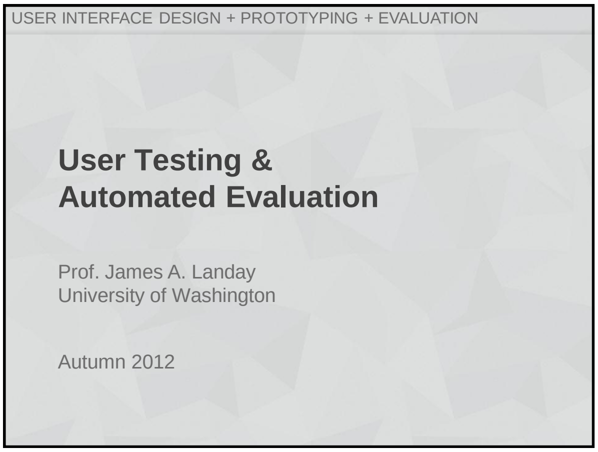
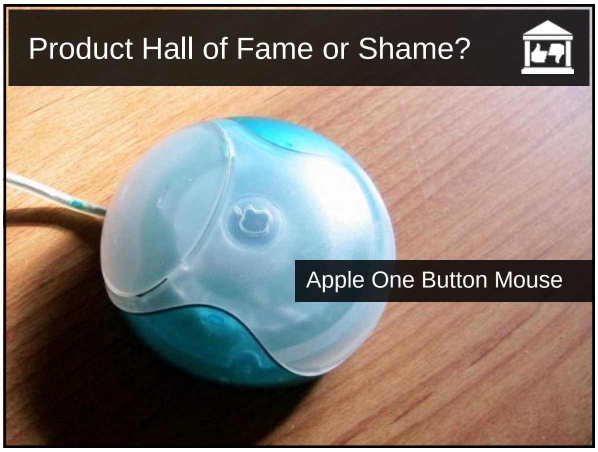


USER INTERFACE DESIGN + PROTOTYPING + EVALUATION

User Testing & Automated Evaluation

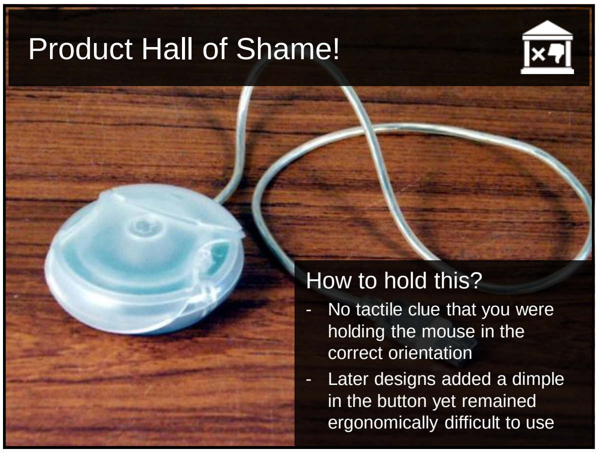
Prof. James A. Landay
University of Washington

Autumn 2012



Product Hall of Fame or Shame?

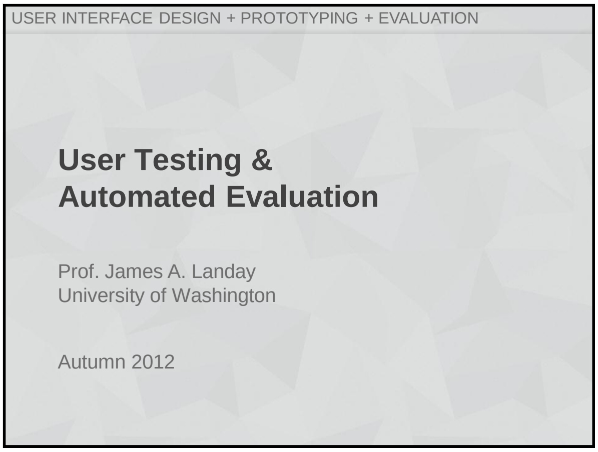
Apple One Button Mouse



Product Hall of Shame!

How to hold this?

- No tactile clue that you were holding the mouse in the correct orientation
- Later designs added a dimple in the button yet remained ergonomically difficult to use

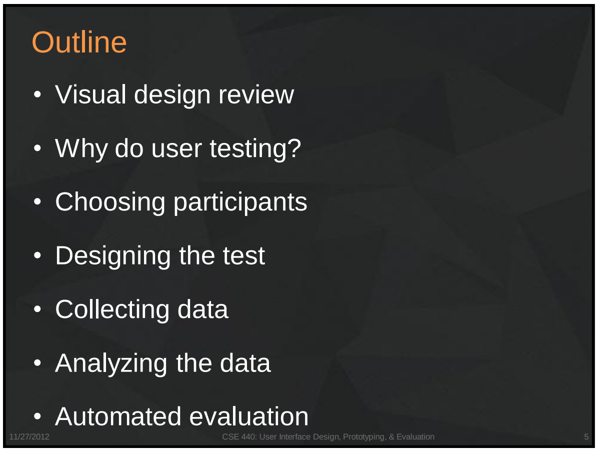


USER INTERFACE DESIGN + PROTOTYPING + EVALUATION

User Testing & Automated Evaluation

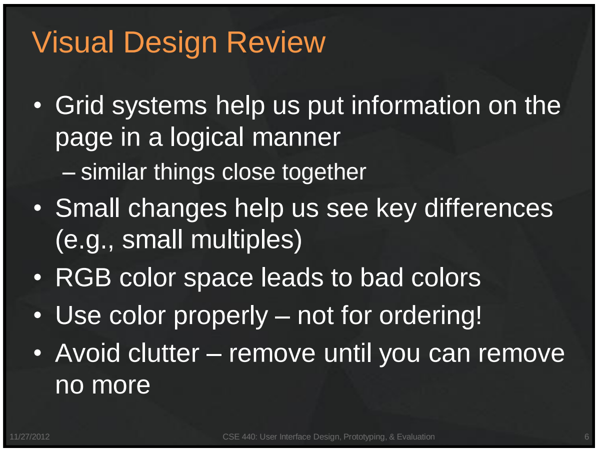
Prof. James A. Landay
University of Washington

