

Usability Testing

刘哲明

Prof. James A. Landay
Computer Science Department
Stanford University

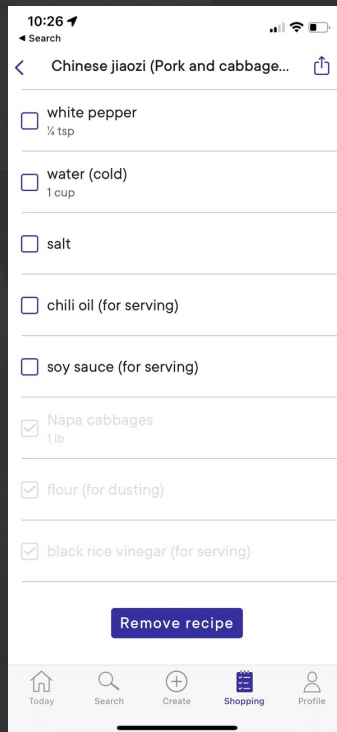
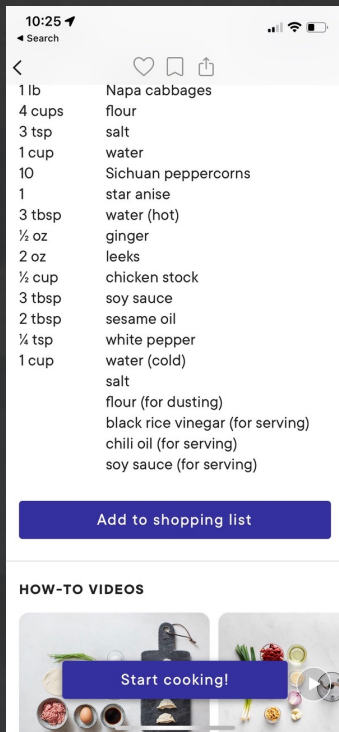
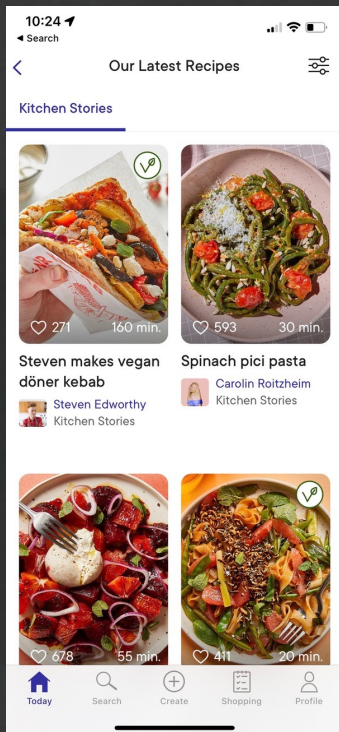
Autumn 2024

November 18, 2024

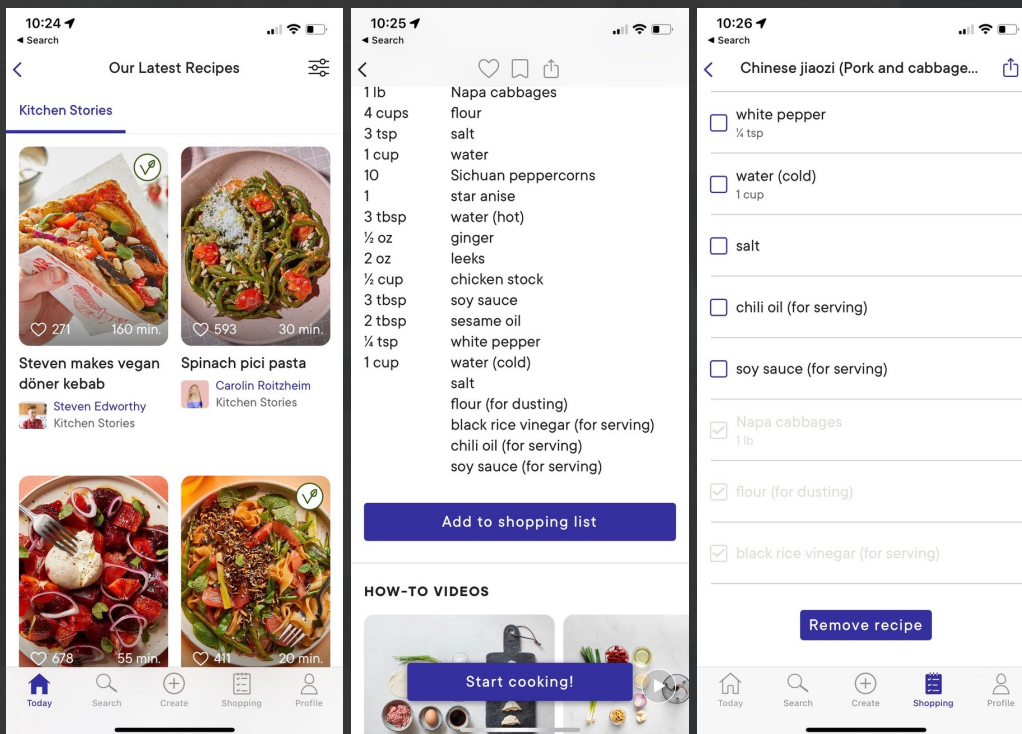
Hall of Fame or Shame?



