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

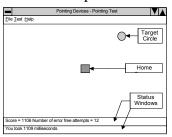
- Hypothesis testing
 - Does a cause b to change? \boldsymbol{a} is independent factor manipulated by the experiment b is dependent factor measured by experiment
- Example
 - Which device is faster to point with: A mouse or a finger-controlled joystick? Does having the joystick on the keyboard save "homing"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

Pointing Device Experiment

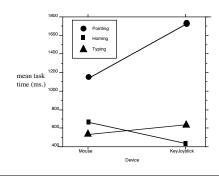
- Hypotheses

 "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	Pointing Homing Typing						
Mouse	1158	667	531					
Key Joy	1746	438	639					
KJ:Mouse	1.51*	0.66*	not significant					

^{*} is significant at p ≤ 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	1		√		· •
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 Delia 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	5 seconds, 0 error	1 second, 0 error
	3 on awareness	4 on awareness	5 on awareness
2. Navigating to a specified exhibit	20 seconds, 1 error	5 seconds, 1 error	3 seconds, 0 error
	3 on directness	4 on directness	5 on directness
3. Positioning an activity rectangle	30 seconds, 3 errors	15 seconds, 1 error	3 seconds, 0 error
	3 on naturalness	4.5 on naturalness	5 on naturalness
4. Opening and manipulating Excel chart	2 minutes, 2 errors	1 minute, 0 error	30 seconds, 0 error
	3 on intuitiveness	4 on intuitiveness	5 on intuitiveness
5. Chat review and input	2 minutes, 2 errors	1 minute, 0 error	30 seconds, 0 error
	3 on comprehension	4 on comprehension	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!

_				
_				
_				
_				
_				
_				

User Sti	udy of A Virtual Science Fair: Consent Form
Virginia Tech Compu	ter Science Department and Center for Human-Computer Interaction
experienced by students, tear (VSF). The VSF is an extens online, communication can ta on. We will use the results of	ted to explore the usefulness, case of learning and use, and satisfaction there, and community members interacting with a Virtual Science Fair sion of physical science fairs, in that exhibits can be created and browse se place among visitors and exhibitors, duding tables place cerinize, and so this study to refine the VSF. Mole: we are interested in how well the ill any inclivical is able to use it.
Procedures:	
work through several tasks de deliberately brief, in order to	brief background survey. Then, after reading some brief instructions you will seigned to introduce you to features of the VSF. These instructions will be see how well the system can support use on its own. After the tasks, you will and then will be given an opportunity to ask any questions you have about, c, or outcomes.
videotaping your actions with One or more evaluators will be that you encounter. To help a asking you to think aloud ab times we may prompt you to	with the VSF, we will be collecting serveral sorts of information; we will be the optam, as well as recording shart happens on each screen of the VSF to taking noise, measuring task start and stop time, and noting any problem so taking noise, measuring task start and stop time, and noting any problem boutly our goals, expectations, and reactions to the system as you work. At let lus what you are thinking. We relative that providing a thrin-aloud ng, but it is important for us to know what you are thinking as you carry out
Participant Consent:	
spend evaluating it. All data will be archived in a secure to your data are reported or des to participation in this experin	stiment is entirely voluntary. There will be no remuneration for the time you guithered from the unability study will be treated in a confidential fastion: it location and will be interpreted only for purposes of this evaluation. When troube, all identifying information will be removed. There are no known risk ment, and you may withdraw at any point. Please feel free to ask the her questions; otherwise, if you are willing to participate, please sign and the properties of the properties o
Name	Date
If there are any questions, p	please contact Dr. Mary Beth Rosson, 231-6470, rosson@vt.edu

