Methodology Overview

Why do we evaluate in HCI?
Why should we use different methods?
How can we compare methods?
What methods are there?

see www.cpsc.ucalgary.ca/~saul/681/

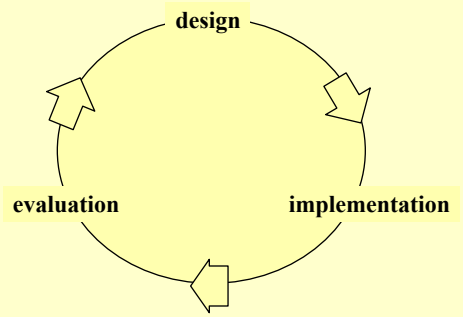
Why Do We Evaluate In HCI?

1. Evaluation to produce generalized knowledge
   - are there general design principles?
   - are there theories of human behaviour?
     - explanatory
     - predictive
   - can we validate ideas / visions / hypotheses?

evaluation produces:
- validated theories, principles and guidelines
- evidence supporting/rejecting hypotheses / ideas / visions…

Why Do We Evaluate In HCI?

2. Evaluation as part of the Design Process



Why Do We Evaluate In HCI?

A. Pre-design stage:

- what do people do?
- what is their real world context and constraints?
- how do they think about their task?
- how can we understand what we need in system functionality?
- can we validate our requirements analysis?

evaluation produces

- key tasks and required functionality
- key contextual factors
- descriptions of work practices
- organizational practices
- useful key requirements
- user type…

Why Do We Evaluate In HCI?

B. Initial design stage:

- evaluate choices of initial design ideas and representations
- usually sketches, brainstorming exercises, paper prototypes
  - is the representation appropriate?
  - does it reflect how people think of their task

evaluation produces:

- user reaction to design
- validation / invalidation of ideas
- list of conceptual problem areas (conceptual bugs)
- new design ideas

Why Do We Evaluate In HCI?

C. Iterative design stage

- iteratively refine / fine tune the chosen design / representation
- evolve low / medium / high fidelity prototypes and products
- look for usability bugs
  - can people use this system?

evaluation produces:

- user reaction to design
- validation and list of problem areas (bugs)
- variations in design ideas

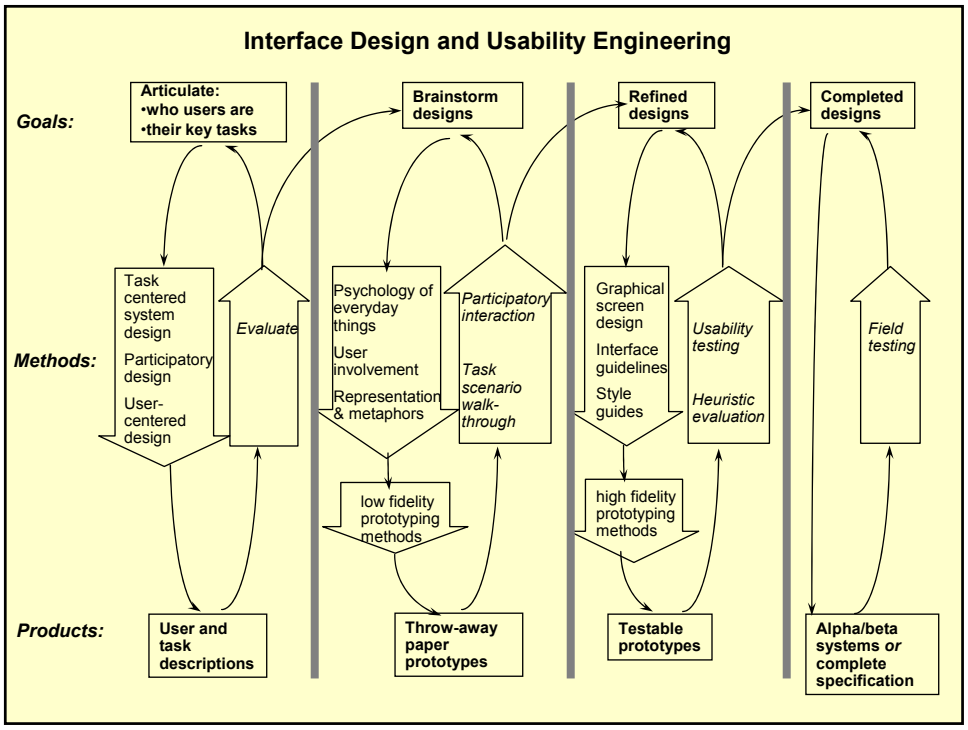
Why Do We Evaluate In HCI?

D. Post-design stage

- *acceptance test:* did we deliver what we said we would?
  - verify human/computer system meets expected performance criteria
  - ease of learning, usability, user's attitude, time, errors…
    - e.g., 9/10 first-time users will successfully download pictures from their camera within 3 minutes, and delete unwanted ones in an additional 3 minutes
- *revisions:* what do we need to change?
- *effects:* what did we change in the way people do their tasks?
- *in the field*: do actual users perform as we expected them to?

evaluation produces

- testable usability metrics
- end user reactions
- validation and list of problem areas (bugs)
- changes in original work practices/requirements



Why Do We Evaluate In HCI?

