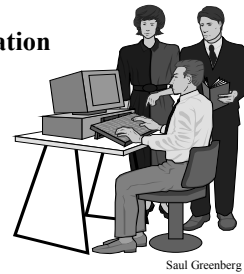# Evaluating Interfaces with Users

Why evaluation is crucial to interface design
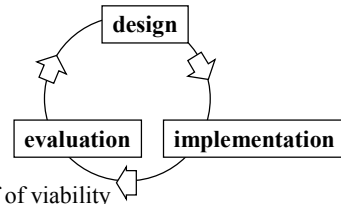
General approaches and tradeoffs in evaluation

The role of ethics

Saul Greenberg

Why Bother?

Tied to the usability engineering lifecycle

design

evaluation   implementation

- Pre-design
  - investing in new expensive system requires proof of viability

- Initial design stages
  - develop and evaluate initial design ideas with the user

- Iterative design
  - does system behaviour match the user's task requirements?
  - are there specific problems with the design?
  - can users provide feedback to modify design?

- Acceptance testing
  - verify that human/computer system meets expected performance criteria
    - ease of learning, usability, user's attitude, performance criteria
    - e.g., a first time user will take 1-3 minutes to learn how to withdraw $50. from the automatic teller
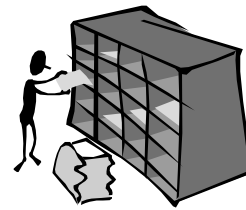
Saul Greenberg

Approaches: Naturalistic

Naturalistic:
- describes an ongoing process as it evolves over time
- observation occurs in realistic setting
  - ecologically valid
- "real life"

External validity
- degree to which research results applies to real situations

Approaches: Experimental

Experimental
- study relations by manipulating one or more *independent* variables
  - experimenter controls all environmental factors
- observe effect on one or more *dependent* variables

Internal validity
- confidence that we have in our explanation of experimental results

Trade-off: Natural *vs* Experimental
- precision and direct control over experimental design *versus*
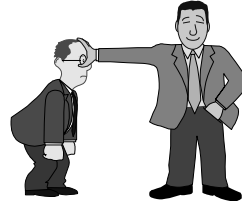- desire for maximum generalizability in real life situations

Approaches: Reliability Concerns

Would the same results be achieved if the test were repeated?

Problem: individual differences:
 • best user 10x faster than slowest
 • best 25% of users ~2x faster than slowest 25%

Partial Solution
 • reasonable number and range of users tested
 • statistics provide confidence intervals of test results
   - 95% confident that mean time to perform task X is 4.5+/-0.2 minutes
       means
     95% chance true mean is between 4.3 and 4.7, 5% chance its outside that

Approaches: Validity Concerns

Does the test measure something of relevance to usability 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
   - physical environment different
       quiet lab vs very noisy open offices vs interruptions
   - social influences different
       motivation towards experimenter vs motivation towards boss

Partial Solution
 • use real users
 • tasks from task-centered system design
 • environment similar to real situation

Ethics

> ...and to think that you want <u>me</u> to test it!!!
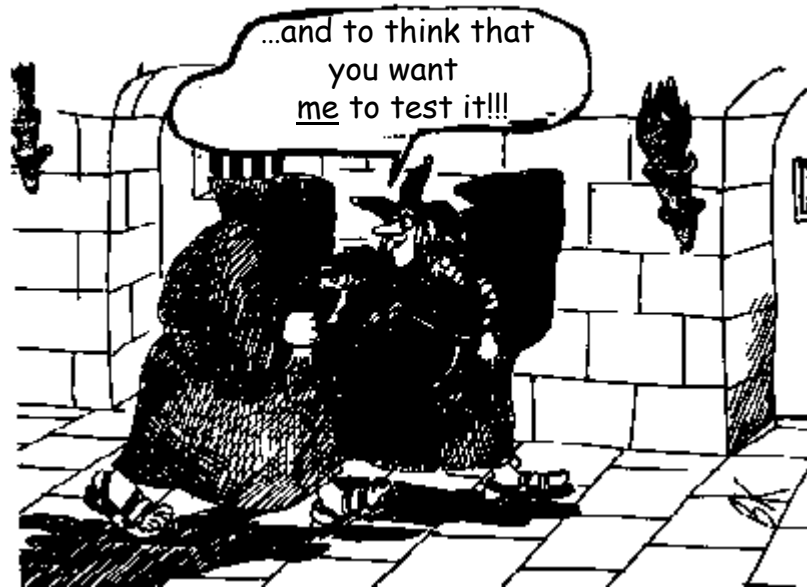
Saul Greenberg
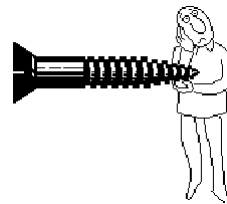
Ethics

Testing can be a distressing experience
- pressure to perform, errors inevitable
- feelings of inadequacy
- competition with other subjects

Golden rule
- subjects should always be treated with respect

Saul Greenberg

Managing subjects in an ethical manner

Before the test

- don't waste the user's time
  - use pilot tests to debug experiments, questionnaires etc
  - have everything ready before the user shows up

- make users feel comfortable
  - emphasize that it is the system that is being tested, not the user
  - acknowledge that the software may have problems
  - let users know they can stop at any time

- maintain privacy
  - tell user that individual test results will be kept completely confidential

- inform the user
  - explain any monitoring that is being used
  - answer all user's questions (but avoid bias)

- only use volunteers
  - user must sign an informed consent form

Managing subjects in an ethical manner

During the test

- don't waste the user's time
  - never have the user perform unnecessary tasks

- make users comfortable
  - try to give user an early success experience
  - keep a relaxed atmosphere in the room
  - coffee, breaks, etc
  - hand out test tasks one at a time
  - never indicate displeasure with the user's performance
  - avoid disruptions
  - stop the test if it becomes too unpleasant

- maintain privacy
  - do not allow the user's management to observe the test

Managing subjects in an ethical manner

After the test
- make the users feel comfortable
  - state that the user has helped you find areas of improvement

- inform the user
  - answer particular questions about the experiment that could have biased the results before

- maintain privacy
  - never report results in a way that individual users can be identified
  - only show videotapes outside the research group with the user's permission

You know now

Evaluation is crucial for designing, debugging, and verifying interfaces

There is a tradeoff in naturalistic *vs* experimental approaches
- internal and external validity
- reliability
- precision
- generalizability

Subjects *must* be treated with respect
- ethical rules of behaviour