Questions?

• Project #3
• HW #7 due Dec 3
• Slides are important too
• Grades
Today

• Introduction to evaluation
Setting the stage for discovery
What does evaluation mean?
Predictions and measurements?

• Type of prediction
  – Point
  – Interval
  – Ordinal
  – Categorical

• Construct and measurement methods
  – Reliability
  – Convergent validity
  – Discriminant validity
Questioning measurements

• Are they reliable?
  – Does the experiment take into account variations between subjects?
    • Need for testing a sample of subjects

• Are they valid?
  – Does the experiment reflects target use?
    • Were users typical?
    • Were tasks typical?
    • Was the setting realistic?
    • Was the experience biased?

• Do they make sense?
  – Setting the stage for discovery!
Are results significant?

• Statistical significance
  – Comparing to the null hypothesis: “There is no effect”
  – Type I errors are the most disruptive

<table>
<thead>
<tr>
<th>Researcher’s Decision</th>
<th>Actual Situation: Null Hypothesis is</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
</tr>
<tr>
<td>Fail to reject the null hypothesis</td>
<td>Correct decision</td>
</tr>
<tr>
<td>Reject the null hypothesis</td>
<td>Type I error</td>
</tr>
</tbody>
</table>

• Design significance?
  – 3.00s versus 3.05s?
The participant standpoint

- Testing is a distressing experience
  - Pressure to perform
  - Feeling of inadequacy
  - Looking like a fool in front of your peers, your boss,…

(from “Paper Prototyping” by Snyder)
Ethics: The Stanford prison experiment

- Was it useful?
  “…that’s the most valuable kind of information that you can have - and that certainly a society needs it” (Zimbardo)

- Was it ethical?
  - Could we have gather this knowledge by other means?