		Jser Backgroun	d Survey	
Virginia Te	ch Computer Scien	nce Department and	Center for Human-Comput	er Enteraction
about your backgro to better understan	und-your experie 5 your interactions	once with science to	e we begin, it will be useful to ins, with technology, and so e system. Homember that a information.	on. This will help us
Name:			Age:	
Occupation:			Years of Education:	
Years a resident in	Blacksburg:			
For how many year	s have you been a	using computers?		
Please describe yo	ur typical compute	ruso (e.g., over a p	eriod of a week):	
			M.Pr. of control	
please describe:	the following 3 iter	es by circling the or	.MUDs, other online comm,	
please describe: Please respond to that in some cases.	the following 3 iter this may require	ns by circling the op you to make a predi	inias that best corresponds often about online activities.	
Please respond to that in some cases. 1. Browning an onl	the following 3 iter this may require	ns by circling the op you to make a predi	ipigs that best corresponds	
Please respond to that in some cases, 1. Browning an onl	the following 3 iter this may require : line science exhibit	ns by circling the op you to make a predi t is like visiting a sci News	izigs that best corresponds ction about online activities. sence fair in the real world.	to your own. Note
Please respond to that in some cases. 1. Browning an ord String's Disagree 2. Online science in Security	the following 3 iter this may require : line science exhibit	ns by circling the op you to make a predi t is like visiting a so Norm! e and of interest to	iziga that best corresponds ction about online activities. ience fair in the real world. Agree	to your own. Note
Please describe: Please respond to 1 that in some cases. 1. Browning an order to the sound to	the following 3 item this may require time science exhibit Desgree solubits are diversi Desgree	ns by circling the op you to make a predi t is like visiting a sci Nevnd e and of interest to Nevnd	ipias that best corresponds ction about online activities, ience fair in the real world. Agree a wide range of visitors.	Shongly Agree
Please describe: Please respond to that in some cases. 1. Browning an ord Strongly Disagree 2. Online science is Strongly Disagree 3. There are many	the following 3 iter this may require tine science whibit Diagne whibits are divers Diagne opportunities for r	ns by citaling the appear to make a predict is the visiting a solution of the second of interest to become involved in the control of the con	initias that best corresponds cition about ordina activities. ience fair in the real world. Agree a wide range of visitors. Agree and in students' science project	Shorely Agree ects. Shorely Agree ects. Shorely
Please describe: Please respond to 1 that in some cases. 1. Browning an ord Strongly Disagree 2. Online science is Strongly Disagree 3. There are many	the following 3 item this may require time science exhibit Desgree solubits are diversi Desgree	ns by circling the op you to make a predi t is like visiting a sci Nevnd e and of interest to Nevnd	izida that best corresponds cition about online activities, ierce fair in the real world. Agree a wide range of visitors. Agree	Shongly Agree ects.
Please respond to that in some cases. 1. Browning an ord Street	the following 3 iter this may require: the science exhibit Disagree exhibits are diverse Disagree opportunities for a Disagree	no by sitcling the ap you to make a predict to like visiting a sci Mental e and of interest to Mental me to become invol	initias that best corresponds cition about ordina activities. ience fair in the real world. Agree a wide range of visitors. Agree and in students' science project	Secretary Agree Secretary Agree Secretary Agree Secretary Agree Secretary Agree
Please respond to that in some cases. 1. Browning an ord Street	the following 3 iter this may require: the science exhibit Disagree exhibits are diverse Disagree opportunities for a Disagree	no by sitcling the ap you to make a predict to like visiting a sci Mental e and of interest to Mental me to become invol	icides that best corresponds cities about ordina activities. ierce fair in the real world. Agree a wide range of visitors. Agree and in students' science propagate.	Secretary Agree Secretary Agree Secretary Agree Secretary Agree Secretary Agree
Please respond to that in some cases. 1. Browning an ord Street	the following 3 iter this may require: the science exhibit Disagree exhibits are diverse Disagree opportunities for a Disagree	no by sitcling the ap you to make a predict to like visiting a sci Mental e and of interest to Mental me to become invol	icides that best corresponds cities about ordina activities. ierce fair in the real world. Agree a wide range of visitors. Agree and in students' science propagate.	Secretary Agree Secretary Agree Secretary Agree Secretary Agree Secretary Agree

Forms

(see Rosson & Carroll chap 7.4)

- General instructions

 - Do the tasks as described
 Ask for help only when give up
 Explain purpose; evaluating system, not participants
 Can leave at any time
- Task instructions
 - don't give away the interface!
- User Reactions Survey

 - Ask only for what you needPilot test it to be sure you're asking the right questions

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 Tanks 1-4:	
Background of Takes 1-1. Image the sty ou are Mr. King, an experienced science teacher who has coached many student projects in the past. This year you are admining Sally Harm on the Malach Note project. You are both very busy, but have arranged on one the VSV or a FF Moury Vivony variety. Sail a salway there sucking, the second of the Mr. William of the VSV or a FF Moury Vivon yave vivo. Sail a salway there sucking.	
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 2:	
 Open the switch element that is an Excel document and modify the title of the chart to be "leteraction of Mass and Movement". Save your change. 	
Task 4: Open the simulation element and add a nested folder named "Veistor Experiments".	
User Reactions Survey	
Virginia Tech Computer Science Department and Content for Human-Computer Interaction Now that you have completed that ISE facils, we would like to linear some of your reactions, both in general, and so the people for leases of the long.	
Name: What free throat did you like most about the VSF? Why?	
1,	
What there things did you like least about the YSF? Why? 1. 2.	
3. If the SF was made available to you, would you use it or not? O Yes O No Why?	
Please respond to the following 10 items by <u>circling the opinion</u> that best corresponds to your own.	
Browsing an online science exhibit is like visiting a science fail in the real world. Browly Browly Browly Browle Contraction concludes any devices and of linkeed to a vide or ganged visitions. Contraction concludes any devices and of interest to a vide or ganged visitions.	
Strongy Disagree Disagree Neutral Agree Agree 3. There are many opportunities for me to become involved in students' science posiects.	
Brough Discope Scale April April A James of other violates (continue and other sehila in the VEF. Compare No. 20 Compare No. 2	
Movement from one location in the VSF to another is awkexard. Binnyly Disagre	
It is easy and robustal to join other violetrs at an exhibit. Through the control of th	
Experimenting with mid-bit components (e.g. (small-bite(t)) in a not enjoyatevity proper to recover to re	

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

Usability Testing Single Camera Videotaping Method Computer VCR monitor

Usability Testing Two Camera Videotaping Method Picture in Pictu

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
28	52
9	10
12	11
18	2
	28 9 12

• 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

_			
_			
_			
_			
_			
_			
_			
_			
_			
_			
_			
_			
_			
_			
_			