

[cs160.valkyriesavage.com](http://cs160.valkyriesavage.com)



# usability testing

July 14, 2015  
Valkyrie Savage

# Genres of assessment

**Automated**

**Usability measures computed by software**

**Inspection**

**Based on skills, and experience of evaluators**

**Formal**

**Models and formulas to calculate measures**

**Empirical**

**Usability assessed by testing with real users**



# Designing Controlled Experiments

jurvetson on flickr

# Steps in Designing an Experiment

---

1. State a lucid, testable hypothesis
2. Identify variables  
(independent, dependent, control, random)
3. Design the experimental protocol
4. Choose user population
5. Apply for human subjects protocol review
6. Run pilot studies
7. Run the experiment
8. Perform statistical analysis
9. Draw conclusions

# Common Metrics in HCI

---

- ☐ **Performance metrics:**

- ☐ **Task success (binary or multi-level)**
- ☐ **Task completion time**
- ☐ **Errors (slips, mistakes) per task**
- ☐ **Efficiency (cognitive & physical effort)**
- ☐ **Learnability**

- ☐ **Satisfaction metrics:**

- ☐ **Self-report on ease of use, frustration, etc.**

# Satisfaction Metric: Likert Scales

---

**“Overall, I am satisfied with the ease of completing the tasks in this scenario”**

**1: Strongly Disagree**

**2: Disagree**

**3: Neither agree nor disagree**

**4: Agree**

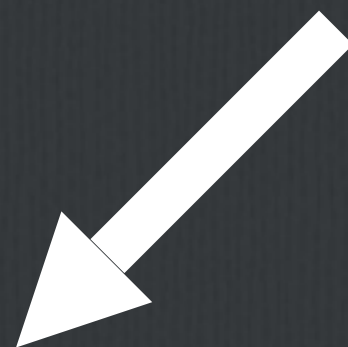
**5: Strongly agree**

☐ Respondents rate their level of agreement to a statement

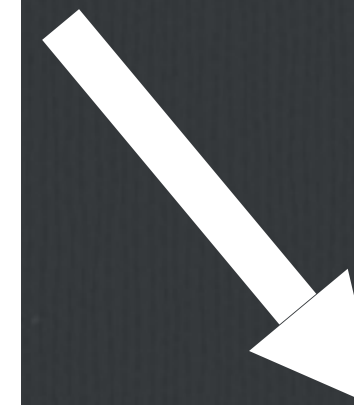
☐ Likert data is ordinal, not continuous (matters for analysis)!

# Between Subjects Design

Wilma and Betty use one interface

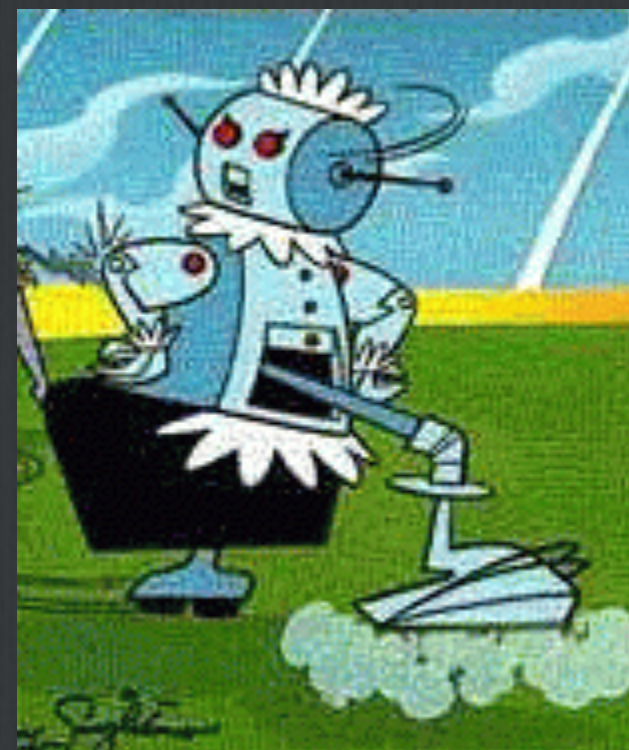


Dino and Fred use the other



# Within Subjects Design

Everyone uses both interfaces



# Between vs. Within Subjects

---

## ☐ Between subjects

### ☐ Each participant uses one condition

☐ +/- Participants cannot compare conditions

☐ + Can collect more data for a given condition

☐ - Need more participants

## ☐ Within subjects

### ☐ All participants try all conditions

☐ + Compare one person across conditions to isolate effects of individual diffs

☐ + Requires fewer participants

☐ - Fatigue effects

☐ - Bias due to ordering/learning effects

# Within Subjects: Ordering Effects

---

- ☐ In within-subjects designs ordering of conditions is a variable that can confound results
  - ☐ Why?
- ☐ Turn it into a random variable
  - ☐ Randomize order of conditions across subjects
  - ☐ Counterbalancing (ensure all orderings are covered)
  - ☐ Latin square (partial counterbalancing)
  - ☐ ...

# Run the Experiment

---

- ☐ Always pilot it first!
  - ☐ Reveals unexpected problems
  - ☐ Can't change experiment design after starting it
- ☐ Always follow same steps – use a checklist
- ☐ Get consent from subjects
- ☐ Debrief subjects afterwards

**ok, now we're going to do it!**

# **Design an Experiment!**

---

- 1. State a lucid, testable hypothesis**
- 2. Identify variables  
(independent, dependent, control,  
random)**
- 3. Design the experimental protocol**
- 4. Choose user population**
- 5. Apply for human subjects protocol  
review**
- 6. Run pilot studies**
- 7. Run the experiment**
- 8. Perform statistical analysis**
- 9. Draw conclusions**

# Design an Experiment!

---

1. State a lucid, testable hypothesis
2. Identify variables  
(independent, dependent, control,  
random)
3. Design the experimental protocol
4. ~~Choose user population~~
5. ~~Apply for human subjects protocol  
review~~
6. Run pilot studies
7. Run the experiment
8. Perform statistical analysis
9. Draw conclusions

**<http://tinyurl.com/littletrans160>**

**<http://tinyurl.com/hellohello160>**

# Design an Experiment!

---

1. State a lucid, testable hypothesis
2. Identify variables  
(independent, dependent, control,  
random)
3. Design the experimental protocol
4. ~~Choose user population~~
5. ~~Apply for human subjects protocol  
review~~
6. Run pilot studies
7. Run the experiment
8. Perform statistical analysis
9. Draw conclusions

# Cognitive Walkthrough

From: Preece, Rogers, Sharp – Interaction Design

---

- ☐ Given an interface prototype or specification, need:
  - ☐ A detailed task with a concrete goal, ideally motivated by a scenario
  - ☐ Action sequences for user to complete the task
- ☐ Ask the following questions for each step:
  - ☐ Will the users know what to do?
  - ☐ Will the user notice that the correct action is available?
  - ☐ Will the user interpret the application feedback correctly?
  - ☐ Record: what would cause problems, and why?

[cs160.valkyriesavage.com](http://cs160.valkyriesavage.com)



# usability testing

July 14, 2015  
Valkyrie Savage