Autumn 2012



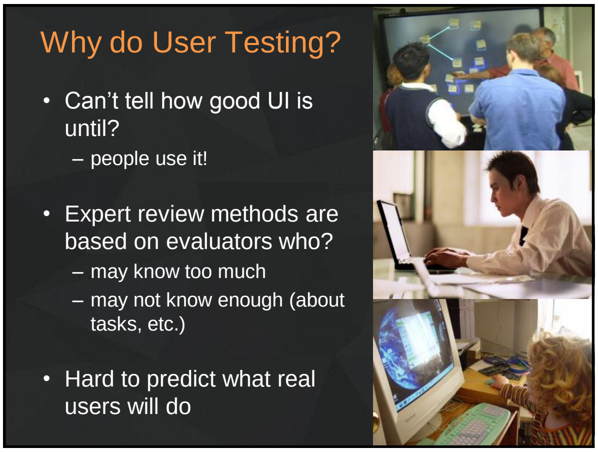
Outline

- Visual design review
- Why do user testing?
- Choosing participants
- Designing the test
- Collecting data
- Analyzing the data
- Automated evaluation



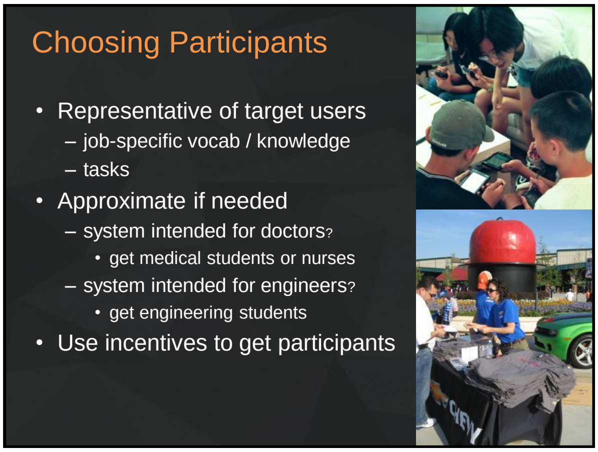
Visual Design Review

- Grid systems help us put information on the page in a logical manner
  - similar things close together
- Small changes help us see key differences (e.g., small multiples)
- RGB color space leads to bad colors
- Use color properly – not for ordering!
- Avoid clutter – remove until you can remove no more



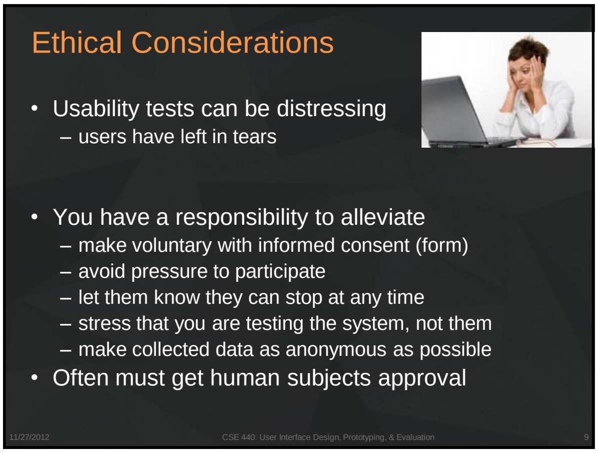
Why do User Testing?

- Can't tell how good UI is until?
  - people use it!
- Expert review methods are based on evaluators who?
  - may know too much
  - may not know enough (about tasks, etc.)
- Hard to predict what real users will do



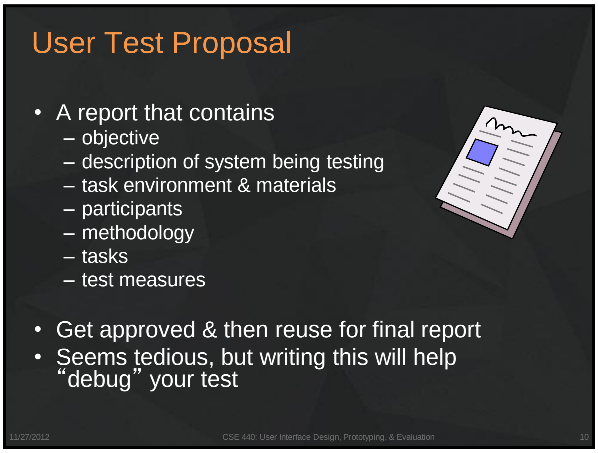
Choosing Participants

- Representative of target users
  - job-specific vocab / knowledge
  - tasks
- Approximate if needed
  - system intended for doctors?
    - get medical students or nurses
  - system intended for engineers?
    - get engineering students
- Use incentives to get participants



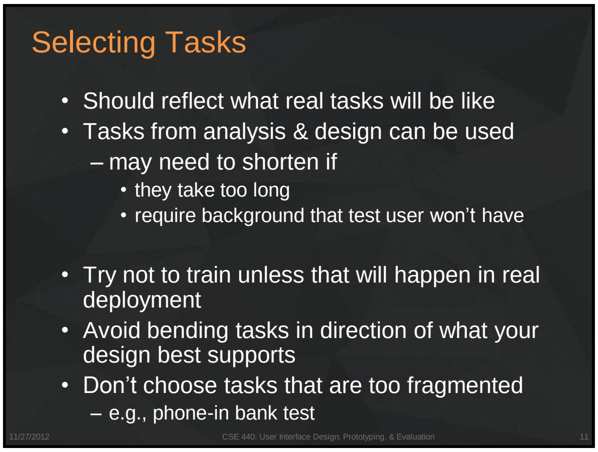
Ethical Considerations

- Usability tests can be distressing
  - users have left in tears
- You have a responsibility to alleviate
  - make voluntary with informed consent (form)
  - avoid pressure to participate
  - let them know they can stop at any time
  - stress that you are testing the system, not them
  - make collected data as anonymous as possible
- Often must get human subjects approval