- Kitchen Stories



Hall of Fame!



- Kitchen Stories

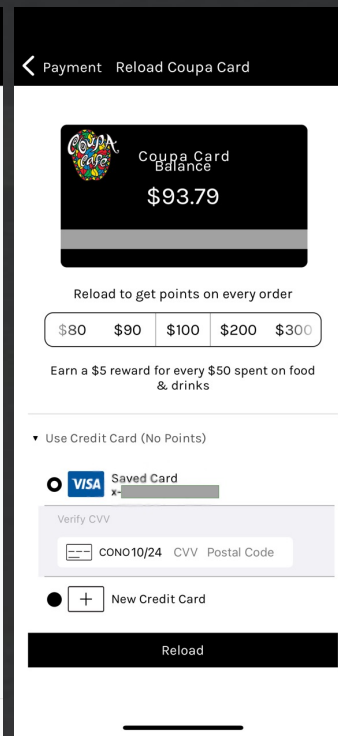
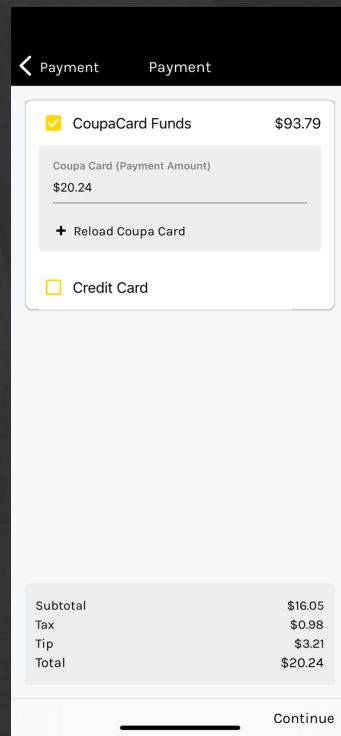
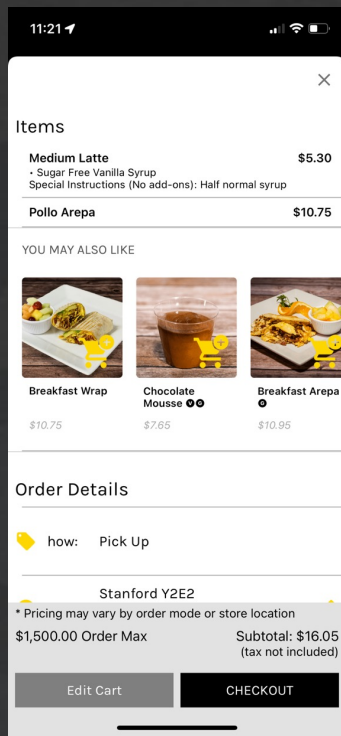
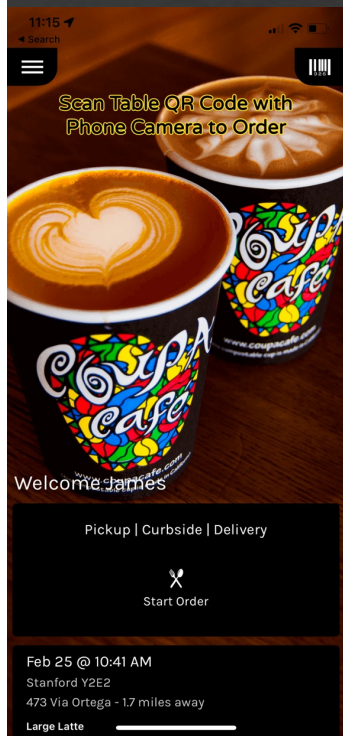
- Like

