

Usability Evaluation with Users

Lecture 9
Chapter 7 Rosson & Carroll

Testing and Evaluation Summary

- Purpose: Evaluation for usability
- Methods
 - Without Users (analytic)
 - Claims analysis
 - Usability inspection
 - Model-Based analysis
 - With Users (empirical)
 - Experiments
 - Field Studies
 - Surveys
 - Usability testing

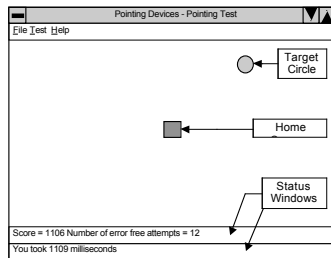
Controlled Experiments

- Hypothesis testing
 - Does a cause b to change?
 a is independent factor manipulated by the experiment
 b is dependent factor measured by experiment
- Example
 - Which device is faster to point with: A mouse or a finger-controlled joystick? Does having the joystick on the keyboard save “homing” time?
 - What is a ? What is b ?

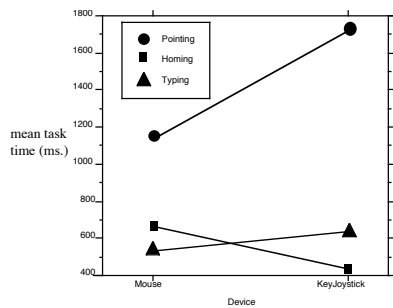
Pointing Device Experiment

- Hypotheses
 - “The key joystick is faster to home.”
 - “The key joystick is faster to point.”
 - “The key joystick is the same at typing.”
- Independent variables
 - Device: mouse vs. key joystick
 - Task: pointing, typing, homing
- Dependent variables (tested by ANOVA)
 - Time
- 12 participants on each device do all tasks

Fitts Pointing Task on the Computer



Task Times for Practiced Users



Task Times for Practiced Users

Device	Mean Trial Time (ms) for Mode Switching Subtasks		
	Pointing	Homing	Typing
Mouse	1158	667	531
Key Joy	1746	438	639
KJ:Mouse	1.51*	0.66*	not significant

* is significant at $p \leq 0.001$

Field Studies

- Observational
 - Use of software in *context* of work or play
 - Method used for studying use over a long period of time (longitudinal)
 - Uses ethnographic methods from anthropology

Surveys

- Data
 - what users say, not what they do!
 - correlational, not causal
- Methods
 - Questionnaires
 - Interviews
 - Caution about focus groups: a marketing method not a usability study
- Can be used in the “small” or “large”
 - Small: Qualitative
 - Large: Sample size representative of population

Usability Testing

- Definition
 - Usability testing is an empirical method which puts typical users in a laboratory, gives them a prototype and a set of tasks, and records their interactions, usually on videotape.

Testing Goals vs. Method

	Guidelines	Walkthrough	Expert	Keystroke Level Model	Usability Testing
Overall Usability					
Detailed Usability					
Completeness					
Correctness					
Consistency					
Performance Time					

Usability Testing

- What can you get from usability testing?
 - Testing against usability requirements: Does the system meet the usability design goals? Usability testing allows measurement of performance time.
 - Design improvements: What changes should be made to the system?
 - Conceptual problems: What misconceptions exist?
 - Repair strategies: What did the person do to recover from failure?
 - Problem solving strategies: What strategies did the person use that could promote a more supportive design?

Usability Testing

- Benefits
 - Focus on first-time users
 - Detects most serious problems
 - Uses real tasks and real users
- Problems
 - Only tests learning
 - Finds problems but doesn't always suggest how to fix them!
 - Expensive and time-consuming

How to do it

- Plan
 - Determine goals of usability testing
 - Usability problems
 - Usability specifications
 - Choose pairs of users
 - Real users, not actors or other developers
 - Representative users and note relationships
 - Get informed consent
 - Choose representative tasks
 - Instructions to participants
 - Working prototype

