

Evaluation

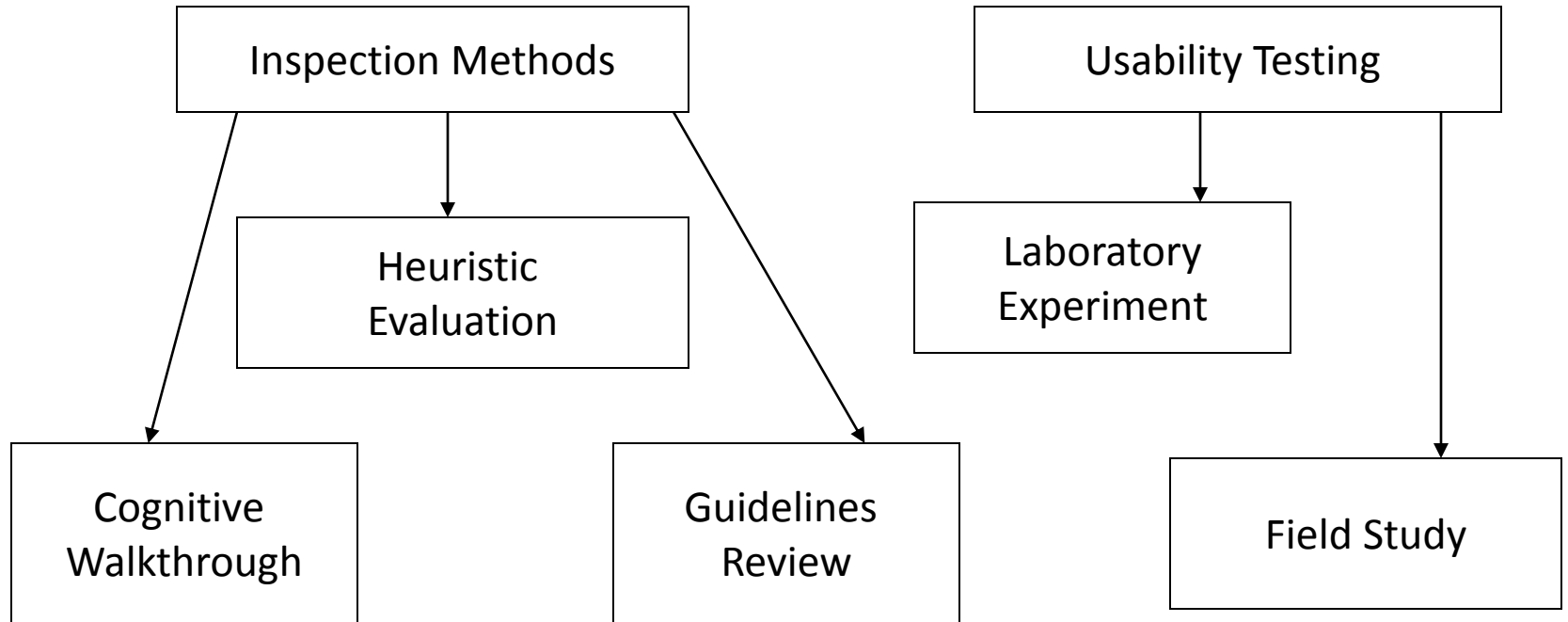
Why Evaluate a User Interface

- Assess effect of interface on user performance and satisfaction
- Identify specific usability problems
- Evaluate users' access to functionality of system
- Compare alternative systems / designs

When Evaluate a User Interface

- Evaluation should be done throughout the usability life cycle – not just at the end - "iterative design"
- Different evaluation methods appropriate at different stages of the cycle

How Evaluate a User Interface



Formative or Summative Evaluation

- Formative Evaluation: Identify usability problems
 - Qualitative measures
 - Ethnographic methods
- Summative evaluation: Measure / compare user performance
 - Quantitative measures
 - Statistical methods

Formative Evaluation: Collecting Usability Problems

- Think aloud protocol
 - User observed performing task
 - User asked to describe what he is doing and why, what he thinks is happening etc.
 - Identify what the user did and expressed: goals, confusions or misunderstandings, errors, reactions expressed
- Advantages
 - Simplicity - requires little expertise
 - Can provide useful insight
 - Can show how system is actually used
- Disadvantages
 - Subjective
 - Difficult to conduct
 - Act of describing may alter task performance

Cognitive Walkthrough

- Evaluates design on how well it supports user in learning task
- Usually performed by expert in cognitive psychology
- Expert "walks through" design to identify potential problems using psychological principles
- Scenarios may be used to guide analysis

Heuristic Evaluation

- Usability criteria (heuristics) are identified
- Design examined by experts to see if these are violated
- Example heuristics
 - System behaviour is consistent
 - Feedback is provided
- Heuristic evaluation "debugs" design.

