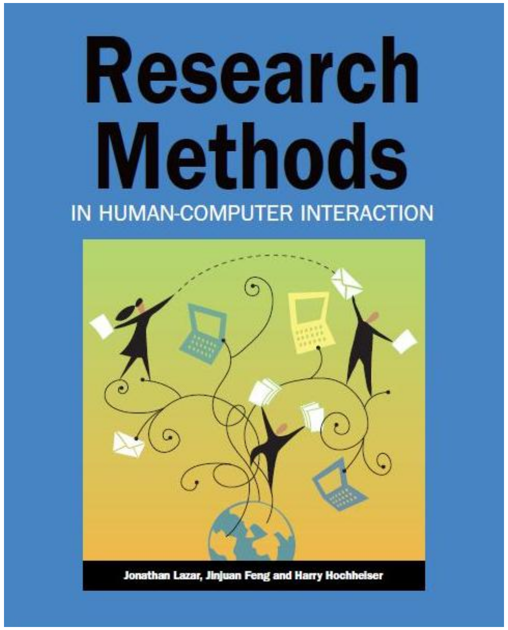
Research Methods in Human-Computer Interaction



Chapter 1- Introduction

History of HCI Research

- The first HCI conference was held in 1982
- But other work was going on before 1982
  - Conferences such as HFES
  - Books such as “Software Psychology”
  - Work at governmental agencies (e.g. SSA)
  - “Office automation” research
- The name “human-computer interaction” or “computer-human interaction” was first applied to the 1982 conference

Why was HCI needed?

- In the late 1970s and early 1980s, there was a shift:
  - From large computers in secured rooms, operated only by engineers
  - To small computers, operated by people without a technical background, in homes and workplaces
- Ease of use, the human side, user acceptance, all became more important!

Changes in HCI research

- The main topics of HCI have shifted over time:
- 1980s
  - Word processing and database interfaces
- 1990s
  - Web usability, e-mail, groupware
- 2000s
  - User-generated content, tagging, social networking

HCI Research

- HCI research is at the intersection of rigor and relevance
  - There must be a good base of theory, and the research methods must be rigorous
  - There must be practical implications, simply influencing theory is not sufficient
- Historically, HCI work has focused on practical results that improve the quality of life

HCI Research

- HCI research involves a number of different disciplines:
  - Computer science
  - Management
  - Cognitive psychology
  - Library science
  - Communication
  - Design
- It’s always going to be complex, and there will always be many different approaches

Measurement in HCI

- There are many different approaches to measurement
- The traditional measurements are:
  - Task performance, time performance, and user satisfaction
- Those measurements do not accurately measure:
  - Why people no longer use an interface
  - Discretionary use for enjoyment (e.g. YouTube)
  - Emotion and trust

New forms of measurement

- A feeling of community?
- Emotion?
- Enjoyment?
- Physiological measures (EEG, EMG)?
- Lower carbon footprint?
- Satisfaction from accomplishments in gaming or in virtual worlds?
- Ease of use and enjoyment of new forms of technology can be challenging to measure

Triangulation

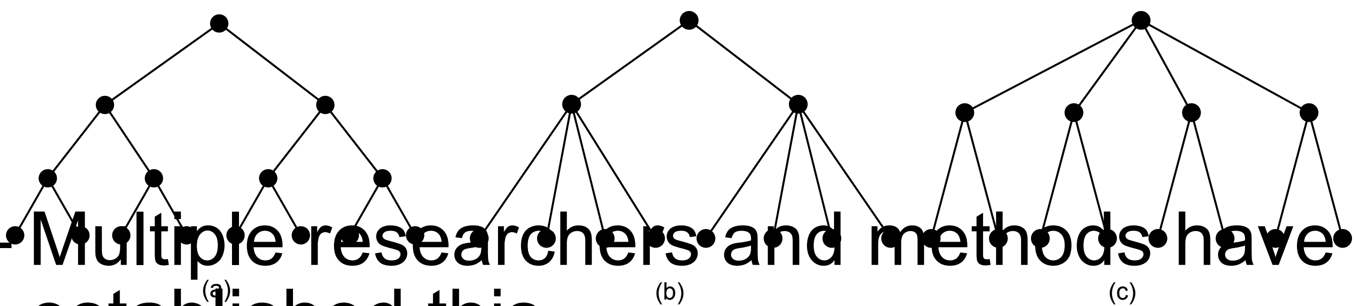
- No data collection method will be perfect
- It is important to have multiple researchers, using multiple methods, investigating the same phenomenon
- We call this triangulation
- One paper ≠ scientific truth
- Different researchers, different methods, all coming to the same conclusion, THAT’S when you find consensus

Consensus in research

- Many core questions in HCI remain unanswered
  - What is the minimum number of people required for usability testing?
- Or the answers change over time due to changing technology
  - What is the biggest frustrations for users on the web? (download speed? Viruses? Browser compatibility?)