Usability Specifications VSF Example

Scenario and Subtasks	Worst Case	Planned	Best Case
Scenario Context: Alicia and Della visit the Fair	3 on usefulness and ease of use	4 on usefulness and ease of use	5 on usefulness and ease of use
1. Finding a co-present visitor	15 seconds, 1 error 3 on awareness	5 seconds, 0 error 4 on awareness	1 second, 0 error 5 on awareness
2. Navigating to a specified exhibit	20 seconds, 1 error 3 on directness	5 seconds, 1 error 4 on directness	3 seconds, 0 error 5 on directness
3. Positioning an activity rectangle	30 seconds, 3 errors 3 on naturalness	15 seconds, 1 error 4.5 on naturalness	3 seconds, 0 error 5 on naturalness
4. Opening and manipulating Excel chart	2 minutes, 2 errors 3 on intuitiveness	1 minute, 0 error 4 on intuitiveness	30 seconds, 0 error 5 on intuitiveness
5. Chat review and input	2 minutes, 3 errors 3 on comprehension	1 minute, 0 error 4 on comprehension	30 seconds, 0 error 5 on comprehension

KLM predicted times to vote

Team	Estimated Time
Balloteers	40.75 sec
BBJ	58.95 sec
CBX	63.95 sec
CMZ	69.85 sec
LCVB	68.95 sec
OVS-04	49.40 sec
Pirates	84.70 sec
mean	62.36 +/- 14.42

Prototype Implementation

- Software Storyboard
- Software Prototype with GUI Builder or Prototyping language
- Partial Target Software Implementation
- Completed Target Software Implementation

Forms

(see Rosson & Carroll chap 7.4)

- Informed Consent
 - What they are going to do
 - How you will evaluate the data
 - Anonymity
 - Can quit at any time
 - Contact information for person doing study
 - Signature
- User background survey
 - Only ask for what you need!

General Instructions for VSF Study

"In the next 30 minutes or so, you will be carrying out ten tasks within the Virtual Science Fair. These tasks are organized into two groups of four and six. Each group will be introduced with a storyline that describes the role and situation we would like you to adopt for that set of tasks, then each task is specified individually. Note that we intentionally leave out some of the detailed task steps so that we can determine how well the system can guide your interactions with it. If you are confused at any point, please just make your best guess about how to proceed, using the information that you have been given. We will intervene if necessary to help you make progress.

At the start of each task, please say out loud: "Beginning Task" followed by the number of the task. When you are done, please say: "Task Complete". Also, please remember to think out loud as you work. It is very important for us to understand your goals, expectations, and reactions as you work through the tasks. Any further questions?"

Specific Task Instructions for Science Fair Study

Virginia Tech Computer Science Department and Center for Human-Computer Interaction

Background to Tasks 1-4:
Imagine that you are Mr. King, an experienced science teacher who has coached many student projects in the past. This year you are advising Sally Harms on her Black holes project. You are both very busy, but have arranged to meet in the VSF at 8 PM tonight. When you arrive, Sally is already there working.

Task 1:

- Find out what exhibit component Sally is working on and synchronize your view with hers.

Task 2:

- While Sally works on her Title Page, upload the Word file "Bibliography.doc" (on your desktop) into the exhibit element named "Bibliography".

Task 3:

- Open the exhibit element that is an Excel document and modify the title of the chart to be "Interaction of Mass and Movement". Save your change.

Task 4:

- Open the simulation element and add a nested folder named "Visitor Experiments".

User Reactions Survey

Virginia Tech Computer Science Department and Center for Human-Computer Interaction

Now that you have completed the VSF tasks, we would like to know some of your reactions, both in general, and to specific features of the system.

Name: _____

What three things did you like **most** about the VSF? Why?

1. _____

2. _____

3. _____

What three things did you like **least** about the VSF? Why?

1. _____

2. _____

3. _____

If the VSF was made available to you, would you use it or not? Yes No

Why? _____

Please respond to the following 10 items by circling the option that best corresponds to your own.

