UX Verification and Validation Methods

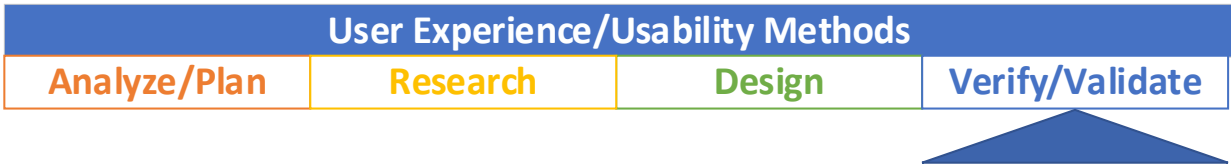**Embedded Interface Design**

with **Bruce Montgomery**

Learning Objectives

Students will be able to…

- Understand considerations for selecting methods and their qualitative vs. quantitative results
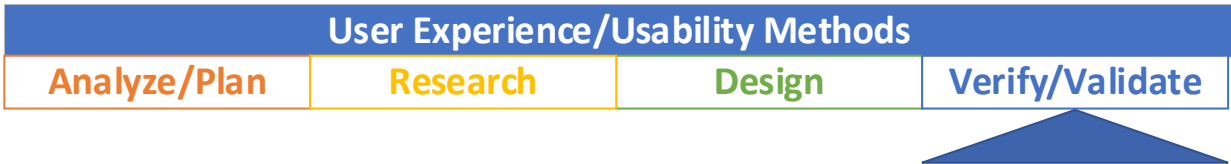- Review, select, and apply selected UX Verify and Validate methods

UX Verify/Validate Methods

Selected Methods for Review
- "Hallway" Usability Test
- Black Hat Session
- First Click Test
- UX Health Check
- A/B Testing
- Heuristic Evaluation (again)
- Formal Usability Tests
- Surveys (separate lecture)
- Statistics and Measures (separate lecture)

Qualitative vs. Quantitative

- The test methods we use here result in qualitative and quantitative results
- Qualitative results are usually based on user comments or tester observations during a given evaluation
  - In some cases, these results could be assigned as classifiers to provide a more quantitative assessment
- Quantitative results of tests can come from numerical measurements, counts, timing, or scaled answers
  - We'll look more at quantifiable results in the UX statistics discussion
- In either case, regardless of statistical strength of results, findings from these test methods can guide the team toward valuable design decisions

Selection Criteria

When looking at which methods are appropriate to your design cycle, consider these following method attributes
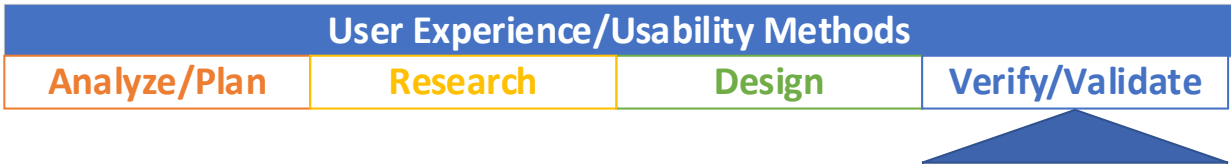
- Time (preparation, execution, follow-up)
- Complexity
- Potential results – qualitative or quantitative
- Fit to your overall project – available resources, skills, level of detail, etc.
- Fit to your goal for the test – what do you need to learn?

University of Colorado **Boulder**

When to test?

- Again, although this is listed as a separate phase, your project may find some verification during any of the earlier phases is valuable
- It's also clearly valuable to find issues as soon as possible, so consider using a method whenever there are design elements that may lend themselves to one of the test methods
- Generally, consider testing sooner than you normally might
- Counterpoint: Does the project support taking action based on test results?
  - i.e., don't waste time testing if changes resulting from the test will not be applied

"Hallway" Usability Test

- Time: Potentially very quick, less than an hour, but may extend with preparation or complexity
- Results: Generally qualitative, but could be used for select quantitative data
- Essentially, this is testing a design or task with anyone available
  - Although fellow developers or engineers should generally not be used
- Steps:
  - Can be with random users, can be from a more targeted group
  - Can be a single tester or observer, moderator, and greeter
  - Ask user to attempt a described task with your design
  - Doesn't require expert users or staff
  - Use as part of iterating test and fix cycle;
    stop and make fixes if severe issues are observed
  - Can be done using online tools in some cases
- Reference [1]

