


The Basics of Conducting a Usability Evaluation



March 26, 2004

Melanie Wright, Ph.D.

Human Factors Engineer, Human Simulation and Patient Safety Center

Assistant Professor, Department of Anesthesiology

Duke University School of Medicine

melanie.wright@duke.edu

<http://simcenter.duhs.duke.edu>

Objectives

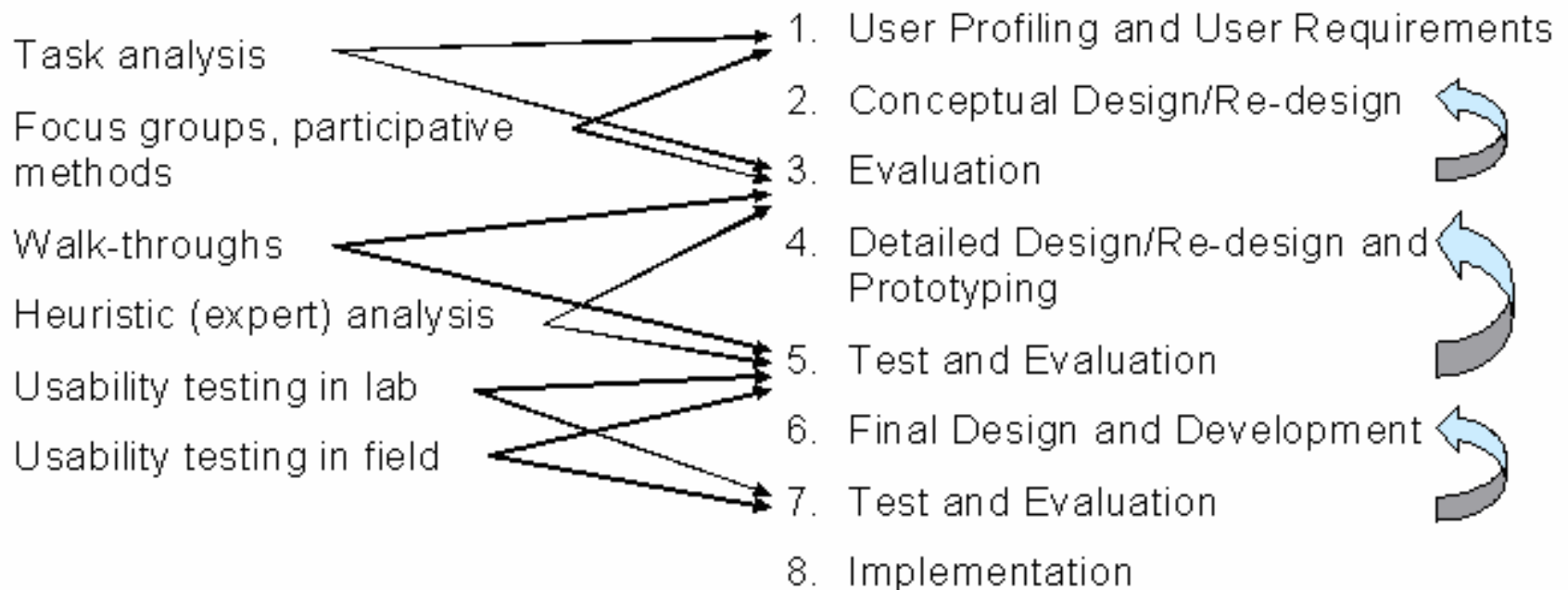
- ❑ Discuss the importance of user studies/usability evaluations within a product design process
- ❑ Present different types of usability evaluations and their purposes
- ❑ Cover in some detail the steps for conducting basic usability test



Human Centered Design Process

Human Factors Methods*

Human Centered Design Process



***Sample** of HF Methods within an HCD process

Types of Usability Evaluations

- Usability test
 - Objective to identify potential usability problems
 - Usually involves a single product
 - Examples: RCA DVD Player, Microsoft Word
- Competitive analysis
 - Objective is to compare two or more products
 - Can be an earlier version of the same product or a competitor
 - Examples: Quicken vs. Money, Ericsson vs. Nokia
- Feature test
 - Compare specific features within a product to make a design decision
 - Example: Text input, Voice control feedback
- Other
 - Other types of comparisons, e.g., overhead console display vs. radio
 - Information gathering, e.g., menu tests



Overview of a Basic Usability Test

- Test Plan
- Product or prototype
- Experimental design
- Test Participants
- Scenarios
- Test setup
- Test procedures
- Measures
- Analysis

Test Plan

- Document what you will test and how
- Include major goals of the test
- Can be informal/internal or formal to share with others
- Include details:
 - Prototype/product fidelity, components and setup
 - Test participant criteria and number
 - Test scenarios
 - Measures to be used
 - Usability goals
 - Test procedures
 - Analysis methods

