HCI: STUDY DESIGN

Dr Kami Vaniea
First, the news...

No Class on Thursday
We have been learning about discount usability because these are easy to do with a small amount of time and budget.
Today: designing studies and evaluating the results
Step 1: Define what “usable” means
Step 2: Identify your variables
Step 3: Setup your study
Step 4: Evaluate the outcome
The problem:
You just built a new widget and now you need to evaluate it
Step 1: Define what “usable” means
A system which is undefined can never be wrong, it can only ever be surprising.
Define your usability goal

- This step is very similar to specifying tasks for a think-aloud type study
- Identify what you think your users need to be able to do using your system
- The goals need to be specific and easy to identify if they have or have not been completed
- Examples:
  - Find a stool on a shopping page and purchase it
  - Locate the nearest bus stop that the 8 bus stops at
- Bad examples:
  - Have fun using the site
  - Find a bus to go somewhere
Step 2: Identify your variables
What are you going to measure?

- In statistics there are classically two types of measurements (variables): dependent and independent

  - Dependent
    - Also known as the outcome variable
    - Measures the usability goal

  - Independent
    - Anything you are directly manipulating
    - An element of the study which is under your control
    - A pre-existing feature of your participant
Let's use this study as an example.

**Button push required**
- Contacts
  - modify your contacts
  - read your contacts

**Only when app is open**
- Calendar
  - add or modify calendar events and send email to guests without owners' knowledge
  - read calendar events plus confidential information

**Anytime in the background**
- Identity *(Ad software)*
  - find accounts on the device
  - add or remove accounts
  - read your own contact card

- Location
  - approximate location *(network-based)*
Goal:
User can identify if an app can or cannot perform an action directly tied to a permission.

- **Contacts**
  - modify your contacts
  - read your contacts

- **Calendar**
  - add or modify calendar events and send email to guests without owners' knowledge
  - read calendar events plus confidential information

- **Identity** *(Ad software)*
  - find accounts on the device
  - add or remove accounts
  - read your own contact card

- **Location**
  - approximate location (network-based)
**Independent variable:**
Which of the two interfaces the participant was shown

**Dependent variable:**
Count of the number of questions the participant answered correctly
Variables that would make sense

- **Goal**: User can identify if an app can or cannot perform an action directly tied to a permission.
- **Dependent**
  - Number of permissions correctly/incorrectly read
  - Time spent reading the permission screen
- **Independent**
  - Study group
  - Order of the permissions
  - Time of day
  - Type of device (laptop, mobile, PC)
  - Demographics of the participants
XKCD ran a study to see what men and women call different colors

- **Dependent**
  - The color name they typed in

- **Independent**
  - Sex (man or woman)
  - Color they were shown

https://blog.xkcd.com/2010/05/03/color-survey-results/
MSc Project on reading config files

• Goal: Does the order of lines in a configuration file impact the way people interpret the file?

• Dependent
  • True/False – did the participant consider order
  • 1-7 – How confident were they in their answer

• Independent
  • Education level for technical professions only
  • Self-efficacy statements around programming and configuration file modification
  • Prior experience with configuration files
  • Other demographics
What I really want you to learn:

Think about what variables you are interested in and what graph / plot / table you want **before** you conduct the study
Common dependent things to measure

- Time to complete task
- Percent of task completed
- Percent of task completed per unit of time
- Ratio of successes to failures
- Time spent in errors
- Percent or number of errors
- Percent or number of competitors better than it
- Number of commands used
- Frequency of help and documentation use
- Percent of favorable/unfavorable user commands
Common dependent things to measure

- Number of:
  - Repetitions of failed commands
  - Runs of successes and failures
  - Times interface misleads the user
  - Good and bad features recalled by users
  - Available commands not invoked
  - Regressive behaviors
  - Users preferring your system
  - Times users need to work around a problem
  - Times the user is disrupted from a work task
  - Times the user loses control of the system
  - Times user expresses frustration or satisfaction
System Usability Scale

- Have the participants interact with the system
- Have them answer the questions on the right
- Follow the scale instructions
- Use the resulting number as a dependent variable

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Step 3: Setup your study
What do you want to be able to say after the evaluation is done?

- X interface is better than Y interface
  - Run an A/B study
  - Randomly assign users to groups
  - Have all users complete the same tasks
- My new interface is better than my old interface
  - Same as above
  - Or use rapid usability approach
- Users can use interface X to accomplish Y
  - Have users accomplish a set of tasks using X
  - Measure the usability (see step 2)
- Using my interface makes people better/smarter
  - Pre/post test – give them the same test before and after using your system
Between vs. Within subjects

• **Between subjects**
  • Your study only shows one interface to one person
  • You are measuring how well the people randomly assigned to the A interface did compared to the people randomly assigned to the B interface
  • Lots of variability with this method

• **Within subjects**
  • Your study shows all interfaces to all people
  • You are measuring the difference in how they do on the two interfaces
  • Less variability (same person) but more learning effects and priming
Scripted vs observational

- Scripted studies are planned in advance
  - Tasks are prepared in advance
  - Participants are in a controlled environment such as a lab
  - Nearly all lab based studies are scripted
  - Think-aloud is scripted

- Observational studies are not planned and simply observe users doing their own tasks
  - Participants may not even be notified that they are part of a study
  - Participants are in their natural environment doing what they would normally do
  - Hard/impossible to prove what task the user was trying to accomplish
Study design

• A/B test between the existing and new interface
• Between subjects
• 10 Tasks shown in the same order to all participants
• Dependent variables
  • Accuracy on task
• Independent variables
  • Which interface
Study design

- Between subjects
- Multiple tasks

Dependent
- The color name they typed in

Independent
- Sex (man or woman)
- Color they were shown

https://blog.xkcd.com/2010/05/03/color-survey-results/
Step 4: Evaluate the outcome
Evaluation options

- Basic
  - Counts of effectiveness on tasks
- Academically sound
  - Statistics
Basic version

- Count the number of tasks where the participant was able to accomplish your goal
- If most participants were able to accomplish the goal then Yay! The interface is usable.

<table>
<thead>
<tr>
<th></th>
<th>Current Interface</th>
<th>New Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Task 2</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>Task 3</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Task 4</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>
We are about to learn about some of the basic statistics used in HCI

These are only needed if you want to prove that a statement is true
Common statistical tests

- Regression
- T-Test
- ANOVA
- $\chi^2$ (Chi Squared)
Chi Squared

\[ \chi^2 = \sum \frac{(Observed \ Value - Expected \ Value)^2}{(Expected \ Value)} \]

• Answers the question:
  • Does the observed data have the same ratio as expected
  OR
  • Do two counts come from the same distribution
Questions?