Usability Testing:
What Have We Overlooked?

Hyoungtae Cho
What is usability?

• “How well users can learn and use a product to achieve their goals and how satisfied they are with that process”

• Cannot be directly measured; quantified by means of indirect measures or attributes:
  – The number of reported problems
  – Ease of learning
  – Efficiency of use
  – Memorability
  – Error frequency and severity
  – Subjective satisfaction
Terminology

• Usability Testing: to evaluate a product by testing it on representative users from the target audience through specific tasks
• Expert Review: performed by usability experts to identify potential usability problems
• Return On Investment (ROI) of Usability: The returns from usability improvement against added efforts/costs (time & money)
# Usability Testing vs. Expert Review

<table>
<thead>
<tr>
<th></th>
<th>Usability Testing</th>
<th>Expert Review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Who conducts?</strong></td>
<td>Representative people from target audience</td>
<td>A couple of Usability Experts</td>
</tr>
<tr>
<td><strong>Useful</strong></td>
<td>When finding real usability problems</td>
<td>When finding violations of usability/design standards</td>
</tr>
<tr>
<td><strong>Length</strong></td>
<td>4-6 weeks</td>
<td>1-3 weeks</td>
</tr>
<tr>
<td><strong>Disadvantage</strong></td>
<td>-Expensive and Time consuming</td>
<td>-False Positives</td>
</tr>
<tr>
<td></td>
<td>-Heavily depends on identifying right target group, accuracy of testing protocol</td>
<td>-Miss the real problems that cause users to fail tasks</td>
</tr>
</tbody>
</table>
Jakob Nielsen: the Magic Number 5

- Usability tests are very costly and complex?
  - Not True
  - “Elaborate usability tests are a waste of resources”

- No more than 5 users and running as many small tests as you can afford
Jakob Nielsen: the Magic Number 5

- “Discount usability” model
- The proportion of Usability
  - $L$: usability problems found by a single user
  - $n$: number of test users

![Graph showing the relationship between the number of test users and the ratio of benefits to costs. The graph indicates that the maximum ratio is achieved with 5 test users.]
Conflicts against the magic number

• Arguing against the five-user guideline in terms of the claim on statistical methods: $1 - (1 - L)^n$
  – A typical L, 31% exists? (varies from 8% to 51%)
  – As the number of participants is increasing, is the proportion of usability problems found increasing?

• Some empirical researches show that testing the first five uncovered much lower percentages (Spool: 35%, Faulkner: 55%) of usability problems than 85% as Nielsen claimed.
Usability Testing: What Have We Overlooked?

• Gitte Lindgaard & Jarinee Chartttratichart
• CUE: Comparative Usability Evaluation, to collect data for usability methods and techniques
• CUE-4: 17 professional teams to evaluate www.hotelpenn.com (9 teams: usability testing, 8 teams: expert review)
Usability Testing: What Have We Overlooked?
Research Questions

• There is a correlation between number of users and the proportion of problems found.
• There is a correlation between number of user tasks and the proportion of problems found.
Methods

• Think-aloud Method

• 9 teams with different number of users

<table>
<thead>
<tr>
<th>Team</th>
<th>A</th>
<th>H</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Users</td>
<td>6</td>
<td>12</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>15</td>
<td>13</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

• Analyze Tasks and Scenarios
  – Task goal, e.g., Find an available room
  – User task, e.g., Check room availability of a particular room type on a certain date; check room availability for the following year, ...
  – User task token, e.g., going back to the home page; making a reservation for a family of three from June 28 to July 5

• Analyze the problems reported by each usability test team
Results

- Number of Users & Usability Problems found
  - No significant correlation
Results

- Number of Users & Usability Problems found
  - Correlation exists

\[ r = 0.731, \ p < 0.05 \ (n=9) \]

\[ r = 0.821, \ p < 0.01 \ (n=9) \]
### Analysis on Results

<table>
<thead>
<tr>
<th>Team</th>
<th>A</th>
<th>H</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Users</td>
<td>6</td>
<td>12</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>15</td>
<td>13</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td># of user tasks</td>
<td>14</td>
<td>11</td>
<td>5</td>
<td>11</td>
<td>12</td>
<td>10</td>
<td>6</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Problems Found(%)</td>
<td>42</td>
<td>43</td>
<td>7</td>
<td>22</td>
<td>27</td>
<td>29</td>
<td>23</td>
<td>24</td>
<td>30</td>
</tr>
<tr>
<td>% New problems</td>
<td>12</td>
<td>8</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

- Correlation between the number of users and the proportion of problems found was not supported
- Correlation between the number of tasks and the proportion of problems found was upheld
- 5 user claim was not supported (85% of the problems found)
- Role of participant recruitment (Team A & Team L)
Summary

• What we have overlooked: Other contributing factors to improve usability such as task coverage and participant recruitment rather than sample size
  – Usability investigation on important target users & the most critical tasks
• If the goal of usability testing is to gather qualitative insights to improve products, one or two individual observations can provide them
• Nielsen’s claim is useless?
  – In overall, the magic number 5 can be applied as one of quantitative assessment
  – To demystify the concept that usability testing is very costly at that time
References

• Gitte Lindgaard, Jarinee Chatratichart, Usability testing: what have we overlooked?, Proceedings of the SIGCHI conference on Human factors in computing systems, April 28-May 03, 2007, San Jose, California, USA


• Faulkner, L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behavior Research Methods, Instruments & Computers, 35, 3, Psychonomic Society (2003), 379—383

• Jared Spool, Will Schroeder, Testing web sites: five users is nowhere near enough, CHI '01 extended abstracts on Human factors in computing systems, March 31-April 05, 2001, Seattle, Washington