CSE440: Introduction to HCI
Methods for Design, Prototyping and Evaluating User Interaction

Lecture 11: Testing
Nigini Oliveira
Manaswi Saha
Liang He
Jian Li Zheng
Jeremy Viny
What we will do today

Design sprint

Testing (beyond paper prototypes)
But first a reminder
But first a reminder

Some of you decided on one idea too fast!
But first a reminder

Some of you are not deciding on one idea fast enough.
But first a reminder

Make sure to meet in your team and decide!
You still have time to adjust/explain things in your final report.
Upcoming

<table>
<thead>
<tr>
<th>Oct 29</th>
<th>Oct 30</th>
<th>Oct 31</th>
<th>Nov 1</th>
<th>Nov 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading 2: Storyboarding and Video Prototyping</td>
<td>Paper Prototyping</td>
<td>2g - Design Review</td>
<td>Testing</td>
<td>Section</td>
</tr>
<tr>
<td>11:30 - 12:50</td>
<td>11:30 - 12:50</td>
<td>11:30 - 12:50</td>
<td>11:30 - 12:50</td>
<td>10:30 - 11:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nigini's office hours</td>
<td>11:30 - 12:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10:00 - 12:00</td>
<td>1:30 - 2:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CSE 218</td>
<td>2:30 - 3:20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nov 5</th>
<th>Nov 6</th>
<th>Nov 7</th>
<th>Nov 8</th>
<th>Nov 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nigini at CSCW</td>
<td>Nigini at CSCW</td>
<td>Nigini at CSCW</td>
<td>Nigini at CSCW</td>
<td>Nigini at CSCW</td>
</tr>
<tr>
<td>3a - Paper Prototype</td>
<td>Heuristic Evaluation</td>
<td>3b - Heuristic Evaluation</td>
<td>Accessibility</td>
<td>Section</td>
</tr>
<tr>
<td></td>
<td>11:30 - 12:50</td>
<td>11:30 - 12:50</td>
<td>11:30 - 12:50</td>
<td>10:30 - 11:20</td>
</tr>
<tr>
<td></td>
<td>OUG 141</td>
<td>OUG 141</td>
<td>OUG 141</td>
<td>11:30 - 12:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1:30 - 2:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2:30 - 3:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3c - Usability Testing Check-In</td>
</tr>
</tbody>
</table>
# Upcoming

<table>
<thead>
<tr>
<th>Oct 29</th>
<th>Oct 30</th>
<th>Oct 31</th>
<th>Nov 1</th>
<th>Nov 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>Paper Prototyping</td>
<td>2g - Design Review</td>
<td>Testing</td>
<td>Section</td>
</tr>
<tr>
<td>2: Storyboarding and Video</td>
<td>11:30 - 12:50</td>
<td>OUG 141</td>
<td>11:30 - 12:50</td>
<td>OUG 141</td>
</tr>
<tr>
<td>Prototyping</td>
<td></td>
<td>Nigini's office hours</td>
<td></td>
<td>10:30 - 11:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10:00 - 12:00</td>
<td></td>
<td>11:30 - 12:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CSE 218</td>
<td></td>
<td>1:30 - 2:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2:30 - 3:20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2h - Getting the Right Design</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nov 5</th>
<th>Nov 6</th>
<th>Nov 7</th>
<th>Nov 8</th>
<th>Nov 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nigini</td>
<td>Nigini at CSCW</td>
<td>Nigini at CSCW</td>
<td>Nigini at CSCW</td>
<td>Nigini at CSCW</td>
</tr>
<tr>
<td>at CSCW</td>
<td>Heuristic Evaluation</td>
<td>Heuristic Evaluation</td>
<td>Accessibility</td>
<td></td>
</tr>
<tr>
<td>3a - Paper Prototype</td>
<td>11:30 - 12:50</td>
<td>OUG 141</td>
<td>11:30 - 12:50</td>
<td>OUG 141</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Reading 3: Usability Testing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3c - Usability Testing Check-In</td>
</tr>
</tbody>
</table>
Design Sprint
Design Sprint (30 minutes)

Form groups of 3
(find at least one member who is not usually in your usual groups)

Read through the handout

5 minutes: Sketch out a few designs and decide on one
15 minutes: Prototype your design
5 minutes: Test your design with one of the TAs, myself, or with a member of another group
Design Sprint (30 minutes)

Form groups of 3
(find at least one member who is not usually in your usual groups)

Read through the handout

5 minutes: Sketch out a few designs and decide on one
15 minutes: Prototype your design
5 minutes: Test your design with one of the TAs, myself, or with a member of another group
Design Sprint (30 minutes)

Form groups of 3
(find at least one member who is not usually in your usual groups)