- large pictures of recipes
- photos & videos
- shopping list that marks off as you purchase

- Wish

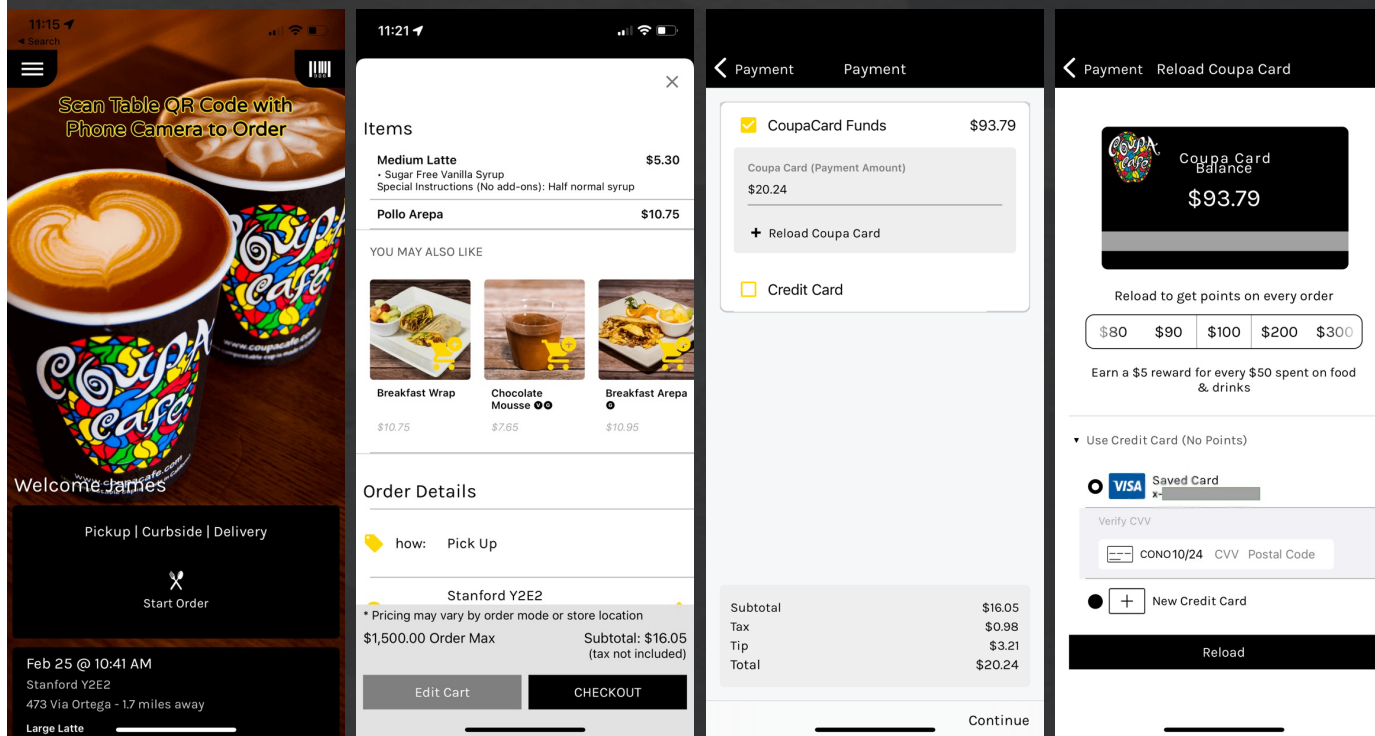
- fonts hard to read for long list
- why the tab for “Kitchen Stories”?

Hall of Fame or Shame?



- Coupa Café

Hall of Shame!