"Hallway" Usability Test
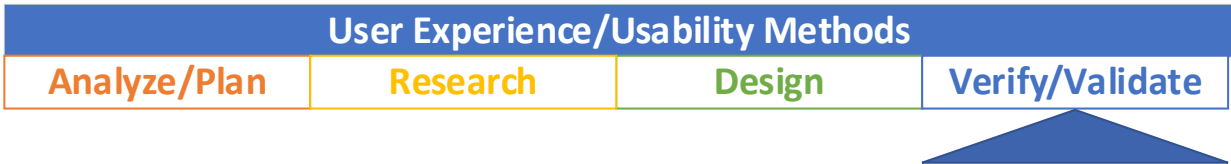
- Tips:
  - Location with foot traffic
  - Pre-preparing materials helps
  - Take some time to setup
  - Use greeters to find testers
  - Be mindful of time (ten minutes is optimum)
  - Minimize number of users
    (three to five is usually sufficient)
  - Explain the purpose (quickly)
  - Reward volunteers (if only with thanks)
  - Debrief after each test
  - Note possible improvements for later
    prioritization or actions
- From [2]

Black Hat Test

- From DeBono's Six Thinking Hats assessment process
  - White Hat – Facts, information
  - Yellow Hat – Positives, value, benefit
  - Red Hat – Feelings, emotions, intuition
  - Green Hat – Creative, new concepts
  - Blue Hat – Control, process
  - Black Hat – Judgement, devil's advocate, why it won't work
  - Reference [3]
- Black Hat Test – spot the risks and weaknesses
- Give permission to reviewers to be negative and direct

Black Hat Test

- Time – hours or less, some setup and preparation
- Results – Generally qualitative, not designed for quantitative results
- Safe process to get negative feedback from reviewers
- Shows you are not protective/defensive, feedback from key groups
- Set up as a group work session
  - explain rules (think grumpy/skeptical user or tough/senior leader)
  - start clock to annotate drawings of design with issues
  - look for themes in issues found (affinity grouping)
- Discuss findings and close with actions to take and/or a review of good elements of design (to end positively)
- Use at any point, focus on specific issues if helpful
- Reference [4]

Try a Black Hat Test

- Spot the risks and weaknesses
- Be negative and direct
- Think grumpy/hurried user
- What do you see as usability issues for this design
- Write down at least three issues

Try a Black Hat Test

- Spot the risks and weaknesses
- Be negative and direct
- Think grumpy/hurried user
- What do you see as usability issues for this design
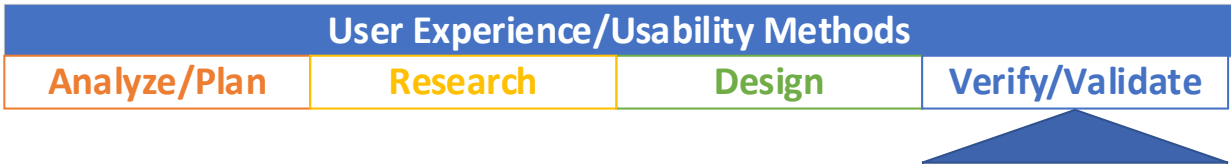- Write down at least three issues

"First Click" Test

- Time: brief test, but some test preparation
- Qualitative results, but could track success/failure vs. expected outcome
- Test looks at what a test participant selects first in an interface to complete a given task
- Comes from web pages, but can be used on any product or interface, or earlier with wireframes or prototypes
- Research shows selecting the first click correctly means a user will complete a task correctly 87% of the time; selecting the wrong path yields a 46% task completion [5]
- Reference [6]

UX Health Check

- Time: Ongoing over project, revisit periodically
- Qualitative measures generally
- Measure baseline quality; assess changes over time
- Steps
  - Break product into sections by functional area
  - Find a comparative or competitive benchmark
  - Rate each area as % as good as the benchmark
  - Identify improvement opportunities
  - Capture measures over time in a spreadsheet
- Use to assess UX improvement over time
- Prioritize features/elements needing help
- Reference [7]