Product/Prototype

- ❑ Prototype must be complete/bug-free as appropriate for the goals of the test
- ❑ In competitive tests, products should be at a similar level of development
- ❑ Pilot test the prototype by trying out the test scenarios
- ❑ Consider all parts of the system



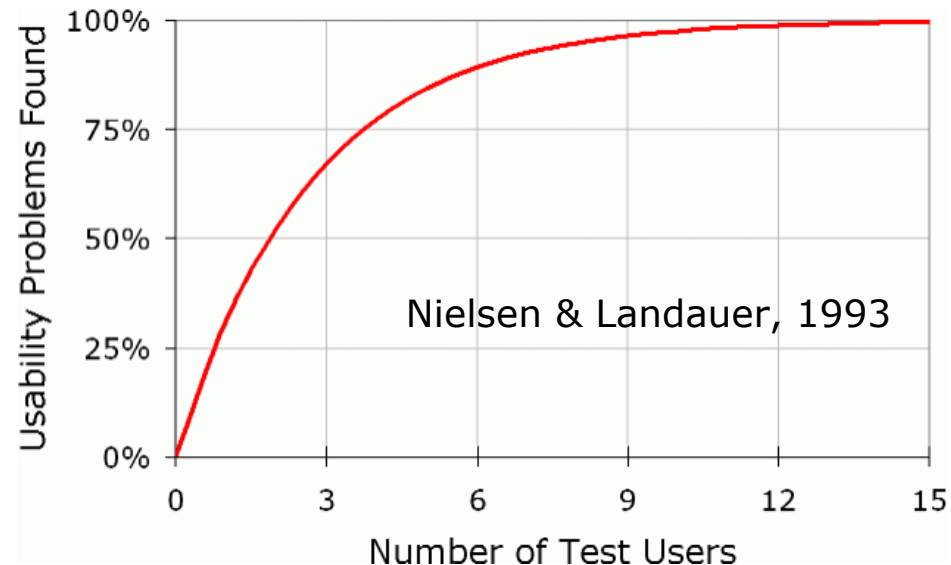
Experimental Design

- Basic usability test of a single product usually an informal (no statistical comparison) design
 - May want to establish some consistent test design and procedures for comparison with earlier or later studies
 - Usability goals can be used as a basis for comparison
- Competitive or comparative studies
 - Within- or between-subjects design depends on potential transfer effects
 - Balance order of presentation in within-subject designs
 - Randomize product presentation for between designs
 - Balance product presentation across different “classes” of users
 - Number of participants depends on experimental design/power analysis



Test Participants

- Participants appropriate to the level of development
 - You may choose to use “internal” individuals for early testing
 - Testing at later stages of development should include “representative” users
 - Create specific criteria (e.g., computer skills) for participants based on user population and screen for those criteria
 - Consider all user types (may require different tests)
- Number of participants depends on purpose of test
 - Early testing for problems with as few as 5
 - Standard usability test ~ 8
 - Competitive/comparison tests may require more if seeking statistically significant differences