- Coupa Café
- Like
 - remembers my usual café
- Wish
 - how much money am I adding?
 - lots of steps to order & checkout

Hall of Fame or Hall of Shame?



جهت دریافت و چاپ مدارک قانونی تان لطفاً به لینک مراجعه فرمائید [I94-OAW-Dari.pdf](#)

د لا موندلو او قانونی اسنادو د چاپولو لپاره مهرباني وکړی لاندی لینک کیکړئ [I94-OAW-Pashtu.pdf](#)

Pre-paying for a provisional I-94 through this web site does not guarantee admission into the United States.

Official Website of the Department of Homeland Security

U.S. Customs and Border Protection I-94 Website
Travel Record for U.S. Visitors

Home Apply Now Get Most Recent I-94 View Travel History View Compliance Help

Official Site for Travelers Visiting the United States: Apply for or retrieve Form I-94, Request Travel History and Check Travel Compliance

International travelers visiting the United States can apply for or retrieve their I-94 admission number/record (which is proof of legal visitor status) as well as retrieve a limited travel history of their U.S. arrivals and departures.

For instructions on accessing and printing your I-94 using your Alien Registration Number (A number), please [Click here.](#)

Para obtener instrucciones sobre como acceder e imprimir su registro de viaje Forma I-94, haga [clic aquí.](#)

Visiting the U.S. & Arriving Via a Land Border or Selected Ferry?

A Form I-94 is needed by all visitors except: U.S. Citizens, returning resident aliens, aliens with immigrant visas, and most Canadian citizens visiting or in transit. Travelers will be issued an I-94 during the admission process at the port of entry. If you are traveling via a land border or by ferry from selected ports, you may apply for an I-94 in advance here, saving time while at the port of entry later.

Reference FAQs for specific ferry ports accepting I-94 Provisionals.

[Need to Learn How to Apply for an I-94? Watch this video here](#)

Apply for new I-94

Already Visiting & Need Proof of Visitor Status?

A traveler lawfully admitted (or paroled) into the U.S. may look up their most recent Form I-94 going back to 1983 for most classes of admission (or parole), and indefinitely for certain classes, such as diplomats and those admitted under the Compacts of Free Association. The corresponding Form I-94 may then be printed here. This is considered your lawful record of admission.

If someone requests your admission information, this is the form you would provide.

Get most recent I-94

Need A History Of Your Arrivals & How Much Longer May I Remain in

Privacy - To

U.S. I-94 Website

I-94: A document that gives you your travel history to and from the U.S.

Hall of Shame!



جهت دریافت و چاپ مدارک قانونی تان لطفاً به لینک مراجعه فرمائید [I94-OAW-Dari.pdf](#)

د لا موندلو او قانونی اسنادو د چاپولو لپاره مهرباني وکړئ لینک کيکړئ [I94-OAW-Pashtu.pdf](#)

Pre-paying for a provisional I-94 through this web site does not guarantee admission into the United States.

Official Website of the Department of Homeland Security

U.S. Customs and Border Protection I-94 Website
Travel Record for U.S. Visitors

Home Apply Now Get Most Recent I-94 View Travel History View Compliance Help

Official Site for Travelers Visiting the United States: Apply for or retrieve Form I-94, Request Travel History and Check Travel Compliance

International travelers visiting the United States can apply for or retrieve their I-94 admission number/record (which is proof of legal visitor status) as well as retrieve a limited travel history of their U.S. arrivals and departures.

Visiting the U.S. & Arriving Via a Land Border or Selected Ferry?

A Form I-94 is needed by all visitors except: U.S. Citizens, returning resident aliens, aliens with immigrant visas, and most Canadian citizens visiting or in transit. Travelers will be issued an I-94 during the admission process at the port of entry. If you are traveling via a land border or by ferry from selected ports, you may apply for an I-94 in advance here, saving time while at the port of entry later.

Reference FAQs for specific ferry ports accepting I-94 Provisionals.

[Need to Learn How to Apply for an I-94? Watch this video here](#)

Apply for new I-94

Already Visiting & Need Proof of Visitor Status?

A traveler lawfully admitted (or paroled) into the U.S. may look up their most recent Form I-94 going back to 1983 for most classes of admission (or parole), and indefinitely for certain classes, such as diplomats and those admitted under the Compacts of Free Association. The corresponding Form I-94 may then be printed here. This is considered your lawful record of admission.

