CSE 440: Introduction to HCI
User Interface Design, Prototyping, and Evaluation

Lecture 17: Closing Thoughts
James Fogarty
Daniel Epstein
Brad Jacobson
King Xia

Tuesday/Thursday
10:30 to 11:50
MOR 234
Today

Informal Prototyping Fun
Experimental Design and Statistics Background
Usability Evaluation Considered Harmful
Presentation Feedback
Exam Q&A
Video Critiques
Informal Prototyping

Sketches are informal

- allowing rapid iteration and greater exploration

Paper prototypes extend that to testing

- person simulates the computing

These are core ideas, central to practice

- so we taught them and you did them
Informal Prototyping

Two related ideas, each can be extended
  keep representation lightweight and fast
  Wizard of Oz simulation for prototype functionality

Heavily explored in research
  think when you might benefit, go looking for ideas
Classic Examples: DENIM
Classic Examples: DENIM

Lin et al, CHI 2000
Classic Examples: Topiary

Li et al., UIST 2004
Classic Examples: Topiary

Li et al., UIST 2004
Classic Examples: SUEDE

Klemmer et al, UIST 2000
Classic Examples: SUEDE
Classic Examples: Phidgets
Classic Examples: Phidgets

Greenberg and Fitchett, UIST 2001
Rapid Fabrication: Constructables
Rapid Fabrication: Constructables
Rapid Fabrication: WirePrint

3D printing: 1:55h
WirePrint: 37min
Rapid Fabrication: WirePrint

3D printing: 1:55h
WirePrint: 37min
Today

Informal Prototyping Fun
Experimental Design and Statistics Background
Usability Evaluation Considered Harmful
Presentation Feedback
Exam Q&A
Video Critiques
Remember our Fitts’s Law study?

This course has focused on quickly getting information to allow you to improve design

Rich methods for experimental design and statistics needed when measurement is goal

Know some of this exists, learn it if you need it

These slides are a bag of keywords

One starting point

https://depts.washington.edu/aimgroup/proj/ps4hci/
Experimental Design and Statistics

Even seemingly simple experiment can be difficult or impossible to correctly analyze.

Design and analysis are inseparable.

Consider your experiment and analyses together so you do not run an experiment you cannot analyze.

Design finds a difference, statistics test it.
Causality and Correlation

We cannot prove causality

We can only show strong evidence for it
Always something outside the scope of an experiment that could be the true cause

We can show correlation

Treatment changes, so does outcome
Hold all things equal but for one
Eliminate possible rival explanations

A negative result means little or nothing
Internal and External Validity

Internal Validity

Convincingly link treatments to effects and the experiment has high internal validity, it shows an effect

External Validity

An experiment likely to generalize beyond the things directly tested is said to have high external validity

Often at odds with each other
Achieving Control

Avoiding other plausible explanations

  Often referred to as confounds

General Strategies

  Remove and/or exclude
  Measure and adjust
  Spread effect equally over all groups
    Randomization (assign randomly)
    Blocking (assign balanced)
Variable Terminology

Factors: Variables of interest
one variable is a single-factor experiment

Levels: Variation within a factor
not necessarily binary

Independent Variables
variables you control

Dependent Variables
outcome measures
(they depend on your independent variables)
Factorial Designs

May have more than one factor
Factors may have multiple levels

A 2x2x3 study has
two factors of two levels each
and a third factor with three levels

Text entry method {Multitap, T9} x
Number of hands {one, two} x
Posture {seating, standing, walking}

Potential dependent variables?
Within and Between Subjects

Within-Subjects Designs
- Each participant experiences multiple levels
- Much more statistically powerful
- Much harder to avoid confounds

Between-Subjects Designs
- Each participant experiences only one level
- Requires more participants
- Avoids possible confounds, easier to analyze
Carryover Effects

Learning, fatigue

anything that transfer between within-subject tasks

Counterbalanced designs help mitigate
e.g., Latin square
\( p \) values

The statistical significance of a result is generally summarized as a \( p \) value (\( N \) is not enough)

\( p \) is the probability the null hypothesis is true (there is no difference)

The same experiment, run \( 1/p \) times, would generate this result by random chance

\( p < .05 \) is an arbitrary but widely used threshold of statistical significance
p and Normal Distributions

Given a mean and a variance, assuming a Gaussian distribution allows estimating the likelihood of a value.

