Predictive Evaluation

John Stasko

Spring 2007

This material has been developed by Georgia Tech HCI faculty, and continues to evolve. Contributors include Gregory Abowd, Al Badre, Jim Foley, Elizabeth Mynatt, Jeff Pierce, Colin Potts, Chris Shaw, John Stasko, and Bruce Walker. Permission is granted to use with acknowledgement for non-profit purposes. Last revision: January 2007.

Agenda

- Evaluation
  - Overview

- Predictive evaluation
  - Heuristic evaluation
  - Discount usability testing
  - Cognitive walkthrough

Evaluation

- Gathering data about usability of a design by a <u>specified group of users</u> for a <u>particular activity</u> within a <u>specified environment</u>

Goals

- 1. Assess extent of system's functionality

- 2. Assess effect of interface on user

- 3. Identify specific problems with system

Forms

- Formative
  - As project is forming. All through the lifecycle. Early, continuous. iterative.
  - "Evaluating the design"

- Summative
  - After a system has been finished. Make judgments about final item.
  - "Evaluating the implementation"

Approaches

- Experimental (Lab studies, quantitative)
  - Typically in a closed, lab setting Manipulate independent variables to see effect on dependent variables

- Naturalistic (Field studies, qualitative)
  - Observation occurs in "real life" setting Watch process over time

Tradeoffs

• Experimental	• Naturalistic
+ Replicable + More "objective"	+ "Ecologically valid" + Cheap, quick
- Expensive, requires real users & lab - Realistic?	- Not reproducible, user-specific results - Not quantitative (how much better?)

Evaluation Methods

- 1. Experimental/Observational Evaluation
  - Typically with users
  - Experiments (usability specifications)

Later in course

- 2. Predictive Evaluation (without users)

Predictive Evaluation

- Basis:
  - Observing users can be time-consuming and expensive
  - Try to predict usage rather than observing it directly
  - Conserve resources (quick & low cost)

Approach

- Expert reviews (often used)
  - HCI experts (not real users) interact with system, try to find potential problems, and give prescriptive feedback

- Best if
  - Haven't used earlier prototype
  - Familiar with domain or task
  - Understand user perspectives

Predictive Eval. Methods

- 1. Heuristic Evaluation

- 2. Discount usability testing

- 3. Cognitive Walkthrough

---

1. Heuristic Evaluation

- Developed by Jakob Nielsen

  (www.useit.com)

- Several expert usability evaluators assess system based on simple and general heuristics (principles or rules of thumb)

  Essay: http://www.useit.com/papers/guerrilla_hci.html

Procedure

- 1. Gather inputs
- 2. Evaluate system
- 3. Debriefing and collection
- 4. Severity rating

Gather Inputs

- Who are evaluators?
  - Need to learn about domain, its practices

- Get the prototype to be studied
  - May vary from mock-ups and storyboards to a working system

Evaluation Method

- Reviewers evaluate system based on high-level <u>heuristics (i.e., usability principles)</u>:

- use simple and natural dialog
- speak user's language
- minimize memory load
- be consistent
- provide feedback

- provide clearly marked exits
- provide shortcuts
- provide good error messages
- prevent errors

Updated Heuristics

- Stresses

- visibility of system status
- aesthetic and minimalist design
- user control and freedom
- consistency and standards
- error prevention

- recognition rather than recall
- flexibility and efficiency of use
- recognition, diagnosis and recovery from errors
- help and documentation
- match between system and real world

Process

- Perform two or more passes through system inspecting
  - Flow from screen to screen
  - Each screen
- Evaluate against heuristics
- Find "problems"
  - Subjective (if you think it is, it is)
  - Don't dwell on whether it is or isn't

Debriefing

- Organize all problems found by different reviewers
  - At this point, decide what are and aren't problems
  - Group, structure
  - Document and record them