A/B Testing

- Time: Hours to days, depending on desired exposure and preparation
- Results: Quantitative (measures taken in A/B)
- In most A/B testing a group of users is presented with two alternate designs for an interface
- Their interaction is measured for time, errors, satisfaction, etc. as appropriate
- Poor use of A/B tests: issues in implementation vs. concept, no research for actual cause of issues, variation based on guesses
- Best to perform usability tests on A/B designs to remove obvious issues before A/B testing
- Reference [8]

Heuristic Evaluation

- Time: Hours, also resource bound to find appropriate reviewers
- Results: Generally qualitative
- In a heuristic evaluation, usability experts review an interface and compare it against accepted usability principles. The analysis results in a list of potential usability issues. (See prior discussions of Nielsen's heuristics and others.)
- Cannot replace user-based testing
- Will likely identify different issues than found in a user-based usability test
- Reference [9]

Formal Usability Testing

- Time and resources high, provides both quantitative and qualitative measures
- Dumas/Redish [10] and Rubin/Chisnell [11] are classic textbooks on more formal usability testing
- Typical key concerns from [11]:
  - when to test
  - skills for testers
  - test planning
  - test environments
  - finding and selecting participants
  - preparing materials
  - conducting tests
  - analyzing data and reporting findings
- Books provide examples templates for materials and presentations

Formal Usability Testing – Guidelines
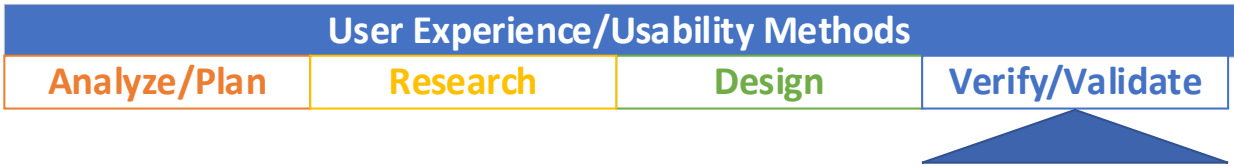
- Different testing approaches may be warranted based on project phase and goals of testing…
- From Usability Test Basics [12]
  - Benchmark Metrics – 8 to 24 users, test current product or process
  - Diagnostic Evaluation – 4 to 6 users for qualitative iterative tests
  - Summative Testing – 6 to 12+ users for formal metrics to assess usability goals
- Numbers of users can change based on need for statistical strength in reporting
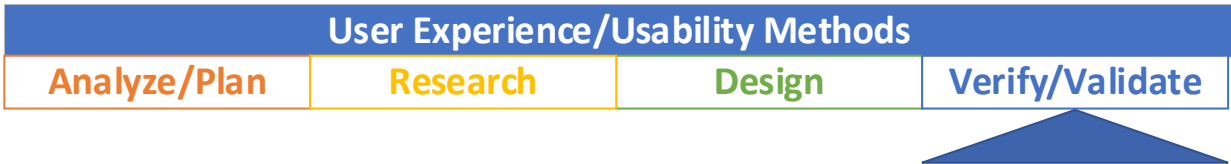
Formal Usability Testing – Materials
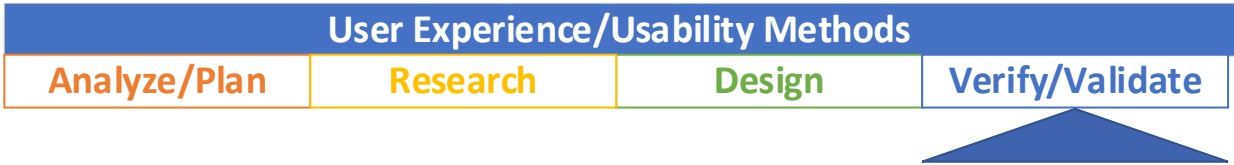
- Recruiting sheets/payment
- IRB Approvals
- Release/consent forms (esp. for recording)
- Test plan
- Study overview
- Process & materials checklist
- Testing scripts
- Report template
- Post-test surveys (can be fee-based)
- Reference [13]