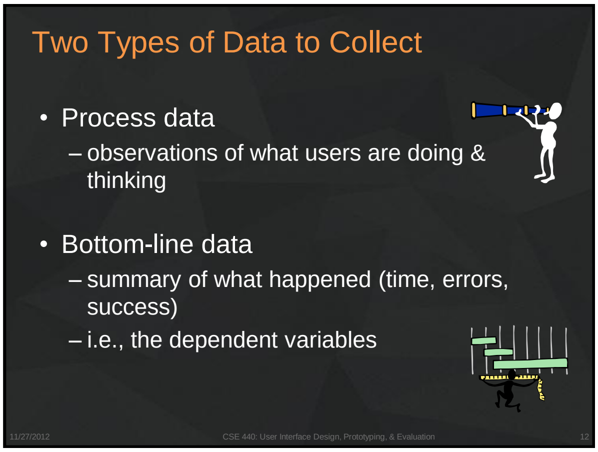
User Test Proposal

- A report that contains
  - objective
  - description of system being testing
  - task environment & materials
  - participants
  - methodology
  - tasks
  - test measures
- Get approved & then reuse for final report
- Seems tedious, but writing this will help "debug" your test



Selecting Tasks

- Should reflect what real tasks will be like
- Tasks from analysis & design can be used
  - may need to shorten if
    - they take too long
    - require background that test user won't have
- Try not to train unless that will happen in real deployment
- Avoid bending tasks in direction of what your design best supports
- Don't choose tasks that are too fragmented
  - e.g., phone-in bank test

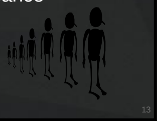


Two Types of Data to Collect

- Process data
  - observations of what users are doing & thinking
- Bottom-line data
  - summary of what happened (time, errors, success)
  - i.e., the dependent variables

Which Type of Data to Collect?

- Focus on process data first
  - gives good overview of where problems are
- Bottom-line doesn’t tell you ?
  - where to fix
  - just says: “too slow”, “too many errors”, etc.
- Hard to get reliable bottom-line results
  - need many users for statistical significance

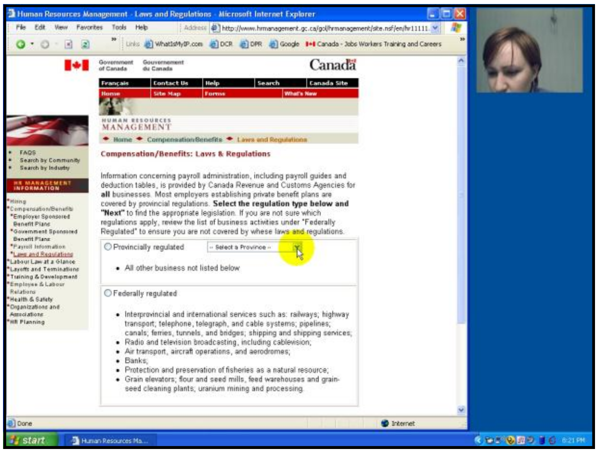


The “Thinking Aloud” Method

- 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 & video (or even event logs)



Human Resources Management - Laws and Regulations - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.hrmanagement.gc.ca/gol/hrmanagement/site.nsf/en/hr11111
Links WhatIsMyIP.com DCR DPR Google Canada - Jobs Workers Training and Careers
Government of Canada Gouvernement du Canada
Canada
Français Contact Us Help Search Canada Site
Home Site Map Forms What's New
HUMAN RESOURCES MANAGEMENT
Home Compensation/Benefits Laws and Regulations
FAQS
Search by Community
Search by Industry
HR MANAGEMENT INFORMATION
*Hiring
*Compensation/Benefits
*Employer Sponsored Benefit Plans
*Government Sponsored Benefit Plans
*Payroll Information
*Laws and Regulations
*Labour Law at a Glance
*Layoffs and Terminations
*Training & Development
*Employee & Labour Relations
*Health & Safety
*Organizations and Associations
*HR Planning
Done
Internet
start
Human Resources Ma...

Compensation/Benefits: Laws & Regulations

Information concerning payroll administration, including payroll guides and deduction tables, is provided by Canada Revenue and Customs Agencies for **all** businesses. Most employers establishing private benefit plans are covered by provincial regulations. **Select the regulation type below and "Next"** to find the appropriate legislation. If you are not sure which regulations apply, review the list of business activities under "Federally Regulated" to ensure you are not covered by whese laws and regulations.

○ Provincially regulated -- Select a Province --

- All other business not listed below

○ Federally regulated

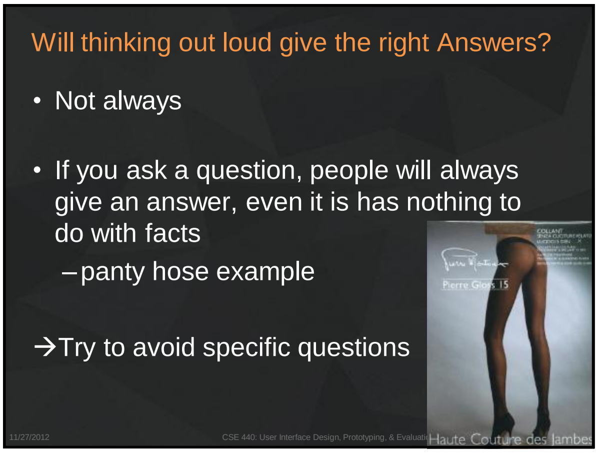
- Interprovincial and international services such as: railways; highway transport; telephone, telegraph, and cable systems; pipelines; canals; ferries, tunnels, and bridges; shipping and shipping services;
- Radio and television broadcasting, including cablevision;
- Air transport, aircraft operations, and aerodromes;
- Banks;
- Protection and preservation of fisheries as a natural resource;
- Grain elevators; flour and seed mills, feed warehouses and grain-seed cleaning plants; uranium mining and processing.

Using the Test Results

- Summarize the data
  - make a list of all critical incidents (CI)
    - positive & negative
  - include references back to original data
  - try to judge why each difficulty occurred
- What does data tell you?
  - UI work the way you thought it would?
    - users take approaches you expected?
  - something missing?

Using the Results (cont.)

