Evaluating Interfaces With Users

Why evaluation is crucial to interface design

General approaches and tradeoffs in evaluation

The role of ethics
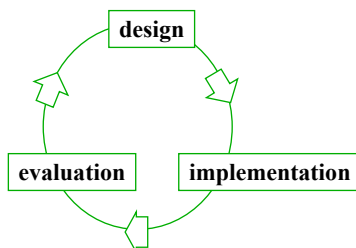

Why Bother?

Tied to all parts of the usability engineering lifecycle

• Pre-design
  - Investing in new expensive system requires proof of viability

• Initial design stages
  - Develop and evaluate initial design ideas with the user

```
              design

  evaluation      implementation
```

Why Bother?

Iterative design
- Does the system match the user's task requirements?
- Are there any specific problems with the design?
- Can users provide feedback to modify the design

Acceptance testing
- Verify that the system meets expected performance criteria: ease of learning, usability, user's attitude
- E.g., A first time user will take between one and three minutes to learn how to withdraw $50 from the ATM.

```
        design
           ↓
evaluation → implementation
```
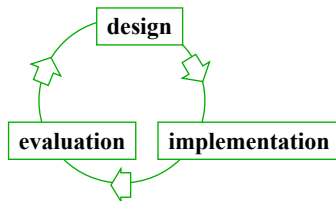
Approaches: Naturalistic

Observation occurs in realistic setting
- Real life

- Problems
  - It may be difficult to arrange and to conduct
  - Results may not be replicated

Approaches: Experimental

Experimental

- Classical lab study
- Study relations by manipulating one or more *independent* variables
  - Experimenter controls all environmental factors (nothing else changes)
- Observe effect on one or more *dependent* variables

Tradeoffs: Natural Vs. Experimental

Internal validity

- Do you measure what you set out to measure (correctness)

External validity

- The degree to which results can be generalized to other situations (realism)

	Naturalistic	Experimental
Internal validity	Low	High
External validity	High	Low

(External) 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
- Use tasks derived from a task-centered approach to system design
- Test in an environment similar to real situation

How Many Participants To Test

Would the same results be achieved if the test were repeated?
- Reliability of data

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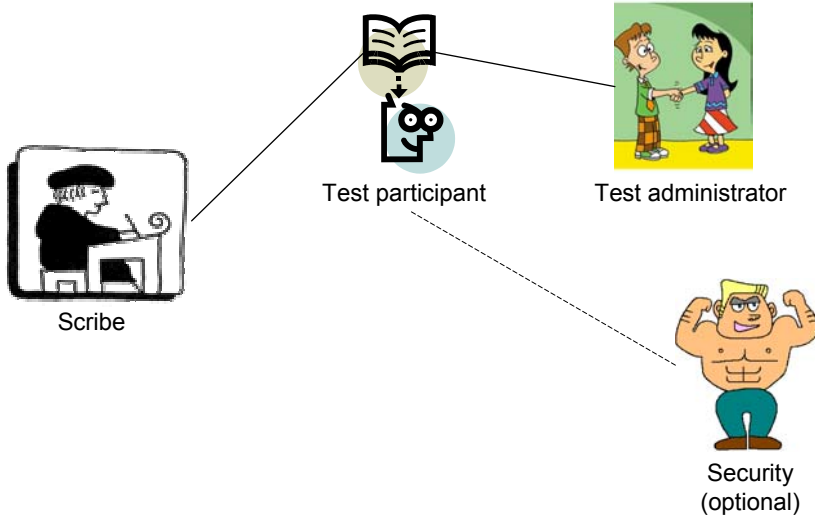- The best 25% of users ~2x faster than slowest 25%

Partial Solution
- Get a reasonable number and range of test participants

Test Roles

Scribe

Test participant

Test administrator

Security
(optional)

Test Procedure

I) Run a pilot study

- "A practice run" of the test
- Purpose: To debug the test
- Results: Used to improve the test

II) Run the main test

- Running the test "for real"
- Purpose: To debug the interface
- Results: Used to improve the interface

Test Procedure (2)

1. **Preparation**

2. **Introduction**

3. **Running the system**

4. **Debriefing**

Ethics

Ethics

Testing can be a distressing experience
- People feel pressure to perform so errors are inevitable
- This can result in:
  - Feelings of inadequacy
  - Competition with other test participants

Golden rule
- Test participants should always be treated with respect

Managing Participants In An Ethical Manner

Before the test
- Don't waste the person's time
  - Use pilot tests to debug experiments, questionnaires etc
  - Have everything ready before the participant shows up

- Make participants feel comfortable
  - Emphasize that it is the system that is being tested, not the person
  - Acknowledge that the software may have problems
  - Let participants know they can stop at any time

- Maintain privacy
  - Tell the participant that individual test results will be kept completely confidential

- Inform the participant
  - Explain any monitoring that is being used
  - Answer all of the person's questions (but avoid bias)

- Only use volunteers
  - Typically the test participant must sign an informed consent form

Managing Participants In An Ethical Manner

During the test

- Don't waste the person's time
  - Never have the user perform unnecessary tasks

- Make test participants comfortable
  - Try to give the person an early success experience
  - Keep a relaxed atmosphere in the room
  - Have coffee, breaks, etc
  - Hand out test tasks one at a time
  - Never indicate displeasure with the person's performance
  - Avoid disruptions
  - Stop the test if it becomes too unpleasant

- Maintain privacy
  - Do not allow the participant's management to observe the test

---

Managing Participants In An Ethical Manner

After the test
- Make the person feel comfortable
  - E.g.,, state that the participant has helped you find areas of improvement

- Inform the participant
  - Answer particular questions about the experiment that could have biased the results before

- Maintain privacy
  - Never report results in a way that individuals can be identified
  - Only show videotapes outside the research group with the participant'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

The number and range of test participants employed will effect the reliability of your results

Test participants *must* be treated with respect
 • The study should be guided by ethical rules of behaviour