If someone requests your admission information, this is the form you would provide.

Get most recent I-94

For instructions on accessing and printing your I-94 using your Alien Registration Number (A number), please [Click here](#).

Para obtener instrucciones sobre como acceder e imprimir su registro de viaje Forma I-94, haga [clic aquí](#).

Need A History Of Your Arrivals & How Much Longer May I Remain in

- Obscure PDFs and use of Dari/Pashto at the top grab attention, but unclear what they are (to non-Afghan users)
- **“Get most recent I-94”** should be central action but is one of many tabs
- Distracting background
- Black overlay behind text doesn’t provide enough contrast
- Too much going on

Usability Testing

刘哲明

Prof. James A. Landay
Computer Science Department
Stanford University

Autumn 2024

November 18, 2024

Outline

- Why do usability testing?
- Choosing participants
- Ethical considerations
- Designing & conducting the test
- Using the results
- Experimental options & details

Why do Usability Testing?

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



Choosing Participants

- Representative of target users. How so?
 - 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
 - t-shirt, mug, free coffee/pizza



Ethical Considerations

- Usability tests can be distressing
 - users have left in tears
- Testing/fieldwork can be coercive if there is a power imbalance (e.g., in under resourced communities)



<http://centread.ucsc.edu/CenTREAD%20photos/BrianDowd2.JPG>

People may feel no option but to speak to you or give you their time even though they may not get anything of value in return.

Ethical Considerations

- You have a responsibility to alleviate these issues
 - 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 (IRB)



<https://www.unthsc.edu/north-texas-regional-irb/institutional-review-board-meeting/>

Usability 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

- Tasks from low-fi design can be used
 - may need to shorten if
 - they take too long
 - require background that test user won't have
- Don't train unless that will occur in real deployment
- Avoid bending tasks in direction of what your design best supports
- Don't choose tasks that are too fragmented ?
 - fragmented = does not represent a complete goal someone would try to accomplish with your application
 - e.g., phone-in bank test or login/create account as a task



Two Types of Data to Collect

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



<http://allazollers.com/discovery-research.php>



<http://www.fusionfarm.com/content/uploads/2012/10/analyzing-data.jpg>

Which Type of Data to Collect?

- Focus on process data first
 - gives good overview of where problems are



http://www.redicecreations.com/ul_img/24592nazca_bird.jpg

Which Type of Data to Collect?

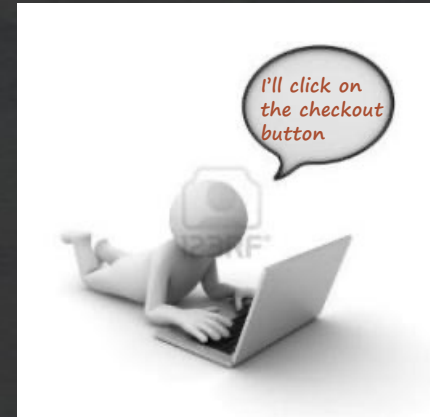
- Focus on process data first
 - gives good overview of where problems are
- Bottom-line data doesn't tell you ?
 - where to fix
 - just says: “too slow”, “too many errors”, etc.
- Hard to get reliable bottom-line results
 - need many users for statistical significance



<http://netbreak-ph.blogspot.com/2012/09/HitLeap-get-traffic-views-on-site.html>

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*



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
- Make a *recording* & take good notes
 - make sure you can tell what they were doing
 - use a digital watch/clock
 - record audio & video
 - or even event logs



<http://jennyham.co.uk/wp-content/uploads/2011/08/200911221250225481.jpg>

Will thinking out loud give the right answers?

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

→ Try to avoid specific questions (especially that have binary answers)



Accessibility Issues with Thinking Out Loud

- May be less comfortable or accessible for certain users
- Factor in time/resources you need to support these interviewees
 - interpreters
 - more time for the interview
 - e.g., someone using sign language can't give feedback while interacting with your prototype



https://blog.gettinghired.com/hubfs/GettingHired_November2019/Images/CSD%20blog93019660x440%20jpeg.jpeg

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

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 these laws and regulations.