- Update task analysis & rethink design
  - rate severity & ease of fixing CIs
  - fix both severe problems & make the easy fixes

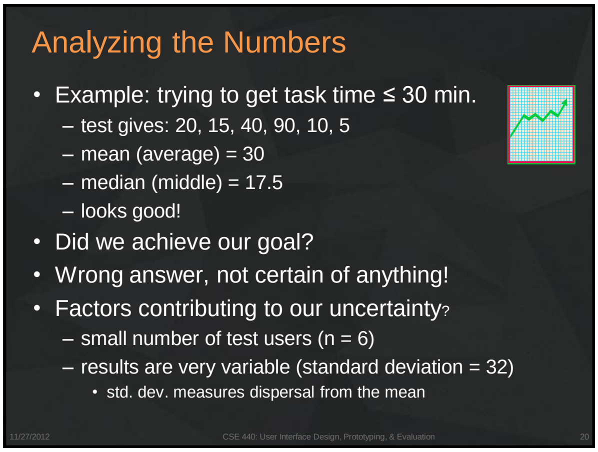


Will thinking out loud give the right Answers?

- Not always
- If you ask a question, people will always give an answer, even it is has nothing to do with facts
  - panty hose example

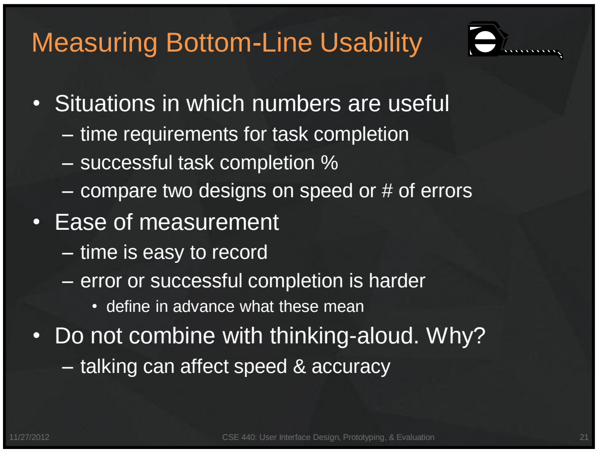
→Try to avoid specific questions

Pierre Gloss 15
Haute Couture des Jambes



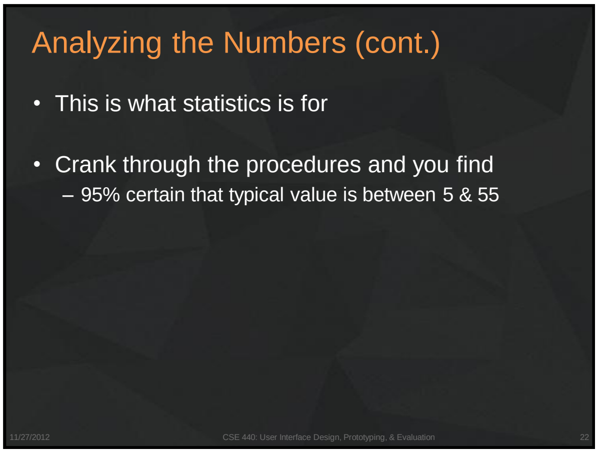
Analyzing the Numbers

- Example: trying to get task time ≤ 30 min.
  - test gives: 20, 15, 40, 90, 10, 5
  - mean (average) = 30
  - median (middle) = 17.5
  - looks good!
- Did we achieve our goal?
- Wrong answer, not certain of anything!
- Factors contributing to our uncertainty?
  - small number of test users (n = 6)
  - results are very variable (standard deviation = 32)
    - std. dev. measures dispersal from the mean



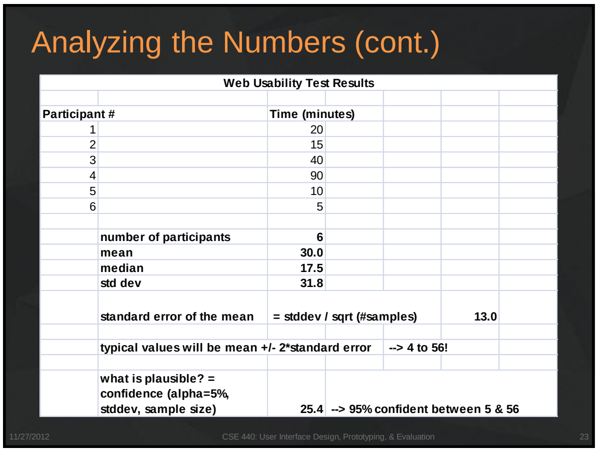
Measuring Bottom-Line Usability

- Situations in which numbers are useful
  - time requirements for task completion
  - successful task completion %
  - compare two designs on speed or # of errors
- Ease of measurement
  - time is easy to record
  - error or successful completion is harder
    - define in advance what these mean
- Do not combine with thinking-aloud. Why?
  - talking can affect speed & accuracy



Analyzing the Numbers (cont.)

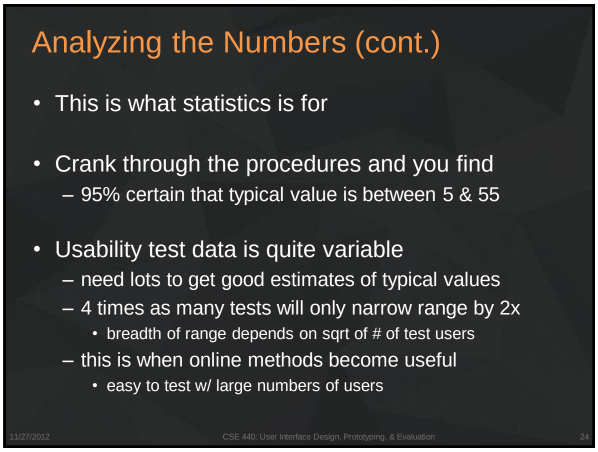
- This is what statistics is for
- Crank through the procedures and you find
  - 95% certain that typical value is between 5 & 55



Analyzing the Numbers (cont.)

