User Testing

CSE 510
Richard Anderson
Ken Fishkin

Today's lecture

- User Testing
  - Overview
  - Examples
  - Techniques
  - Case study

User testing

- Why do you want to test users?

Motivation
Answer scientific question
Promote a product
Develop a product
Evaluate a product
Understand a work process
Market research

Obvious point – goals of testing
Influences methodology

Product development vs. scientific studies

Both valid – don't knock dev

- Products
- Science

Emphasize different goals and methods
Products will be evaluated in the market place
Scientific results must stand on their own
Breadth of scientific results
Aim to be broader than the particular artifact

What is user testing?

Laundry list
Interviews, observations,
Surveys, logs, measurements
Video analysis, verbal
Protocols, experiments,
Artifact examination

Damned if you do, damned if you don't

- You will be critical of almost every user study that you read

And almost everyone will be critical of yours, too
"The food is awful and the portions are too small"

Why are user studies hard?

- Achieving statistical significance
- Confounding factors
  - Often trying to measure a low order effect
- Users are not always easy to deal with
- Experimental design is not easy
- Large resource requirements

User studies for IO devices

- How would you evaluate the NYU Quikwriting

This is a speculative idea
What do you want to figure Out?
One obvious challenge is that It is slow to learn the device.
Can an expert use it effectively
Is it learnable? Why or Why not.

User studies for Tablet PC grading tool (paperless grading)

- TAs annotate CS1 assignments using Tablet PC

Quality of grading
Efficiency of Grading

Design and use of Annotation system

Bring up Hawthorne effect

Ethical Considerations

- Do not harm the participants

Broad definition of Harm
Pain, embarassment
Discomfort

Do not rant about HSB
If HSB discussion
There have been serious problems
HSB often oriented
Towards medicine

Informed consent

- Participant must be given full information
- Ability to opt out (without penalty)
- Free from coercion

Optional – HS form – "Is Deception used in this study"
Levels of deception

Privacy

- HSB very concerned about participant privacy
- Concerns about data linked to individuals
  - Access to records
  - Retention of information

Audio and video recordings, logged data

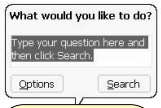
Basic techniques

Emphasize that other disciplines Have done a lot of work on these Its easy to get these wrong

- Surveys
- Unstructured interviews
- Semi-structured interviews

Mr. Wizard Testing

- Testing before building



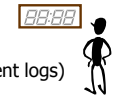
Examples of mockups
Paper based studies
Classic Clipee study

Verbal Protocols

- Need to know what users are thinking, not just what they are doing
- Ask users to talk while performing tasks
  - tell us what they are thinking
  - tell us what they are trying to do
  - tell us questions that arise as they work
  - tell us things they read
- Make a recording or take good notes
  - make sure you can tell what they were doing

Thinking Aloud (cont.)

- Prompt the user to keep talking
  - "tell me what you are thinking"
- Only help on things you have pre-decided
  - keep track of anything you do give help on
- Recording
  - use a digital watch/clock
  - take notes, plus if possible
    - record audio and video (or even event logs)



Design Experiments

Done in real setting

- Qualitative
- Inform design of educational intervention

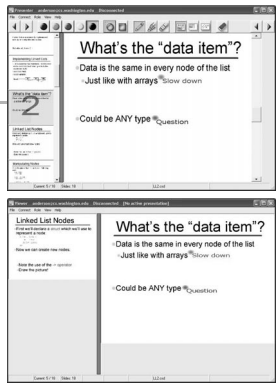
Ethnography

- Immersive study

Limit discussion of the system

Case study

- Classroom Feedback System
- Student devices give real time feedback to lecturer
- Feedback associated with slide content



Main point of discussion – Mechanics of classroom evaluation

Questions

- Does this work?
  - Does this improve the large lecture class?
  - Is this information valuable for instructors
  - Will students give useful feedback?

Questions of interest – but Avoid discussion of them

Why we did not look at "learning outcomes"

Methodology

- Design experiment
  - Gather information from multiple sources
  - Study the application in the real setting
  - Use results to alter the design of the intervention
  - Qualitative, not quantitative

The studies

Major undertaking – even for a Small number of classes

- Pen and paper
- CSE 100, 4 classes observed, 1 with CFS
- CSE 142, 20 classes observed, 6 with CFS
- Classroom experiments, 10 students with laptops for feedback

Methodology

- Surveys
- Instructor interviews
  - With tape recording and transcript
- Detailed classroom observations
  - Instructor and student utterances
- System logging

Data analysis

- Survey tabulation
- Transcription
- Coding of observations
- Analysis of logs
- Correlating events

Results

- Raw data analysis –
  - Basic class had low interaction rate
  - The system received modest usage – but did increase communication rate
- Detailed analysis
  - Identified specific episodes of interaction
    - Cases where instructor used feedback
  - Discovery of usage patterns
    - Feedback lag
    - Anticipatory feedback

Next Week

- February 17. President's day, no class
- February 19. Tangible Interfaces
  - Hiroshi Ishii and Brygg Ullmer, "Tangible Bits: Towrads Seamless Interfaces between People, Bits, and Atoms". CHI '97 Conference Proceedings, March 1997
  - Roy Want, Kenneth Fishkin, Anuj Gujar, Beverly Harrison, "Bridging Physical and Virtual Worlds with Electronic Tags", CHI '99 Conference Proceedings, April 1999.