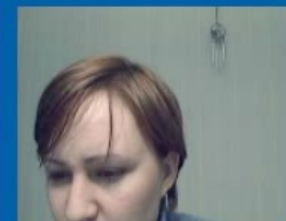
Provincially regulated -- Select a Province --

- 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.

Done Internet



Try it out!

Use the **think aloud protocol** to test **one task** in your medium-fi prototype.

In a moment, we will break you out into your teams. You have 3 minutes to identify which task to study. Select **one person** to rotate into the group next to you to serve as their participant. If possible, this should be someone who ***has not done a HE*** on that team's prototype.

When they are ready, ask your participant to complete the task while speaking their thought process aloud. Take notes!

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?



<https://cdn.careerfoundry.com/en/wp-content/uploads/old-blog-uploads/data-analyst-skills-and-tasks.jpeg>

Using the Results (cont.)

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



<http://www.thetomorrowplan.com/exchange/policies-prairie-chickens-and-parking/>

Administrivia

- Group project grades (60% of your overall grade)
 - your project grade can be impacted by your contribution to the team → pull your weight!
 - midway due studio this week (show ≥ 1 task working!)
- Web sites now linked from class projects page
 - let @stardoby know if you updated your tagline or project name (check)
- Assignments due the final week
 - Poster & pitch slide drafts (Mon) & finals (Wed at noon)
 - Video demo (Wed)
 - Prototype (Fri)
 - Report (Sun)

TEAM BREAK

Work on (or plan for)
building your hi-fi prototype

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

- Example: trying to get task time ≤ 30 min.
 - test gives: 40, 5, 20, 90, 10, 15
 - 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



Analyzing the Numbers (cont.)

- This is what basic statistics can be used 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 4.6 & 55.4

Analyzing the Numbers (cont.)

- This is what basic statistics can be used for
- Crank through the procedures and you find
 - 95% certain that typical value is between 5 & 55
- Usability test data is *highly variable*
 - need lots to get good estimates of typical values
 - 4x 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, unrealistic 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 (cheaper)
 - can't use the same tasks or order (learning)
 - best for low-level interaction techniques
 - e.g., new mouse, new swipe interaction, ...



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 → counterbalance order
- 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



Heuristic Evaluation vs. User Testing

- HE is much faster
 - 2-4 hours each evaluator vs. days-weeks
- HE doesn't require interpreting user's actions
- User testing is far more accurate (by def.)
 - takes into account actual users and tasks
 - HE may miss problems & find “false positives”
- Good to alternate between HE & user testing
 - find different problems
 - don't waste participants

Summary

- User testing is important, but takes time/effort
- 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: each subject participates in only one of n conditions
 - within groups: everyone participates in multiple conditions

Further Reading on Ethical Issues With Community-based Research

- Children and Families “At Promise, Beth B. Swadener, Sally Lubeck, editors, SUNY Press, 1995, <http://www.sunypress.edu/p-2029-children-and-families-at-promis.aspx>
- “Yours is better!” Participant Response Bias in HCI, Proceedings of CHI 2012, by Nicola Dell, et al., <http://research.microsoft.com/pubs/163718/CHI2012-Dell-ResponseBias-proc.pdf>
- “Strangers at the Gate: Gaining Access, Building Rapport, and Co-Constructing Community-Based Research”, Proceedings of CSCW 2015, by Christopher A. Le Dantec & Sarah Fox, <http://dl.acm.org/citation.cfm?id=2675133.2675147&coll=DL&dl=ACM>
- “Imperialist Tendencies” blog post by Jan Chipchase, <http://janchipchase.com/content/essays/imperialist-tendencies/>
- “To Hell with Good Intentions” by Ivan Illich, speech to the Conference on InterAmerican Student Projects (CIASP), April 20, 1968, <https://www.southwestern.edu/live/files/1158>

Exit ticket



<https://tinyurl.com/2024au-exit-ticket-9-372>

Next Time

- Design Patterns
 - Read from the *Design of Sites*
 - “Making the Most of Web Design Patterns” (Ch 2)
 - “Up-Front Value Proposition” (Pattern C2)
 - “Process Funnel” (Pattern H1)
 - “Meaningful Error Messages” (Pattern K13)
- Studio
 - A8 midway – show at least 1 of your tasks implemented