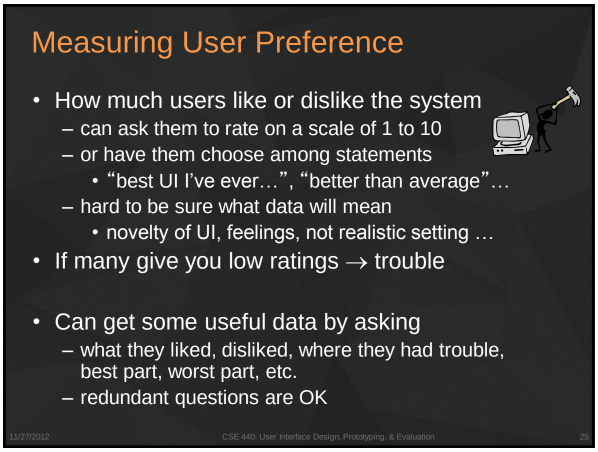
Web Usability Test Results

Participant #		Time (minutes)				
1		20				
2		15				
3		40				
4		90				
5		10				
6		5				
	number of participants	6				
	mean	30.0				
	median	17.5				
	std dev	31.8				
	standard error of the mean	= stddev / sqrt (#samples)			13.0	
	typical values will be mean +/- 2*standard error			--> 4 to 56!		
	what is plausible? = confidence (alpha=5%, stddev, sample size)	25.4	--> 95% confident between 5 & 56			



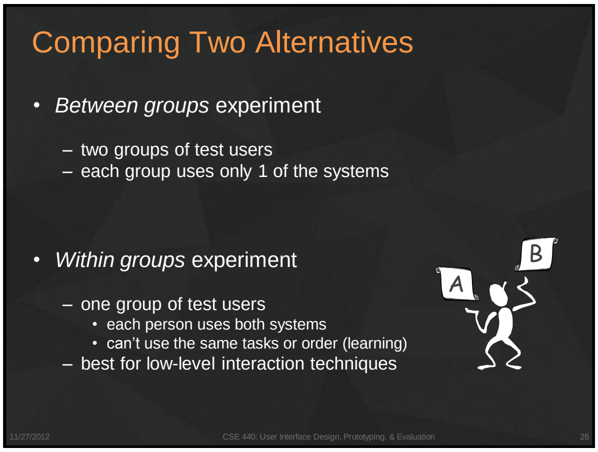
Analyzing the Numbers (cont.)

- This is what statistics is for
- Crank through the procedures and you find
  - 95% certain that typical value is between 5 & 55
- Usability test data is quite variable
  - need lots to get good estimates of typical values
  - 4 times as many tests will only narrow range by 2x
    - breadth of range depends on sqrt of # of test users
  - this is when online methods become useful
    - easy to test w/ large numbers of users



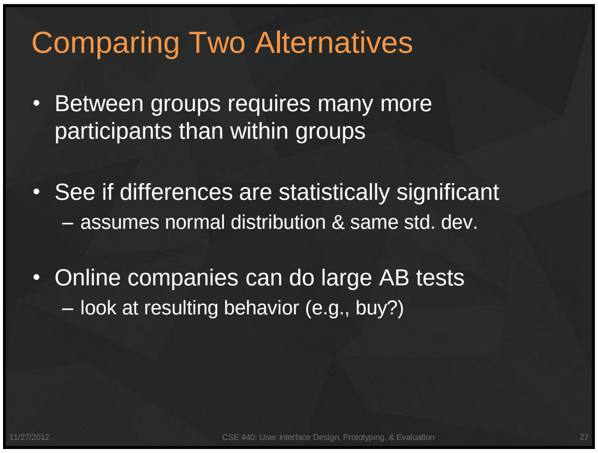
Measuring User Preference

- How much users like or dislike the system
  - can ask them to rate on a scale of 1 to 10
  - or have them choose among statements
    - “best UI I’ve ever…”, “better than average”…
  - hard to be sure what data will mean
    - novelty of UI, feelings, not realistic setting …
- If many give you low ratings → trouble
- Can get some useful data by asking
  - what they liked, disliked, where they had trouble, best part, worst part, etc.
  - redundant questions are OK



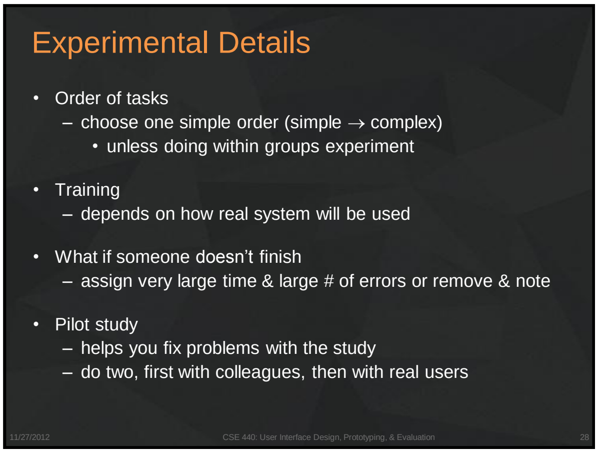
Comparing Two Alternatives

- *Between groups* experiment
  - two groups of test users
  - each group uses only 1 of the systems
- *Within groups* experiment
  - one group of test users
    - each person uses both systems
    - can’t use the same tasks or order (learning)
  - best for low-level interaction techniques



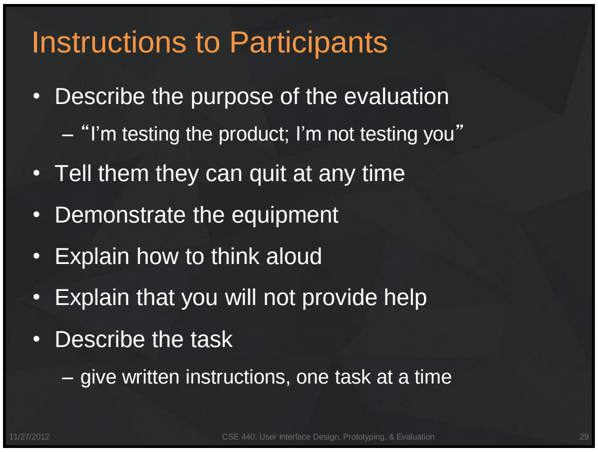
Comparing Two Alternatives

- Between groups requires many more participants than within groups
- See if differences are statistically significant
  - assumes normal distribution & same std. dev.
- Online companies can do large AB tests
  - look at resulting behavior (e.g., buy?)