Severity Rating

- 0-4 rating scale
  - 4 is the most severe

- Based on
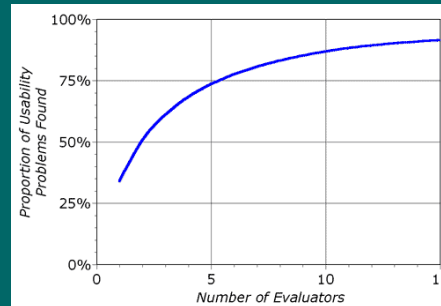 - frequency
  - impact
  - persistence
  - market impact

Advantages

- Cheap, good for small companies who can't afford more

- Getting someone practiced in method is valuable

Application

- Nielsen found that about 5 evaluations found 75% of the problems

- Above that you get more, but at decreasing efficiency

Somewhat Controversial

- Very subjective assessment of problems
  - Depends of expertise of reviewers
- Why are these the right heuristics?
  - Others have been suggested
- How to determine what is a true usability problem
  - Some recent papers suggest that many identified "problems" really aren't

2. Discount Usability Testing

- Hybrid of empirical usability testing and heuristic evaluation

- Have 2 or 3 think-aloud user sessions with paper or prototype-produced mock-ups

Discount Usability in Action

- Mockups are not supposed to be perfect!

- A variety of approaches for mockups:
  - Must be quick to create; economical in use of resources
  - Sketches most common
  - Paper has its limitations; tends to focus on the visual elements
  - Sometimes awkward to use in usability testing

3. Cognitive Walkthrough

- Assess learnability and usability through simulation of way users explore and become familiar with interactive system

- A usability "thought experiment"

- Like code walkthrough in s/w engineering

- From Polson, Lewis, et al at UC Boulder

CW Process

- Construct carefully designed tasks from system spec or screen mock-up

- Walk through (cognitive & operational) activities required to go from one screen to another

- Review actions needed for task, attempt to predict how users would behave and what problems they'll encounter

Requirements

- Description of users and their backgrounds

- Description of task user is to perform

- Complete list of the actions required to complete task

- Prototype or description of system

Assumptions

- User has rough plan

- User explores system, looking for actions to contribute to performance of action

- User selects action seems best for desired goal

- User interprets response and assesses whether progress has been made toward completing task

Methodology

- Step through action sequence
  - Action 1
  - Response A, B, ..
  - Action 2
  - Response A
  - ...

- For each one, ask <u>four questions</u> and try to construct a <u>believability story</u>

CW Questions

- 1. Will users be trying to produce whatever effect action has?

- 2. Will users be able to notice that correct action is available?

- 3. Once found, will they know it's the right action for desired effect?

- 4. Will users understand feedback after action?

Answering the Questions

- 1. Will user be trying to produce effect?
    - Typical supporting Evidence
        - It is part of their original task
        - They have experience using the system
        - The system tells them to do it
    - No evidence?
        - Construct a failure scenario
        - Explain, back up opinion

Next Question

- 2.Will user notice action is available?
    - Typical supporting evidence
        - Experience
        - Visible device, such as a button
        - Perceivable representation of an action such as a menu item

Next Question

- 3.Will user know it's the right one for the effect?
  - Typical supporting evidence
    - Experience
    - Interface provides a visual item (such as prompt) to connect action to result effect
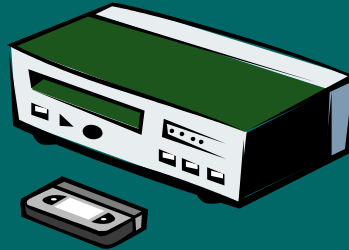   - All other actions look wrong

Next Question

- 4.Will user understand the feedback?
  - Typical supporting evidence
    - Experience
    - Recognize a connection between a system response and what user was trying to do

Example

- Program VCR
  - List actions
  - Ask questions

IRB

- Need to move ahead for project now

- Prepare human subjects submission by next Tuesday
  - Sample consent forms available
  - Do best job with survey instruments
  - Must be forwarded to me
  - Can be amended later

Administratia

- Missing survey forms

Upcoming

- Requirements gathering & Understanding users
  - Contextual inquiry
  - Ethnography
- Task Analysis & User requirements