Thus, parametric tests (most common tests) assume data is from normal distributions.
Some Tests

t test
  single factor, possibly multiple levels

F test
  multiple factors
  linear regressions fits equation to variables
  main effects (impact of single factor)
  interactions (relationship between factors)

Chi Square test
  comparing proportions

Non-Parametric tests
  data from non-normal distributions
Concern for Fishing

Bad form to simply test things until you find something significant, then to report that

Comparisons should be theoretically motivated

Recall the definition of $p$

Unprincipled comparisons increase risk of falsely identifying a result

Because if you test enough things, something is bound to be significant

See Tukey’s Honestly Significant Difference

See Sequential Bonferroni Procedure
Today

Informal Prototyping Fun
Experimental Design and Statistics Background
Usability Evaluation Considered Harmful
Presentation Feedback
Exam Q&A
Video Critiques
Usability Evaluation Considered Harmful

http://dx.doi.org/10.1145/1357054.1357074
Usability Evaluation Considered Harmful
Some of the time

Saul Greenberg
University of Calgary

Bill Buxton
Microsoft Research
Warning: Opinions Ahead
Warning: Opinions Ahead
Warning: Opinions Ahead
An anti usability rant?

Bill


Saul

Usability evaluation *if wrongfully applied*

In early design
- stifle innovation by quashing (valuable) ideas
- promote (poor) ideas for the wrong reason

In science
- lead to weak science

In cultural appropriation
- ignore how a design would be used in everyday practice
the choice of evaluation methodology - if any – must arise and be appropriate for the actual problem, research question or product under consideration
Changing how you think

- Usability evaluation
- CHI trends
- Theory
- Early design
- Science
- Cultural appropriation
Part 1. Usability Evaluation
Usability Evaluation

assess our designs and

test our systems to ensure that they actually

behave as we expect and

meet the requirements of the use

Dix, Finlay, Abowd, and Beale 1993
Usability Evaluation Methods

Most common (research):
- controlled user studies
- laboratory-based user observations

Less common
- inspection
- contextual interviews
- field studies / ethnographic
- data mining
- analytic/theory
- ...
Part 2. CHI Trends
CHI Trends (Barkhuus/Rode, Alt.CH1 2007)
CHI Trends (Barkhuus/Rode, Alt.CHICHI 2007)

usability evaluation in industry
User evaluation is now a pre-requisite for CHI acceptance.
CHI Trends  (Call for papers 2008)

Authors
“you will probably want to demonstrate ‘evaluation’ validity, by subjecting your design to tests that demonstrate its effectiveness”

Reviewers
“reviewers often cite problems with validity, rather than with the contribution per se, as the reason to reject a paper”
HCI Education
HCI Practice

Dogma

Usability evaluation = validation = CHI = HCI
Part 3. Some Theory
Discovery
- uncover facts
- detailed evaluation

Invention
- create new things
- refine invention

Understand what is

Influence what will be
Time

Breakthrough | Replication | Empiricism | Theory | Automation | Maturity

Learning

Brian Gaines
Time

Learning

Breakthrough | Replication | Empiricism | Theory | Automation | Maturity

early design & invention | science | cultural appropriation
Part 4. Early Design

Breakthrough Replication
Memex

Bush

Concept
Unimplemented and untested design. Microfilm is impractical. The work is premature and untested.

Resubmit after you build and evaluate this design.
We usually get it wrong
Early design as working sketches

Sketches are innovations valuable to HCI

- Evocative
- Suggest
- Explore
- Question
- Propose
- Provoke
- Tentative
- Noncommittal
Early design

Early usability evaluation can kill a promising idea
- focus on negative ‘usability problems’
Early designs

Iterative testing can promote a mediocre idea
Early design

Generate and vary ideas, then reduce

Usability evaluation
the better ideas
Early designs as working sketches

Getting the design right

Getting the right design
Early designs as working sketches

Methods:
- idea generation, variation, argumentation, design critique, reflection, requirements analysis, personas, scenarios contrast, prediction, refinement, ...
Part 6. Science
I need to do an evaluation
What's the problem?
It won't get accepted if I don't. Duh!
Research process

Choose the method *then* define a problem

*or*

Define a problem *then* choose usability evaluation

*or*

Define a problem *then* choose a method to solve it
Research process

Typical usability tests
- show technique is better than existing ones

Existence proof: one example of success
Research process

Risky hypothesis testing
- try to disprove hypothesis
- the more you can’t, the more likely it holds

What to do:
- test limitations / boundary conditions
- incorporate ecology of use
- replication
Part 6. Cultural Appropriation

Automation  Maturity
Hypertext

From Wikipedia, the free encyclopedia

"MetaText" redirects here. For the literary concept, see Metafiction.