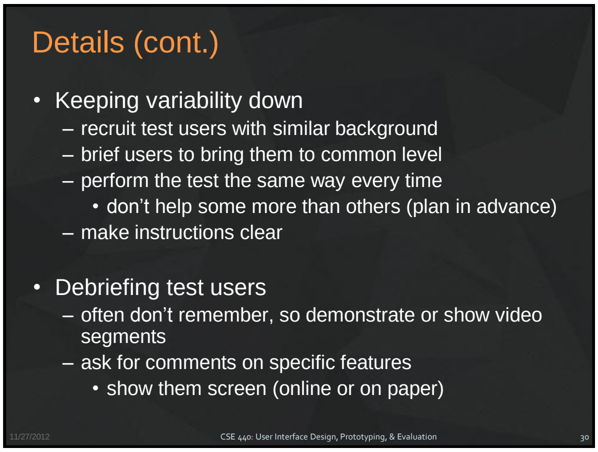
Experimental Details

- Order of tasks
  - choose one simple order (simple → complex)
    - unless doing within groups experiment
- Training
  - depends on how real system will be used
- What if someone doesn’t finish
  - assign very large time & large # of errors or remove & note
- Pilot study
  - helps you fix problems with the study
  - do two, first with colleagues, then with real users



Instructions to Participants

- Describe the purpose of the evaluation
  - “I’m testing the product; I’m not testing you”
- Tell them they can quit at any time
- Demonstrate the equipment
- Explain how to think aloud
- Explain that you will not provide help
- Describe the task
  - give written instructions, one task at a time



Details (cont.)

- Keeping variability down
  - recruit test users with similar background
  - brief users to bring them to common level
  - perform the test the same way every time
    - don’t help some more than others (plan in advance)
  - make instructions clear
- Debriefing test users
  - often don’t remember, so demonstrate or show video segments
  - ask for comments on specific features
    - show them screen (online or on paper)

Reporting the Results

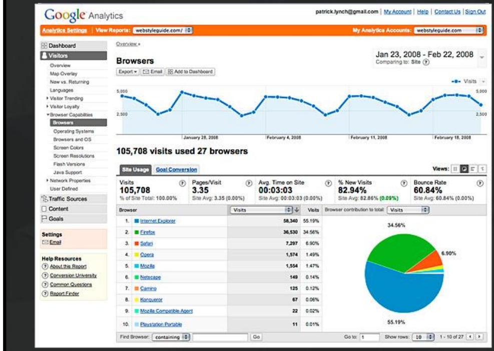
- Report what you did & what happened
- Images & graphs help people get it!
- Video clips can be quite convincing

AUTOMATED & REMOTE USABILITY EVALUATION

Automated Analysis & Remote Testing

- Log analysis
  - infer user behavior by looking at web server logs
- A-B Testing
  - show different user segments different designs
  - requires live site (built) & customer base
  - measure outcomes (profit), but not why?
- Remote user testing
  - similar to in lab, but online (e.g., over Skype)

Web Logs Analysis Difficult



Google Analytics patrick.lynch@gmail.com | My Account | Help | Contact Us | Sign Out

Analytics Settings | View Reports: webstyleguide.com/ My Analytics Accounts: webstyleguide.com

Dashboard
Visitors
Overview
Map Overlay
New vs. Returning
Languages
Visitor Trending
Visitor Loyalty
Browser Capabilities
Browsers
Operating Systems
Browsers and OS
Screen Colors
Screen Resolutions
Flash Versions
Java Support
Network Properties
User Defined
Traffic Sources
Content
Goals

Settings
Email

Help Resources
About this Report
Conversion University
Common Questions
Report Finder

Overview »

Browsers

Export ▾ Email Add to Dashboard

Jan 23, 2008 - Feb 22, 2008
Comparing to: Site

105,708 visits used 27 browsers

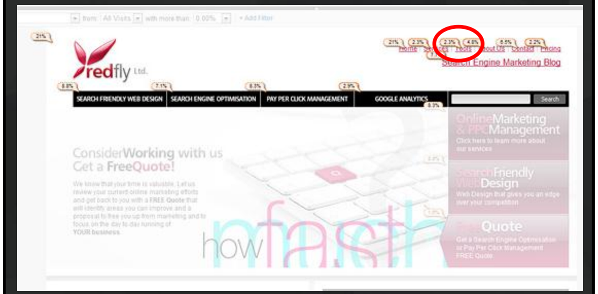
Site Usage | Goal Conversion Views:

Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
105,708	3.35	00:03:03	82.94%	60.84%
% of Site Total: 100.00%	Site Avg: 3.35 (0.00%)	Site Avg: 00:03:03 (0.00%)	Site Avg: 82.86% (0.09%)	Site Avg: 60.84% (0.00%)

	Browser	Visits	Visits
1.	Internet Explorer	58,340	55.19%
2.	Firefox	36,530	34.56%
3.	Safari	7,297	6.90%
4.	Opera	1,574	1.49%
5.	Mozilla	1,554	1.47%
6.	Netscape	149	0.14%
7.	Camino	125	0.12%
8.	Konqueror	67	0.06%
9.	Mozilla Compatible Agent	22	0.02%
10.	Playstation Portable	11	0.01%

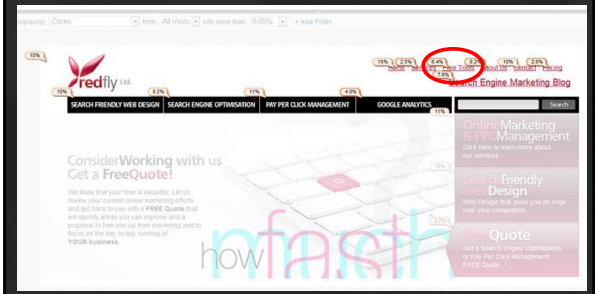
Find Browser: containing Go Go to: 1 Show rows: 10 1 - 10 of 27

Google Analytics – Server Logs++



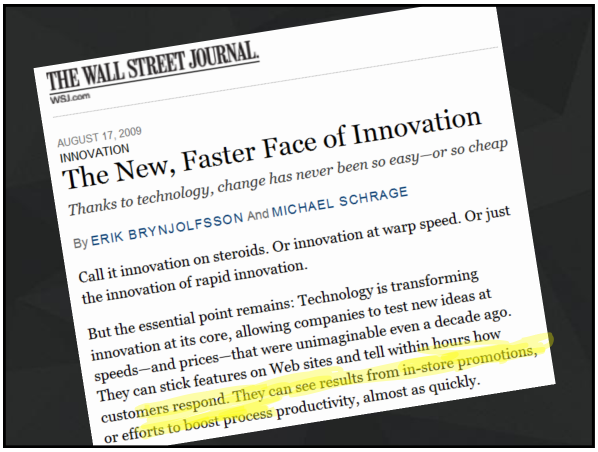
http://www.redflymarketing.com/blog/using-google-analytics-to-improve-conversions/

