Empirical Evaluation

Agenda

- Announcements
- Lecture: Empirical Evaluation
- Evaluation Activity: Usability Testing
- Next class

Announcements, Questions

- A3 due (but hold on to it for now)
- P1 will be returned at the end of class (but lets talk about it a bit now)

P1: Common Mistakes

- Not giving your reader any context or any sense of the outline of the work
- Not being clear about why you chose the three methods you chose and how they work together
- Providing results without any interpretation of what these results mean
- Not connecting your results to your requirements
- Extremely poor grammar
- Sloppy and unprofessional writing and formatting

Empirical Evaluations

- **Usability Evaluations**
  - Typically in lab
  - Tests usability metrics
  - Earlier in design process
- **Field Studies**
  - Out in the "real world"
  - Tests user experience metrics
  - Later in the design process

Usability Evaluations

- Testing Plans
- Usability Lab Studies

- Example: how do we evaluate this site?

  http://historywired.si.edu/index.html

Usability Test Plan

A. Objectives
B. User profile
C. Method
D. Task list
E. Evaluation measures

From Rubin, J. (1994). *Handbook of Usability Testing*. New York: Wiley

A. Test Objectives

- Create *very specific* objectives for your evaluation

- Poor examples
  - Can users identify trends?
  - Is the user-interface usable?

- Good examples
  - Can users employ the slider associated with the timeline to identify outlying dates?
  - Can users select filters and select colors so that the relationship between X & Y is readily seen?
  - Can users find material more quickly in the visual or textual version of the table of contents?

B. User Profile

- Enumerate attributes for your target users and select users that meet the profile
- Can be based on people who fit your persona types

- Example:
  - Age — 25-30
  - Gender — 50% men; 50% women
  - Computer skills — Daily use of IE 6.0
  - Background — Intro class in statistics (STATS-101)
  - Interests — Track stocks online

For your projects …

- What are good test objectives?
- What is a good user profile?

C. Method

- Many different approaches to structuring a test design

- The ‘best’ approach depends on
  - Resources (time & money)
  - Objectives of the study

D. Task List

- A detailed list of tasks
- Each task
  1. Description:
     What you prompt users with?
  2. Machine state:
     Where users begin from?
  3. Successful completion:
     When is the task completed?

Task Unit

1. **Description:**
   - Name three important events that took place in the 1770s in America
2. **Machine state:**
   - Timeline is set to the 1980s
   - Article on space shuttle is being shown
3. **Successful completion:**
   - Participants verbally reports the names of three events by using the timeline

For your project …

What is a good task unit?

1. Description
2. Starting machine state
3. Successful completion

E. Evaluation Measures

- Quantitative count data
  - Time, errors, confusions, breakdowns, workarounds, success/failure.
- Your observations
  - Notes about *where, when, why and how the* above things occurred.
- Users' comments and feedback
  - Often a questionnaire is used at the end
  - User quotes "I LOVE it, except when it crashes"

For your projects …

- What could be good evaluation measures?

Running the Test

- Introduce the test
  - ***"The interface is being tested, not you."***
  - "I didn't design or build this; I was just asked to find out what the problems are."
- Prompt them to continually think-aloud
- Observe task times, errors, confusions, breakdowns, workarounds, and success/ failure
  - Make notes, video-record, audio-record

Allowing Them to Stray

- If you build extra time into your tests, you can allow users to stray a bit as they work
  - They should stay on task
  - But they might wander down a rabbit hole
  - This can yield good data, but takes time
- Eventually, you may have to interrupt and prompt them to find their way back. If they can't, help them, and note a major failure.

Answering a User's Questions

- Basically, you really, really shouldn't
  - You wouldn't be there "in real life"
  - You want to see if they can figure it out
  - You want to see how hard it is
  - You want to see how catastrophic the outcome is if they keep struggling
- Answering users' questions for help ruins your data and contaminates them
  - "Why don't you try something else?"

Being a Good Moderator

- Spend almost all your time listening, observing carefully, and planning what to say (or not say) next
- ‘Encourage’ participants in a neutral fashion
- When people become quiet say
  - “Can you keep talking?”

Think Aloud Prompts

- “Tell me what you are thinking.”
- “Tell me what you are trying to do.”
- “Are you looking for something? What?”
- “What did you expect to happen just now?”
- “What do you mean by that?”

Debrief

- Tell them more details about what you were interested in discovering, with their help
- Answer any questions they have
- Now you can show them how to accomplish tasks that they had failures on
- Thank them for their time
- Pay them $$! :)

Human Subject Ethics

- Being in a user test can be uncomfortable for some
- Guidelines
  - Acknowledge that that system is being tested, not the participant (remind repeatedly)
  - Tell the participant that she is free to leave at any time
  - Reveal who is watching & what is being recorded
  - Do not report results such that a participant is identified
  - Avoid telling the participant that he is making mistakes or doing things wrong
  - Acknowledge participants efforts but in a neutral fashion
- Bottom line: Treat people with great respect

Tips for Usability Evaluations

- Keep it simple
- Keep your objectives specific
- Be consistent with all participants
  - Create a script and follow it carefully
- Conduct a pilot test to uncover problems
- Have detailed plan for analyzing the data

Field Studies

- Give users a functional prototype of your system and let them use naturally for a set amount of time
  - Also called "in situ" studies, "real world deployments," or studies "in the wild"

Considerations

- How long? / The Novelty Effect
- How many people?
- How to recruit?
- How to retain participants?
- Experimental or exploratory?
- What data to collect?

How long? / The Novelty Effect

- Any new technology will get the most use when it first is introduced, then interest wanes
- How long?
  - Nathan Eagle (MIT) estimates it is about 2 weeks
  - May be longer
- Field deployments should last long enough for the novelty effect to wear off to understand more realistic use

Participants

- The more the better, but think about your resources
- Try to recruit as diverse of sample as possible
  - Think about recruiting proportional to your personas
  - If experimental, may want to recruit homogenous sample to reduce variables
- Recruitment – internet ads, word of mouth, "snowball" sampling
  - Consider offering payment to attract and retain

Experimental or Exploratory?

- Comparing new product against an old one can be very powerful
  - Your new product: experimental
  - Existing product: control
  - "participants using product Y preferred it over product X 9 times out of 10"
- Exploratory: just give out your product and see what happens
  - Often better in initial stage evaluation

What data to collect?

- Pre- and post- evaluation interviews and surveys
  - Depending on length of study, consider mid-study interviews as well
- Log usage data (if possible)
  - Automatically, with a computer script
  - Time stamped
- Diary entries after each use
  - Can automatically be prompted after usage

Summary - Empirical Evaluations

- **Usability Evaluations**
  - Typically in lab
  - Tests usability metrics
  - Earlier in design process
- **Field Studies**
  - Out in the "real world"
  - Tests user experience metrics
  - Later in the design process

Class Activity: Usability Test

- Testing your A3 paper prototypes
  - Break into groups of 3 (I REALLY MEAN THREE!!!!!)
    - 1 evaluator, 1 usability tester, 1 evaluator's assistant (e.g., take notes
  - You'll run user tests on each other using the tasks you prepared
  - Remind your participant to think aloud
  - Then, swap roles

Next Class

- HCI in the "real world"
  - LOTS OF READING, Get started NOW if you haven't already.
  - There will be discussion. Prepare.
- Upcoming Work
  - R8 due Thursday (last reading reflection!)
  - P3 demos 12/4