Formal Usability Testing – Usability Test Reports

- ISO/IEC 25062:2005 ANSI/NIST Common Industry Format (CIF) for Usability Test Reports
- CIF content:
  - the description of the product
  - the goals of the test
  - the test participants
  - the tasks the users were asked to perform
  - the experimental design of the test
  - the method or process by which the test was conducted
  - the usability measures and data collection methods
  - the numerical results
- Reference [14]

University of Colorado **Boulder**

Formal Usability Testing – Lab-based Tests

Usability Testing Lab
emeraldsong.com
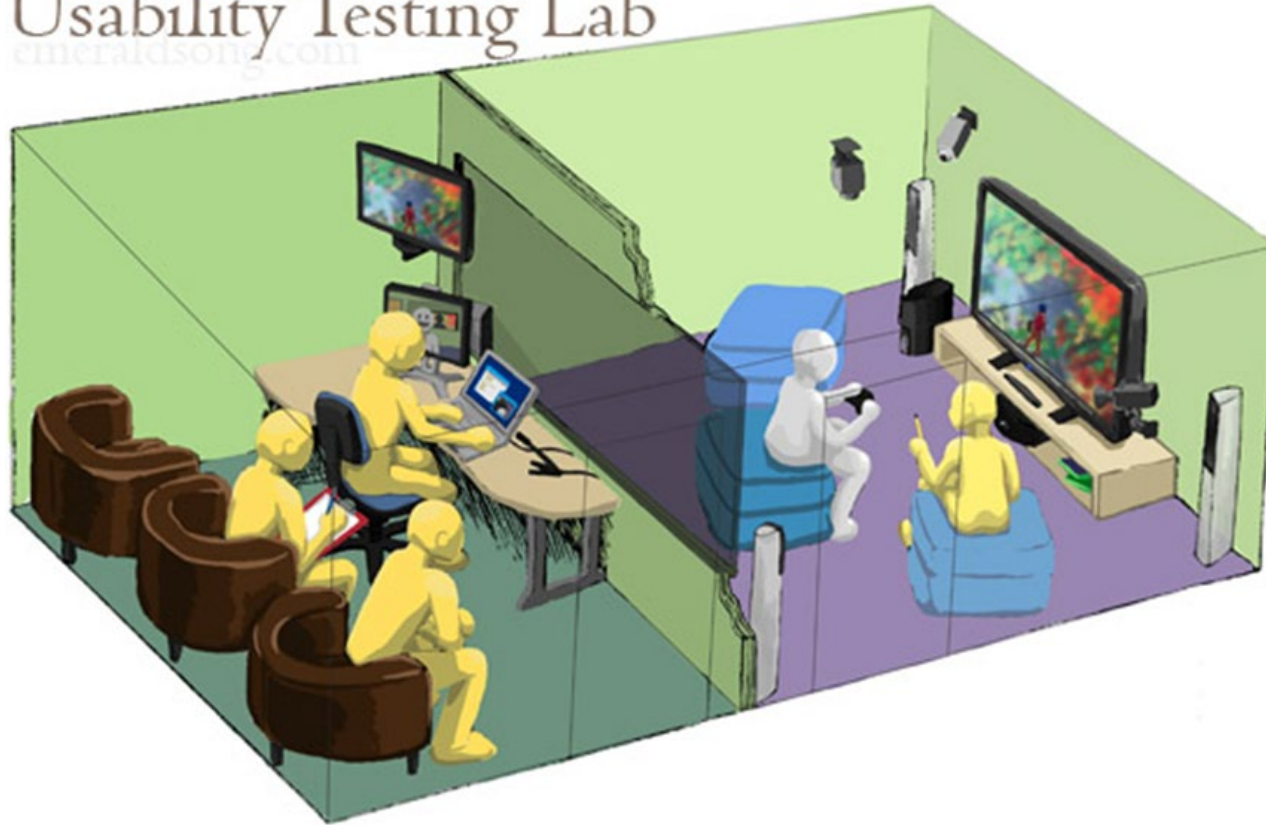
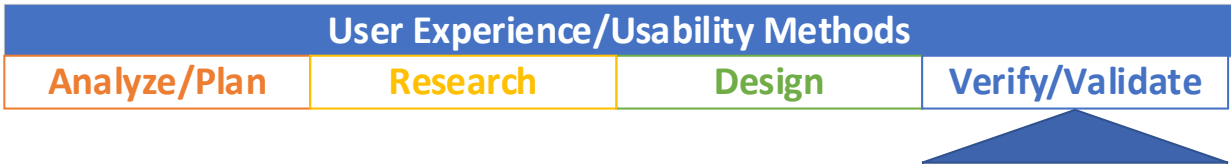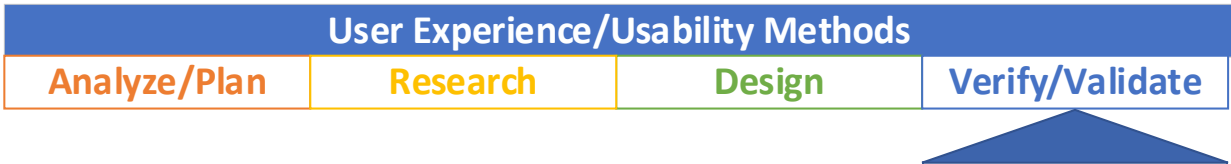| User Experience/Usability Methods			
Analyze/Plan	Research	Design	Verify/Validate