Google Analytics – Server Logs++



Displaying: Clicks from: All Visits with more than: 0.00% + Add Filter
15%
redfly Ltd.
15%
2.5%
6.4%
6.2%
10%
2.6%
7.6%
Home
Services
Free Tools
About Us
Contact
Pricing
Search Engine Marketing Blog
10%
8.2%
11%
4.0%
SEARCH FRIENDLY WEB DESIGN
SEARCH ENGINE OPTIMISATION
PAY PER CLICK MANAGEMENT
GOOGLE ANALYTICS
Search
Consider Working with us
Get a Free Quote!
Online Marketing & PPC Management
Search Friendly Web Design
Quote
how fast

http://www.redflymarketing.com/blog/using-google-analytics-to-improve-conversions/



THE WALL STREET JOURNAL.
WSJ.com

AUGUST 17, 2009
INNOVATION

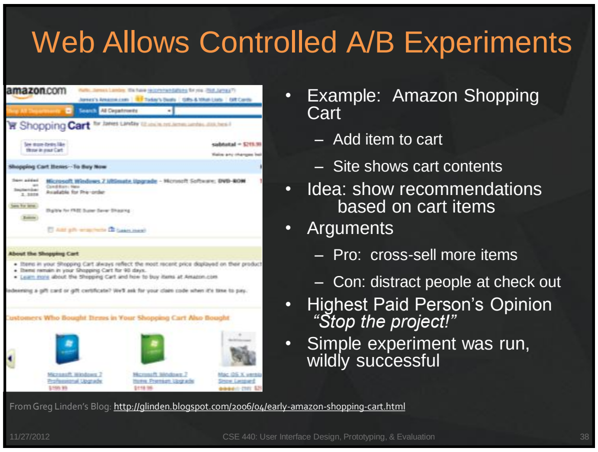
The New, Faster Face of Innovation

Thanks to technology, change has never been so easy—or so cheap

By ERIK BRYNJOLFSSON And MICHAEL SCHRAGE

Call it innovation on steroids. Or innovation at warp speed. Or just the innovation of rapid innovation.

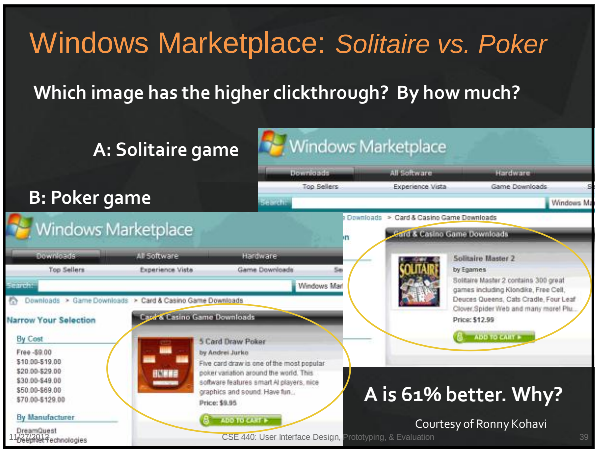
But the essential point remains: Technology is transforming innovation at its core, allowing companies to test new ideas at speeds—and prices—that were unimaginable even a decade ago. They can stick features on Web sites and tell within hours how customers respond. They can see results from in-store promotions, or efforts to boost process productivity, almost as quickly.



Web Allows Controlled A/B Experiments

- Example: Amazon Shopping Cart
  - Add item to cart
  - Site shows cart contents
- Idea: show recommendations based on cart items
- Arguments
  - Pro: cross-sell more items
  - Con: distract people at check out
- Highest Paid Person's Opinion *"Stop the project!"*
- Simple experiment was run, wildly successful

From Greg Linden's Blog: http://glinden.blogspot.com/2006/04/early-amazon-shopping-cart.html



Windows Marketplace: *Solitaire vs. Poker*

Which image has the higher clickthrough? By how much?

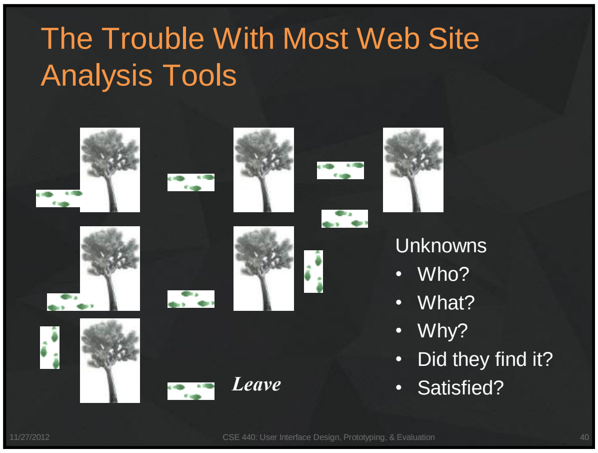
A: Solitaire game

B: Poker game

Windows Marketplace
Downloads
All Software
Hardware
Top Sellers
Experience Vista
Game Downloads
Search:
Downloads > Game Downloads > Card & Casino Game Downloads
Narrow Your Selection
By Cost
Free -$9.00
$10.00-$19.00
$20.00-$29.00
$30.00-$49.00
$50.00-$69.00
$70.00-$129.00
By Manufacturer
DreamQuest
DeepNet Technologies
Card & Casino Game Downloads
5 Card Draw Poker
by Andrei Jurko
Five card draw is one of the most popular poker variation around the world. This software features smart AI players, nice graphics and sound. Have fun...
Price: $9.95
ADD TO CART
Solitaire Master 2
by Egames
Solitaire Master 2 contains 300 great games including Klondike, Free Cell, Deuces Queens, Cats Cradle, Four Leaf Clover, Spider Web and many more! Plu...
Price: $12.99
ADD TO CART

