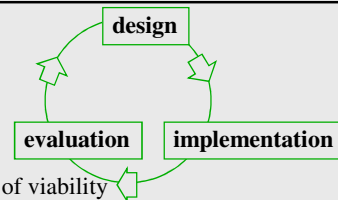


## Evaluating Interfaces with Users

### Why Bother?

#### Tied to the usability engineering lifecycle

- Pre-design
  - investing in new expensive systems 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



## What Defines Success?

**We want a “usable” system. What are some metrics that can be used to measure whether a system is usable?**

- Time to learn
- Speed of performance
- Rate of errors by users
- Retention over time
- Subjective Satisfaction

**Often, there will be tradeoffs between these goals.**

Evan Golub / Ben Bederson / Saul Greenberg

## Approaches: Naturalistic/Qualitative

### **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



Evan Golub / Ben Bederson / Saul Greenberg

## Approaches: Experimental/Quantitative

### **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

Evan Golub / Ben Bederson / Saul Greenberg

## 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

Evan Golub / Ben Bederson / Saul Greenberg

## Validity Concerns

### **Does the test measure something of relevance to usability of real products in real use outside of lab?**

- Some typical validity 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

Evan Golub / Ben Bederson / Saul Greenberg