Hypertext most often refers to text on a computer that will lead the user to other, related information on demand. Hypertext represents a relatively recent innovation to user interfaces, which overcomes some of the limitations of written text. Rather than remaining static like traditional text, hypertext makes possible a dynamic organization of information through links and connections (called hyperlinks). Hypertext can be designed to perform various tasks; for instance when a user "clicks" on it or "hovers" over it, a bubble with a word definition may appear, a web page on a related subject may load, a video clip may run, or an application may open.

Contents
1 Etymology
2 Types and uses of hypertext
3 History
   3.1 Early precursors to hypertext
   3.2 The Memex
   3.3 The invention of hypertext
   3.4 Applications
   3.5 Hypertext and the World Wide Web
4 Implementations
5 Academic conferences
6 HyperText fiction
   6.1 Critics and theorists
7 See also
8 References
9 External links

Etymology

The prefix hyper- ("over" or "beyond") signifies the overcoming of the old linear constraints of written text. The term "hypertext" is often used where the term hypermedia might seem appropriate. In 1992 Ted Nelson - who coined both terms in 1965 - wrote:
Dear Customers,

A strange thing happened on the way to the paperless society. We humans created more paper than ever before. Computer printers (and their evil companion, the ink-toner cartridge) have proliferated, and most of us routinely print out and lug around loads of personal and professional documents. Why? It’s not that buying printers or changing ink-toner cartridges is fun. It’s because reading on paper is better than reading on traditional computer displays. Printing has been worth the hassle.

Kindle starts to change that. People who see Kindle’s display for the first time do a double-take. It looks and reads like real paper. People who swore they would never read books on computers are reading books on Kindle in numbers far greater than we ever expected. And they’re now starting to ask: If I can carry my whole library around on my Kindle, how about I carry all my personal and professional documents there too?

We’re excited to announce Kindle DX, the large-screen addition to the Kindle family.
To create a new entry in the blog, just type here...

I've nearly finished my FAQ on blogs. Many of the comments and suggestions. It'll be completed on Tuesday, Aug. 5th.

A friend at Cisco told me that two weeks ago, he heard about a 1,500 person layoff. All the jobs are being offered for $500,000, but the only way to get this out of the newspapers, Cisco is laying off people for $200 per week.

Kathryn Ortland

Name: Kathryn Ortland
Member Since: August 2, 2005
Last Update: November 10, 2005

Account Info

Basic Info

Account Info

Contact Info

Personal Info

Professional Info

We made
Announcements
better
Check out the new Facebook Flyers.
Starting at only $5!
Advertise to your campus now.

View More Photos of Me (18)
Edit My Profile
Edit My Privacy

This is you.

Friends at Oregon
Kathryn has 22 Oregon friends.
Part 7. What to do
No Evaluation
More Appropriate Evaluation
The choice of evaluation methodology - if any – must arise and be appropriate for the actual problem or research question under consideration.

argumentation
design critiques
design competitions
visions
inventions
prediction
reflection
design rationales

...
We decide what is good research and practice
There is no them
Only us
Remember Both Sides of this Course

This course emphasize both

Getting the Right Design
Getting the Design Right

Many people fall into a trap of the latter

Be mindful of your methods
Today

Informal Prototyping Fun
Experimental Design and Statistics Background
Usability Evaluation Considered Harmful
Presentation Feedback
Exam Q&A
Video Critiques
Initial Paper Prototype
Initial Paper Prototype
Task 1: Finding a SmartMatch

Criteria
Level √
Avg Dist ✘
Avg Time ☐
Route Pref ☐

Criteria
Level √
Avg Dist ☐
Avg Time ☐
Route Pref ☐

Frank
About: Happy guy who runs casually
Level: Novice
Rating: ★★★
Add + Find Another

Jenn
No more matches found!!
Rating: ★★★
Add + Find Another
Testing - Results

- **Heuristic Evaluation**
  - High Severity Issues
  - **Example:** “Go Shopping” mode was useless

- **User Testing**
  - High and Medium Severity Issues
  - **Example:** Takes too long to get to “Add Item”

- **Design Mockup Critique**
  - Low Severity and Aesthetic Issues
  - **Example:** Home screen too cluttered
GOOD MORNING, JASON.