Guidelines Inspection

- Written guidelines recommended for larger projects:
 - Screen layout
 - Appearance of objects
 - Terminology
 - Wording of prompts and error messages
 - Menus
 - Direct manipulation actions and feedback
 - On-line help and other documentation

Designing a Usability Experiment

- Usability testing in a controlled environment
 - There is a test set of users
 - They perform pre-specified tasks
 - Data is collected (quantitative and qualitative)
 - Take mean and/or median value of measured attributes
 - Compare to goal or another system
- Contrast with “expert review” evaluation methodology

Experimental Design: Subjects

- Subjects
 - Representative
 - Sufficient sample

Experimental Design: Variables

- Independent variable (IV)
 - Characteristic changed to produce different conditions.
e.g. interface style, number of menu items.
- Dependent variable (DV)
 - Characteristics measured in the experiment
e.g. time to perform task, number of errors.

Experimental Design: Groups

- Within groups design
 - Each subject performs experiment under each condition.
 - Transfer of learning possible
 - Fewer subjects needed
 - Less likely to suffer from user variation
- Between groups design
 - Each subject performs under only one condition
 - No transfer of learning
 - More subjects required
 - User variation can bias results

Experimental Design: Factors

- Total task time
- User “think time”
- Time spent not moving toward goal
- Ratio of successful actions:errors
- Commands used/not used
- Frequency of user expression of confusion, frustration, satisfaction
- Frequency of reference to *help*
- Percentage of time *help* provided the needed answer

Experimental Design: Measuring User Performance

- Measuring learnability
 - Time to complete a set of tasks by novice
 - Learnability/efficiency trade-off
- Measuring efficiency
 - Time to complete a set of tasks by expert
 - How to define and locate “experienced” users

Experimental Design: Measuring User Performance

- Measuring user satisfaction
 - Likert scale (agree or disagree)
- Measuring errors
 - Classification of minor v. serious

Experimental Design: Documenting the Test Protocol

- What tasks?
- Criteria for completion?
- What will users be asked to do?
- Interaction with experimenter?
- What data will be collected?

Experimental Design: Identifying Tasks

- Cover most important parts of UI
- Don't take too long to complete
- Goal or result oriented (possibly with scenario)
- Not frivolous or humorous (unless part of product goal)
- First task should build confidence
- Last task should create a sense of accomplishment

Experimental Design: Identifying Tasks

- Cover most important parts of UI
- Don't take too long to complete
- Goal or result oriented (possibly with scenario)
- Not frivolous or humorous (unless part of product goal)
- First task should build confidence
- Last task should create a sense of accomplishment

Experimental Design: Ethics

- Re-assure that individual results not revealed
- Re-assure that user can stop any time
- Provide comfortable environment
- Don't laugh or refer to users as subjects or guinea pigs
- Don't volunteer help, but don't allow user to struggle too long
- Provide de-briefing
 - Answer all questions
 - Thank for helping

Field Studies: Approaches

- Direct observation in actual use
 - discover new uses
 - take notes, don't help, chat later
- Logging actual use
 - objective, not intrusive
 - great for identifying errors
 - which features are/are not used
 - privacy concerns

Field Studies: Approaches

- Questionnaires and interviews with real users
 - ask users to recall critical incidents
 - questionnaires must be short and easy to return
- Focus groups
 - 6-9 users
 - skilled moderator with pre-planned script
 - Video conferencing

Field Studies: Approaches

- On-line direct feedback mechanisms
 - initiated by users
 - may signal change in user needs
 - trust but verify
- Bulletin boards and user groups

Field Studies: Pro and Con

- Advantages
 - natural environment
 - context retained (though observation may alter it)
 - longitudinal studies possible
- Disadvantages
 - distractions
 - noise
- Appropriate
 - where context is crucial for longitudinal studies