<td>1 Never</td>
<td>unclear</td>
<td></td>
</tr>
<tr>
<td>Post about fun things you are doing</td>
<td>1 Never</td>
<td>unclear</td>
<td></td>
</tr>
<tr>
<td>Find places to eat or services to use</td>
<td>1 Never</td>
<td>unclear</td>
<td></td>
</tr>
<tr>
<td>Work together toward a shared goal</td>
<td>1 Never</td>
<td>unclear</td>
<td></td>
</tr>
</tbody>
</table>

Switching from a left-to-right rating scale to a top-to-bottom rating scale was confusing; bottom = low value.
Case 4: Department of Defense

- **Problem:** What difficulties do respondents have with the survey?
- **Method:** In-the-field ethnography
Case 4: Department of Defense

- Ethnography
  - Respondents had many distractions
  - Respondents completed it in more than one sitting
  - It took a lot longer than expected
Case 5: National Institute of Standards and Technology

- **Problem:** What makes a password secure and usable?
- **Method:** In-lab user UX study on tablet and smartphone
Case 5: National Institute of Standards and Technology

- Usability testing
  - People had greater difficulties with passwords that required multiple screens
  - Number of screens was more impactful than number of characters
Case 6: National Cancer Institute

- **Problems:** What do users want? Do the labels make sense? Which designs work best?
- **Methods:** In-lab and remote usability testing
Case 6: National Cancer Institute

- In-lab usability testing
  - Sessions with survivors and supporters
- Remote testing
  - Sessions with oncologists, advocates, and PIs
- Feedback from all user groups
- Redesigns were better informed
Conclusion
Better UX means...

- Higher user satisfaction
  - Increased efficiency and accuracy
  - Repeat visits and recommendations
- Decreased costs for the organization
  - Reduce call center phone calls and staffing
- Data you can trust
  - Empirically tested products
- From the end users’ perspective
It’s up to you.

- Throughout product development
- Any size budget
- You have the power to improve your product
Thank you!

- Twitter: @forsmarshgroup
- LinkedIn: http://www.linkedin.com/company/fors-marsh-group
- Blog: www.forsmarshgroup.com/index.php/blog

Jennifer Romano Bergstrom
@romanocog
jbergstrom@forsmarshgroup.com

FCSM 2013 | Washington, DC