1. Browsing an online science exhibit is like visiting a science fair in the real world.	<input type="radio"/> Strongly Disagree	<input type="radio"/> Disagree	<input type="radio"/> Neutral	<input type="radio"/> Agree	<input type="radio"/> Strongly Agree
2. Online science exhibits are diverse and of interest to a wide range of visitors.	<input type="radio"/> Strongly Disagree	<input type="radio"/> Disagree	<input type="radio"/> Neutral	<input type="radio"/> Agree	<input type="radio"/> Strongly Agree
3. There are many opportunities for me to become involved in students' science projects.	<input type="radio"/> Strongly Disagree	<input type="radio"/> Disagree	<input type="radio"/> Neutral	<input type="radio"/> Agree	<input type="radio"/> Strongly Agree
4. I am aware of other visitors' locations and activities while in the VSF.	<input type="radio"/> Strongly Disagree	<input type="radio"/> Disagree	<input type="radio"/> Neutral	<input type="radio"/> Agree	<input type="radio"/> Strongly Agree
5. Movement from one location in the VSF to another is awkward.	<input type="radio"/> Strongly Disagree	<input type="radio"/> Disagree	<input type="radio"/> Neutral	<input type="radio"/> Agree	<input type="radio"/> Strongly Agree
6. It is easy and natural to join other visitors at an exhibit.	<input type="radio"/> Strongly Disagree	<input type="radio"/> Disagree	<input type="radio"/> Neutral	<input type="radio"/> Agree	<input type="radio"/> Strongly Agree
7. Experimenting with exhibit components (e.g. spreadsheets) is fun and enlightening.	<input type="radio"/> Strongly Disagree	<input type="radio"/> Disagree	<input type="radio"/> Neutral	<input type="radio"/> Agree	<input type="radio"/> Strongly Agree
8. Chatting with multiple visitors at the same time is confusing.	<input type="radio"/> Strongly Disagree	<input type="radio"/> Disagree	<input type="radio"/> Neutral	<input type="radio"/> Agree	<input type="radio"/> Strongly Agree

<more Likert scales, assessing reactions to other subtasks>

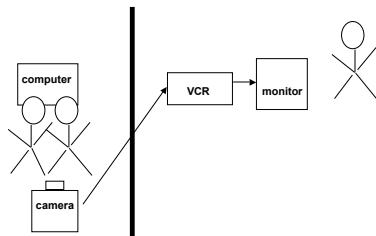
How to do it (cont.)

- Data Collection: Videotaping Method
- Analysis of Data

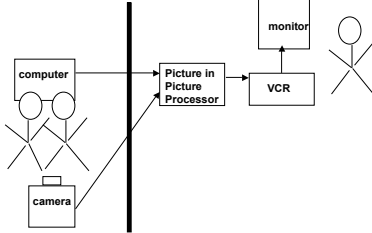
Videotaping

- Number of participants
 - Thinking aloud (one person)
 - Constructive interaction (two or more people)
- Number of cameras
 - One camera
 - Two cameras
- Don't watch what they are doing.
 - Set up camera and leave
 - Put camera on tripod over their heads aimed at display
- Don't answer their questions unless they are very stuck

Usability Testing Single Camera Videotaping Method



Usability Testing Two Camera Videotaping Method



Usability Testing: Analysis

- Review tape, noting where participants have problems
- Try to understand reasons for problems
- Prioritize problems by severity
- Summarize findings in terms of overall interface characteristics
 - Were the usability requirements met?
 - Overall impression of the software

Usability Testing: Analysis Worksheet

- Interface Name, Date, Time, Participants
- Location of problem on tape
- Task attempting to do
- What were the users' stated or assumed goal?
- What were the users' actions?
- What did the users think happened? (perceived effect)
- Priority of the problem
- Analysis and recommendation

Integrating Usability Testing back into the Design

- Usability Testing
 - Overall usability requirements achieved?
 - Identify and prioritize usability problems
 - Explain causes
- Generate design alternatives to solve most important problems
 - Justify advantages/disadvantages/tradeoffs of each solution
- Choose one solution for each problem
- Implement in prototype and test again

Testing & Evaluation: What is the best method? (Jeffries et al., CHI '91)

- Which method discovers the most severe usability problems?

Expert Evaluation	3.59
Walkthrough	3.44
Guidelines	3.61
Usability Testing	4.15
- Range: 1= trivial 9= critical
- mean shown for severity of problems discovered

Testing & Evaluation: What is the best method?

- Number of severe problems found

	Most Severe	Least Severe
Expert Evaluation	28	52
Walkthrough	9	10
Guidelines	12	11
Usability Testing	18	2
- most= top third least= bottom third

Usability Testing in Practice

(Nielsen, *Usability Engineering*, 1993)

Survey of 28 projects from broad variety of companies and industries

Findings:

	<u>Median</u>
Project size in person-years	24
Actual share of budget for usability	6%
Ideal share of budget for usability	10%
Actual usability effort in person-years	1.5
Ideal usability effort in person-years	2.1

Usability Testing in Practice

cont.

- Findings:

- Usability effort is independent of project size!
- Why? Many usability activities take about the same time to perform, no matter how difficult the program is to develop

Testing and Evaluation of Users

Summary

- Methods

- Without Users
 - Guidelines
 - Interface (Cognitive) walkthrough
 - Heuristic evaluation
 - Keystroke Level Model
- With Users
 - Usability testing
 - Questionnaire & Interview

- Use *all* methods for a more usable interface