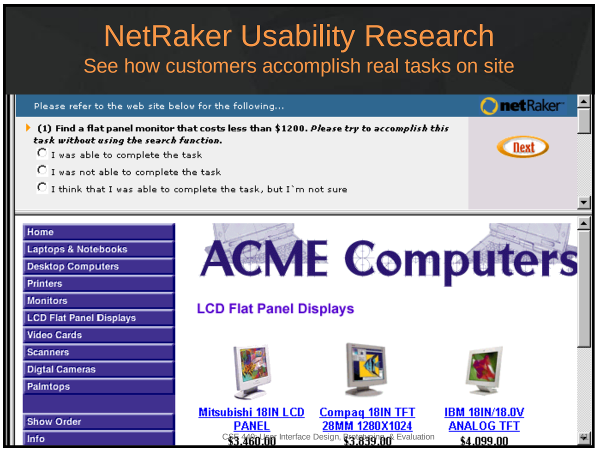
A is 61% better. Why?

Courtesy of Ronny Kohavi



The Trouble With Most Web Site Analysis Tools

Unknowns
- Who?
- What?
- Why?
- Did they find it?
- Satisfied?



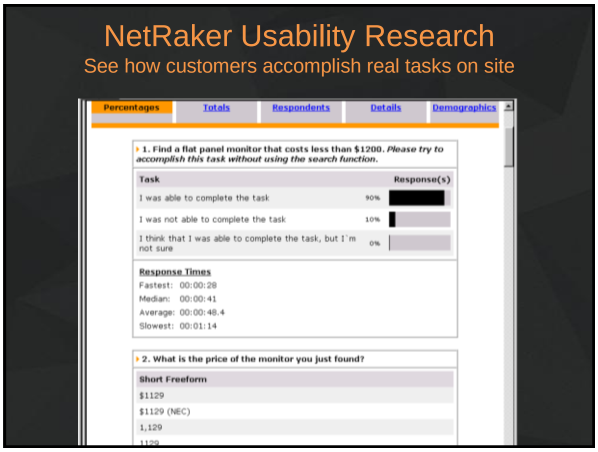
NetRaker Usability Research

See how customers accomplish real tasks on site

Please refer to the web site below for the following...

(1) Find a flat panel monitor that costs less than $1200. *Please try to accomplish this task without using the search function.*

○ I was able to complete the task
○ I was not able to complete the task
○ I think that I was able to complete the task, but I`m not sure



NetRaker Usability Research

See how customers accomplish real tasks on site

1. Find a flat panel monitor that costs less than $1200. *Please try to accomplish this task without using the search function.*

Task		Response(s)
I was able to complete the task	90%	■
I was not able to complete the task	10%	■
I think that I was able to complete the task, but I`m not sure	0%	

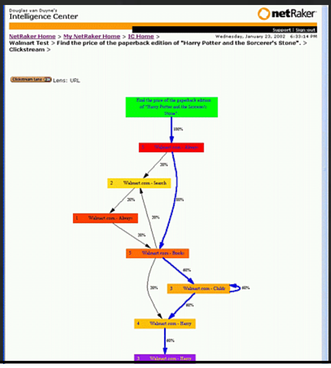
Response Times
Fastest: 00:00:28
Median: 00:00:41
Average: 00:00:48.4
Slowest: 00:01:14

2. What is the price of the monitor you just found?

Short Freeform
$1129
$1129 (NEC)
1,129
1129

NetRaker Usability Research

See how customers accomplish real tasks on site



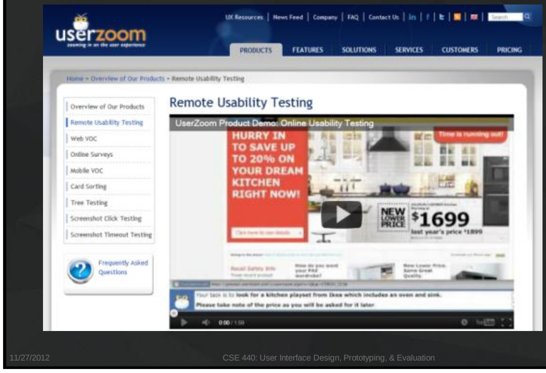
Douglas van Duyne's
Intelligence Center

netRaker

Support | Sign-out

NetRaker Home > My NetRaker Home > IC Home > Wednesday, January 23, 2002 6:33:14 PM
Walmart Test > Find the price of the paperback edition of "Harry Potter and the Sorcerer's Stone". > Clickstream >

UserZoom



userzoom
zooming in on the user experience
UX Resources | News Feed | Company | FAQ | Contact Us
Search
PRODUCTS
FEATURES
SOLUTIONS
SERVICES
CUSTOMERS
PRICING

Home » Overview of Our Products » Remote Usability Testing

Remote Usability Testing

UserZoom Product Demo: Online Usability Testing

Your task is to look for a kitchen playset from Ikea which includes an oven and sink.
Please take note of the price as you will be asked for it later

Advantages of Remote Usability Testing

- Fast
  - can set up research in 3-4 hours
  - get results in 36 hours
- More accurate
  - can run with large samples (50-200 users → stat. sig.)
  - uses real people (customers) performing tasks
  - natural environment (home/work/machine)
- Easy-to-use
  - templates make setting up easy
- Can compare with competitors
  - indexed to national norms

Disadvantages of Remote Usability Testing

- Miss observational feedback
  - facial expressions
  - verbal feedback (critical incidents)
- Need to involve human participants
  - costs some amount of money (typically $20-$50/person)
- People often do not like pop-ups
  - need to be careful when using them

Summary

- User testing is important, but takes time/effort
- Early testing can be done on mock-ups (low-fi)
- Use ????? tasks & ????? participants
  - real tasks & representative participants
- Be ethical & treat your participants well
- Want to know what people are doing & why? collect
  - process data
- Bottom line data requires ???? to get statistically reliable results
  - more participants
- Difference between between & within groups?
  - between groups: everyone participates in one condition
  - within groups: everyone participates in multiple conditions
- Automated usability
  - faster than traditional techniques
  - can involve more participants → convincing data
  - easier to do comparisons across sites
  - tradeoff with losing observational data

Next Time

Interactive Prototype Presentations