Read through the handout

5 minutes: Sketch out a few designs and decide on one
15 minutes: Prototype your design
5 minutes: Test your design with one of the TAs, myself, or with a member of another group
Design Sprint (30 minutes)

Form groups of 3
(find at least one member who is not usually in your usual groups)

Read through the handout

5 minutes: Sketch out a few designs and decide on one
15 minutes: Prototype your design
5 minutes: Test your design with one of the TAs, myself, or with a member of another group
Design Sprint (30 minutes)

Form groups of 3
(find at least one member who is not usually in your usual groups)

Read through the handout

5 minutes: Sketch out a few designs and decide on one
15 minutes: Prototype your design
5 minutes: Test your design with one of the TAs, myself, or with a member of another group
Discussion

What worked well during the paper prototyping?

What did you learn about coffee ordering?

How would you change your design?

What did you come up with?
Testing
Testing

User tests vs. Usability tests

Different ways for gaining insights
- Quick and dirty: informal discussions with users
  sometimes done using paper prototypes/Wizard-of-Oz testing
- Usability tests in lab
  e.g., Human Task Performance Measures
- Remote usability tests
  e.g., using recording software
- A/B testing
Remote Usability Testing

Conferencing-based testing
  Use tools like video conferencing, instant messaging, and screencasting to test with a remote participant

Unmoderated remote testing
  Automatic logging and some analysis of usage

Controlled online A/B experiments
  Carefully measure results of showing different versions to different sets of live customers
Unmoderated Remote Usability Test

Move usability testing online

- participants access the “lab” via web
- answer questions & complete tasks in “survey”
- records actions or screens for playback
- can test many people completing many tasks

Analyze data individually or in aggregate

- playback individual sessions
- find general problem areas
- if needed, look more closely with traditional methods

More: https://www.nngroup.com/articles/unmoderated-user-testing-tools/
Unmoderated Remote Usability Test
Unmoderated Remote Usability Test

https://www.loop11.com/
Unmoderated Remote Usability Test
Controlled A/B Experiments

Many names for concept

A/B tests or
Control/Treatment

Randomized
Experimental Design

Controlled experiments

Split testing

Parallel flights
Controlled A/B Experiments

Example: Amazon Shopping Cart Recommendations

Add an item to your shopping cart
Most sites show the cart

At Amazon, Greg Linden had idea to show recommendations based on cart items
Controlled A/B Experiments

Evaluation
   Pro: cross-sell more items
   Con: distract people from checking out

Highest Paid Person’s Opinion:
   Stop the project

Simple experiment run:
   Wildly successful
Checkout Page

Conversion rate is percentage of visits that include purchase

Which version has a higher conversion rate?
Checkout Page

*Conversion rate* is percentage of visits that include purchase

Which version has a higher conversion rate?
Checkout Page

Conversion rate is percentage of visits that include purchase

Which version has a higher conversion rate?

Coupon Code decreases by factor of 10
Office Online Feedback

Which one has a higher response rate? By how much?
Which one has a higher response rate? By how much? B gets more than double...
Another Feedback Variant

Call this variant C. Like B, also two stage. Which one has a higher response rate, B or C?
Another Feedback Variant

Call this variant C. Like B, also two stage.
Which one has a higher response rate, B or C?

C outperforms B by a factor of 3.5
Obama Campaign

http://o04151985.kylerush.net/blog/optimization-at-the-obama-campaign-ab-testing/
Obama Campaign

http://o04151985.kylerush.net/blog/optimization-at-the-obama-campaign-ab-testing/
Romney Campaign

http://o04151985.kylerush.net/blog/optimization-at-the-obama-campaign-ab-testing/
Data-Driven Methods Not Just Online

1999
Made from cardboard, the first Netflix mailer weighs more than an ounce. But with only 100,000 customers, reducing material and shipping costs is not yet a priority for the company.

2000
Thick paper replaces cardboard. DVDs are inserted and removed from the top rather than the side.

2000
Full-color printing is introduced. Top-loading is abandoned in favor of side-loading, which is judged more convenient.
Data-Driven Methods Not Just Online
Data-Driven Methods Not Just Online

2001
Foam padding is dropped because the benefits don't justify the cost. The company gives top-loading another try.

2001
Marking a return to side-loading, this mailer is a direct ancestor of the one the company uses today.

2003
Instead of sealing the entire top and bottom, Netflix introduces a circular sticker, affixed only on the top.

2004
A window shows the disc barcode. Speculation is that this enables storing discs in mailers prior to shipping.
Limitations

Drives hill-climbing, but not overall design
   A design may be better, but is it good?

Impossible for new designs to compete

Can be difficult to scale to many features
Ask me something!
Ideation and Evaluation

Case Video:
Phone Button Lay-Out

http://www.youtube.com/watch?v=kCSzjExvbTQ