Design and evaluation

- Best if they are done **together**
  - evaluation suggests design
  - design suggests evaluation
  - use evaluation to create as well as critique
- Design and evaluation methods **must fit** development constraints
  - budget, resources, time, product cost…
  - do triage: what is most important given the constraints?
- Design usually needs quick approximate answers
  - precise results rarely needed
  - close enough, good enough, informed guesses,…
- See optional reading by Don Norman
  - Applying the Behavioural, Cognitive and Social Sciences to Products.

Why Use Different Methods?

Method definition (Baecker, McGrath)

- Formalized procedures / tools that guide and structure the process of gathering and analyzing information

Different methods can do different things.

- Each method offers potential opportunities not available by other means,
- Each method has inherent limitations…

Why Use Different Methods?

All methods:

- enable but also limit what can be gathered and analyzed
- are valuable in certain situations, but weak in others
- have inherent weaknesses and limitations
- can be used to complement each other's strengths and weaknesses.

-McGrath (Methodology Matters)

Why Use Different Methods?

Information requirements differ

- pre-design, iterative design, post-design, generalizable knowledge…

Information produced differs

- outputs should match the particular problem/needs

Relevance

- does the method provide information to our question / problem?
- its not what method is best,
  its what method is best to answer the question you are asking

How Can We Compare Methods?

Naturalistic

- is the method applied in an ecologically valid situation?
  - observations reflect real world settings
    - real environment, real tasks, real people, real motivation

Repeatability

- would the same results be achieved if the test were repeated?

Validity

- External validity:
  - can the results be applied to other situations?
  - are they generalizable?
- Internal validity:
  - do we have confidence in our explanation?

How Can We Compare Methods?

Product relevance

- Does the test measure something relevant to the usability and usefulness of real products in real use outside of lab?

- Some typical **reliability problems** of testing vs real use
  - non-typical users tested
  - tasks are not typical tasks
  - tests usability vs usefulness
  - physical environment different
    - quiet lab vs very noisy open offices vs interruptions
  - social influences different
    - motivation towards experimenter vs motivation towards boss

How Can We Compare Methods?

Partial Solution for product relevance

- use real users
- user real tasks (task-centered system design)
- environment similar to real situation
- context similar to real situation

Why Use Different Methods?

Cost/benefit of using method

- cost of method should match the benefit gained from the result

Constraints and pragmatics

- may force you to chose quick and dirty discount usability methods

How Can We Compare Methods?

Quickness
- can I do a good job with this method within my time constraints?

Cost
- Is the cost of using this method reasonable for my question?

Equipment
- What special equipment / resources required?

Personnel, training and expertise
- What people / expertise are required to run this method?

How Can We Compare Methods?

Subject selection
- how many do I need, who are they, and can I get them?

Scope of subjects
- is it good for analyzing individuals? small groups? organizations?

Type of information (qualitative vs quantitative)
- is the information quantitative and amenable to statistical analysis?

Comparative
- can I use it to compare different things?

How Can We Compare Methods?

Control
- can I control for certain factors to see what effects they have?

Cross-sectional or Longitudinal
- can it reveal changes over time?

Setting
- field vs laboratory?

Support
- are there tools for supporting the method and analyzing the data?

How Can We Compare Methods?

Routine application
- is there a fairly standard way to apply the method to many situations

Theoretic
- is there a theoretic basis behind the method?

Result type
- does it produce a description or explanation?

Metrics
- are there useful, observable phenomena that can be measured

How Can We Compare Methods?

Measures
- can I see processes or outcomes

Organizational
- can they be included within an organization as part of a software development process

Politics
- are there 'method religion wars' that may bias method selection?

What methods are there?

Laboratory tests

requires human subjects that act as end users

- Experimental methodologies
  - highly controlled observations and measurements to answer very specific questions i.e., hypothesis testing
- Usability testing
  - mostly qualitative, less controlled observations of users performing tasks

What methods are there?

Interface inspection

done by interface professionals, no end users necessary

- Usability heuristics
  - several experts analyze an interface against a handful of principles
- Walkthroughs
  - experts and others analyze an interface by considering what a user would have to do a step at a time while performing their task

What methods are there?

Field studies

requires established end users in their work context

- Ethnography
  - field worker immerses themselves in a culture to understand what that culture is doing
- Contextual inquiry
  - interview methodology that gains knowledge of what people do in their real-world context

What methods are there?

Self reporting

requires established or potential end users

- interviews
- questionnaires
- surveys

What methods are there?

Cognitive modeling

requires detailed interface specifications

- Fitt's Law
  - mathematical expression that can predict a user's time to select a target
- Keystroke-level model
  - low-level description of what users would have to do to perform a task that can be used to predict how long it would take them to do it
- Goms
  - structured, multi-level description of what users would have to do to perform a task that can also be used to predict time

Goals of Behavioural Evaluation

Designer:
- user-centered iterative design

Customer
- selecting among systems

Manager
- assisting effectiveness

Marketer
- building a case for the product

Researcher
- developing a knowledge base

(From Finholt & Olsons CSCW 96 Tutorial)

Course goal

To provide you with a toolbox of evaluation methodologies for both research and practice in Human Computer Interaction

To achieve this, you will:
- investigate, compare and contrast many existing methodologies
- understand how each methodology fits particular interface design and evaluation situation
- practice several of these methodologies on simple problems
- gain first-hand experience with a particular methodology by designing, running, and interpreting a study.