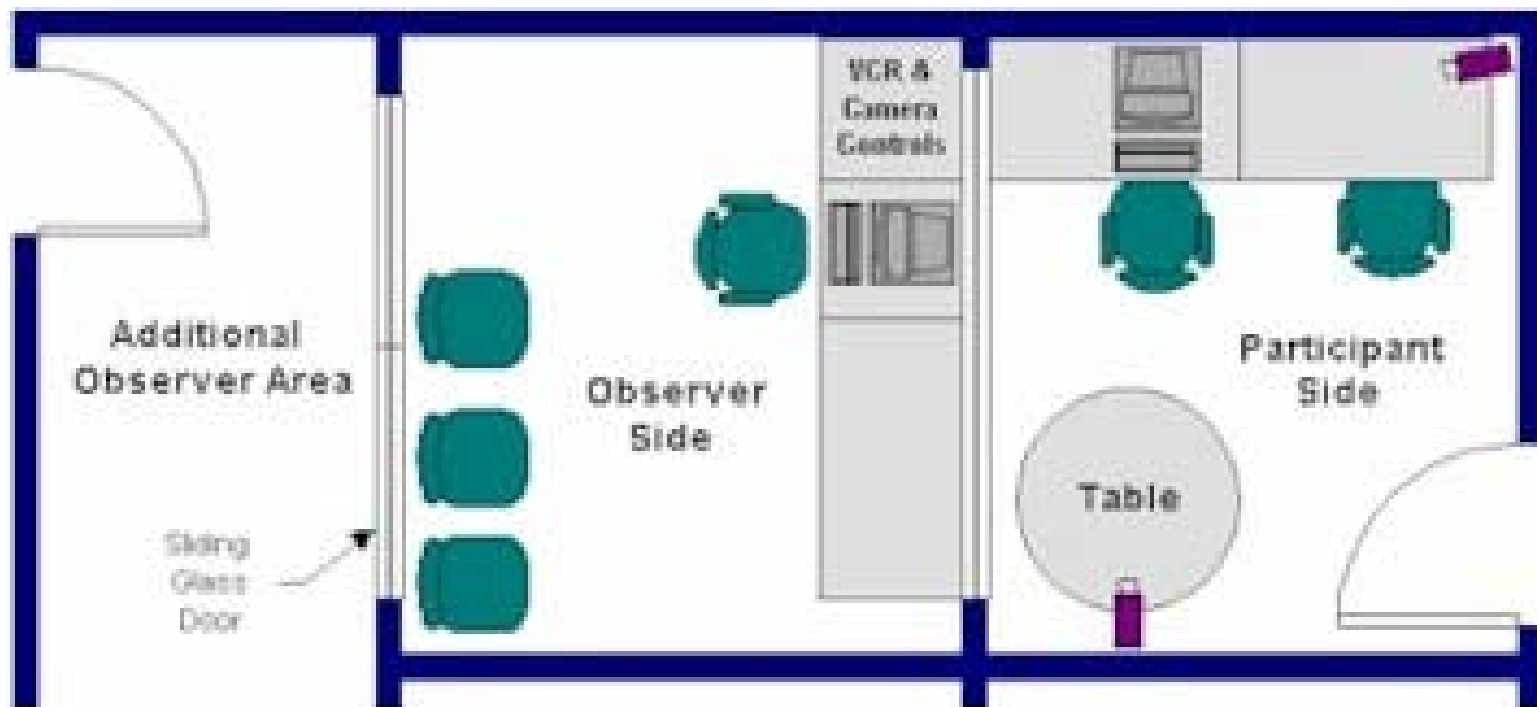


Scenarios

- ❑ Scenarios are essentially end user “use cases” for testing the product
- ❑ Consider the major tasks and sub-tasks
- ❑ “Core” tasks may be selected that can be used across tests for comparisons
- ❑ Select a sample of tasks that is representative of key user interface features you wish to test
- ❑ Write the tasks into realistic scenarios and include any necessary details to allow the participants to complete the tasks
- ❑ Number of tasks to be tested will affect the length of the test

Lab Setup

- Fidelity of test setup based on goals of the test
 - Laboratory environment provides a level of control
 - Laboratory setup can be informal/formal
 - Field environment provides higher fidelity but less control
 - Testing impact of product use on other tasks (e.g., driving) may require unique laboratory setup or field environments
- Common laboratory setup
 - Separate test room
 - Observation through (one-way) glass
 - Communication via intercom
 - Video and audio-taping to include taping of test participant and taping of product interaction



Test Procedures

- Typical test
 - Introduction, background questionnaires, consent forms
 - Training session
 - Review of testing instructions
 - Test scenarios/task (end of task questionnaires)
 - Breaks
 - End of test questionnaire
 - Debrief
- Consider limits of attention in test time
 - 4 hours or less works well
 - Full day and multi-day tests require planning and breaks

Measures

- Performance measures
 - E.g., Time on task, task success/failure, errors, assistance
 - Define specific criteria for performance measures
- Secondary task performance
- Subjective ratings
 - Likert scales for rating satisfaction (e.g., ease of use, task time, help if used, usefulness of function, etc.)
 - End of test ratings of details including hardware qualities and software qualities
 - Ranking is useful in competitive tests
 - Workload ratings (NASA-TLX)
- Notes – error/assistance details



Data Collection Do's and Don'ts


- ❑ DO have participant interact with one test administrator throughout test
- ❑ DO have patience
- ❑ DO repeat instructions as necessary
- ❑ DON'T coach or instruct participants during tasks
- ❑ DO allow participants to communicate difficulties and to achieve some level of success with each task
- ❑ DON'T allow observers to disturb the data collection process
- ❑ DO know the interface very well prior to testing
- ❑ Remember that you are there to observe – there is no real “right” or “wrong”



Analysis & Findings

- Basic usability testing often just use simple descriptive statistics
 - E.g., number pass/fail for each scenario, average time on task, average number of errors/assistance per scenario, average ratings or number of participants satisfied/dissatisfied
 - Use these statistics plus notes to identify problem areas
 - Recommendations naturally fall out from the testing experience
- Statistical comparisons for competitive tests
 - Simple statistics such as t-tests
 - Non-parametric statistics for small sample sizes
- Key part of findings is recommendations for design changes

The Basics of Conducting a Usability Evaluation



March 26, 2003

Melanie Wright, Ph.D.

Human Factors Engineer, Human Simulation and Patient Safety Center

Assistant Professor, Department of Anesthesiology

Duke University School of Medicine

melanie.wright@duke.edu

<http://simcenter.duhs.duke.edu>