8:10 AM  67°
11.5.2014  96 BPM

MOOD CHART
ADD ACTIVITY
SEND DATA

AROUSAL:
CALM ➔ EXCITED

BORED

Activity added: 1粥

Activity: Tennis
Starting date: 11.5.2014
Starting time: 10:00 AM
Duration: 1 hr

Activity:
Starting date: 11.5.2014
Starting time: 7:00 AM
Duration: 1 hr

ADD DAY

Calendar:

1. Tennis Practice
2. Ice Cream Social

Day: March 12, 2014
Final Paper Prototype

IEP-Connect Classroom

Many Screens on One Slide
Final Paper Prototype

Task 2 - Record Water Intake

Many Screens on One Slide
Fewer Screens, Show Connections

Band-it

- Begin Workout
- Workout History
- Medical Profile
- Health Tips

Last Workout: Nov 16 2014 Left Ankle
- 12:20pm Hamstring Stretch
- 11:55am Tennis
- 11:52am Squats
- 11:35am Jumping Jacks

Medical Profile
- Age: 25
- Weight: 150 lbs
- Height: 6 ft 1 in
- Sex: Male
- Previous Injuries and Conditions:
  - Sprained Ankle: 60%
  - Dislocated Shoulder: 40%

Family History:
- Osteoarthritis
- Flat Feet
Fewer Screens, Show Connections
IMPROVED DESIGN

Sunday, November 2nd, 2014

Overall

Sessions

Distracted  Productive

Session 1  Session 2

2 h 10 m 4 h 20 m
Facebook Conversations

1 h 50 m
MatLab

Click on activity for more actions.
IMPROVED DESIGN

Sunday, November 2nd, 2014

Click on activity for more actions.
IMPROVED DESIGN

Sunday, November 2\textsuperscript{nd}, 2014

Session 1 (9 am - 1:20 pm) - 4 h 20 m
- 2 h 10 m Facebook
- 20 m Conversations
- 1 h 50 m MatLab

Click on activity for more actions.
IMPROVED DESIGN

Sunday, November 2nd, 2014

Session 1 (9 am - 1:20 pm) - 4 h 20 m
2 h 10 m facebook
20 m conversations
1 h 50 m MatLab

Click on activity for more actions.
IMPROVED DESIGN

Sunday, November 2nd, 2014

Session 1 (9 am - 1:20 pm) - 4 h 20 m

Session 1 - 60% productive
Session 2 - 40%

2 h 10 m facebook
20 m conversations
1 h 50 m MatLab

Click on activity for more actions.
IMPROVED DESIGN

Sunday, November 2nd, 2014

Overall | Sessions
---|---

<table>
<thead>
<tr>
<th>Distracted</th>
<th>Productive</th>
</tr>
</thead>
</table>

Session 1 (9am - 1:20pm) = 4h 20m

2h 10m | Facebook
---|---
20m | Conversations
1h 50m | MatLab

Click on activity for more actions.
Initial Paper Prototype

Task 1: Is Netflix worth it?

1. View the Koala homepage
2. Navigate to Netflix Detailed View
3. View your Usage Score for Netflix
4. Go to Settings
5. Click “Unsubscribe”
6. Return to homepage
Initial Paper Prototype

Task 1: Is Netflix worth it?

1. View the Koala homepage
2. **Navigate to Netflix Detailed View**
3. View your Usage Score for Netflix
4. Go to Settings
5. Click “Unsubscribe”
6. Return to homepage
Initial Paper Prototype

**Task 1: Is Netflix worth it?**

1. View the Koala homepage
2. Navigate to Netflix Detailed View
3. **View your Usage Score for Netflix**
4. Go to Settings
5. Click “Unsubscribe”
6. Return to homepage
Initial Paper Prototype

*Task 1: Is Netflix worth it?*

1. View the Koala homepage.
2. Navigate to Netflix Detailed View
3. View your Usage Score for Netflix
4. **Go to Settings**
5. Click “Unsubscribe”
6. Return to homepage
Initial Paper Prototype

Task 1: Is Netflix worth it?

1. View the Koala homepage.
2. Navigate to Netflix Detailed View
3. View your Usage Score for Netflix
4. Go to Settings
5. **Click “Unsubscribe”**
6. Return to homepage
Initial Paper Prototype

Task 1: Is Netflix worth it?

1. View the Koala homepage.
2. Navigate to Netflix Detailed View
3. View your Usage Score for Netflix
4. Go to Settings
5. Click “Unsubscribe”
6. Return to homepage
PROBLEM

- LUNCH $5.49
- COFFEE -$3.49
- MOVIES -$11.20
- COFFEE $3.89
- ICE CREAM $4.42
- DINNER -$7.79
- COFFEE $4.09
- BOWLING -$10.20
- KIT KAT $0.99
- BRUNCH $11.42
- BEER -$4.00
Today

Informal Prototyping Fun
Experimental Design and Statistics Background
Usability Evaluation Considered Harmful
Presentation Feedback
Exam Q&A
Video Critiques
CSE 440: Introduction to HCI
User Interface Design, Prototyping, and Evaluation

Lecture 15: Interface Implementation
James Fogarty
Daniel Epstein
Brad Jacobson
King Xia

Tuesday/Thursday
10:30 to 11:50
MOR 234