- Typical elements: Observers, users, moderators, technicians; video and audio recording; comfortable area for users [15]

Formal Usability Testing vs. Discount Testing

- Comparing formal vs. discount test methods
- Days and high effort vs. hours
- Discrete points vs. continuous iteration
- Recruiting carefully vs. loosely
- Laboratory vs. ad hoc setup
- Formal reports vs. action lists
- Video vs. observation
- High vs. low cost to perform
- Reference [16]

Summary

- UX verification and validation tests range from informal to formal, general to focused, and provide both quantitative and qualitative feedback
- Always know what the goal of a test is prior to its design, and be prepared to address issues found
- But also be open to discovery that will likely occur in user interactions
- Test early and often, and enjoy the benefits of actual user feedback on designs
- Lets dig into UX surveys and statistics next

Next Steps

- Project 4 is active…
- Working on ordering gear, I'll get with you
- Next week: Surveys, Statistics, Designing Connected Products, Prototypes to Product, Project 5
- Quiz – two sketches – due ~~Wed~~ Friday 11/1, voting the next Monday – assignment is on Canvas
- Class staff available to help
  - Shubham - Tues 12-2 PM, Fri 3-5 PM in ECEE 1B24
  - Sharanjeet - Tues 2-3 PM, Thur 2-3 PM in ECEE 1B24
  - Bruce - Tue 9:30-10:30 AM, Thur 1-2 PM in ECOT 242
- Final Exam is set
  - Tuesday Dec 17 7:30 PM - 10 PM ECCR 1B51
  - Final will be open notes and Canvas based, you'll need a PC

References

[1] https://medium.com/@maureenkerwin/note-this-post-was-originally-published-on-my-internal-company-blog-at-work-5a3ee7b496ae

[2] https://digital.gov/2014/02/19/10-tips-for-better-hallway-usability-testing/

[3] http://www.debonogroup.com/six_thinking_hats.php

[4] https://www.testuff.com/black-hat-software-testing-its-not-what-you-think/

[5] https://measuringu.com/first-click/

[6] https://www.usability.gov/how-to-and-tools/methods/first-click-testing.html

[7] User Experience Ream of One, Buley, 2013, Rosenfeld

[8] https://www.nngroup.com/articles/ab-testing-and-ux-research/

[9] https://www.usability.gov/how-to-and-tools/methods/heuristic-evaluation.html

[10] A Practical Guide to Usability Testing, Dumas & Redish, 1999, Intellect

[11] Handbook of Usability Testing, Rubin & Chisnell, 2008, Wiley

[12] http://webservices.itcs.umich.edu/drupal/wwwsig/sites/webservices.itcs.umich.edu.drupal.wwwsig/files/Usability-Testing-Basics.pdf

[13] https://www.usability.gov/how-to-and-tools/resources/templates.html

[14] http://zing.ncsl.nist.gov/iusr/documents/cifv1.1b.htm

[15] http://www.emeraldsong.com/2011/08/02/usability-testing-article-over-at-altdevblogaday/zrsc_usability_lab_cutaway_ypak-copy/

[16] Rocket Surgery Made Easy, Krug, 2010, New Riders