Consensus in research

- One area where there is clear consensus in the HCI research:
  - Broad, shallow tree structures (c) in information spaces are superior to narrow, deep structures (a and b)



  - Multiple researchers and methods have established this

Inherent conflicts in HCI

- HCI research is complex
- There often is not one optimal solution
- There are trade-offs and multiple stakeholders with conflicting goals
  - Users prefer consistency over change
  - Often, there is a trade-off between usability and security (the highest ease of use would be with no security, which isn’t possible)
  - HCI research can be hard to cost-justify

Interdisciplinary nature

- HCI research involves many disciplines:
  - Human factors, engineering, psychology (were heavily involved from the start)
  - Art and design, library and information science (very involved right now)
- Each discipline has a tradition of research methods
- These methods are modified for use in HCI research, some new ones are created

Interdisciplinary nature

- We have a long history of controlled studies in laboratories measuring task and time performance from psychology and engineering
- As new technologies are developed involving communication, socializing and emotion, we need to modify and use methods from the social sciences
- New approaches to research, and multi-method research, is needed

Interdisciplinary research

- Interdisciplinary research can be challenging
  - Focus on journals vs. conferences
  - Single author vs. group author publications
  - Focus on theory vs. applied, practical work
  - Focus on grant money vs. anti-grant money
  - Some disciplines are very self-reflective
  - Even the appropriate attire varies
- This is true even if all disciplines call themselves “HCI researchers”

Different Focus on each step

- Different disciplines may focus more on one step of the research process than others
  - Theory
  - Research Methods
  - Participants/context
  - Statistics/analysis
  - Implications for interfaces
- We expect this comment to be somewhat controversial, but based on our experience, we believe strongly that it is true

Communicating your ideas

- You need to be familiar with research methods from different disciplines
- Know what the “sensitive spots” are from other disciplines
- You need to communicate your results in a way that others can understand
- People from other disciplines will be reviewing your paper or your grant
- Be prepared to answer: “why did we use method X instead of methods Y or Z?”

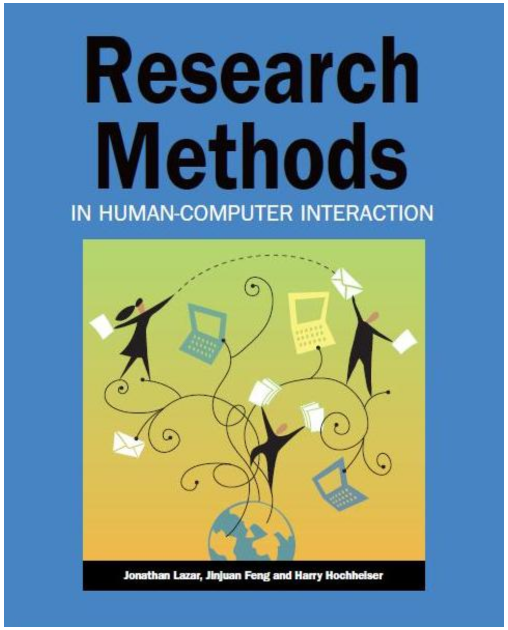
Research and Usability Testing

- Is usability testing the same as research?
- The methods used in UT and research can be the same (experimental model, task and time measurements, lab setting)
- The GOALS of UT and research are often different:
  - UT: find and fix flaws in specific interfaces, using any method needed, and few users
  - Research: larger number of users, results that can be generalized
  - For more information on differences, see chapter 10

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 2- Experimental Research

Overview

- Types of behavioral research
- Research hypotheses
- Basics of experimental research
- Significance tests
- Limitations of experimental research

Types of behavioral research

- Descriptive investigations focus on constructing an accurate description of what is happening.
- Relational investigations enable the researcher to identify relations between multiple factors. However, relational studies can rarely determine the causal relationship between multiple factors.
- Experimental research allows the establishment of causal relationship.

Types of behavioral research

Type of research	Focus	General claims	Typical methods
Descriptive	Describe a situation or a set of events	X is happening	Observations, field studies, focus groups, interviews
Relational	Identify relations between multiple variables	X is related to Y	Observations, field studies, surveys
Experimental	Identify causes of a situation or a set of events	X is responsible for Y	Controlled experiments

Table 2.1 Relationship between descriptive research, relational research, and experimental research.

Research hypotheses

- An experiment normally starts with a research hypothesis.
- A hypothesis is a precise problem statement that can be directly tested through an empirical investigation.
- Compared with a theory, a hypothesis is a smaller, more focused statement that can be examined by a single experiment

Types of hypotheses

- Null hypothesis: typically states that there is no difference between experimental treatments.
- Alternative hypothesis: a statement that is mutually exclusive with the null hypothesis.
- The goal of an experiment is to find statistical evidence to refute or nullify the null hypothesis in order to support the alternative hypothesis.
- A hypothesis should specify the independent variables and dependent variables.

Research hypotheses

- Independent variables (IV) refer to the factors that the researchers are interested in studying or the possible "cause" of the change in the dependent variable.
  - IV is independent of a participant's behavior.
  - IV is usually the treatments or conditions that the researchers can control.
- Dependent variables (DV) refer to the outcome or effect that the researchers are interested in.
  - DV is dependent on a participant's behavior or the changes in the IVs
  - DV is usually the outcomes that the researchers need to measure.

Typical independent variables in HCI

- Those that relate to technology
  - Types of technology or device
  - Types of design
- Those that relate to users: age, gender, computer experience, professional domain, education, culture, motivation, mood, and disabilities
- Those that relate to context of use:
  - Physical status
  - User status
  - Social status

Typical dependent variables in HCI

- Efficiency:
  - e.g., task completion time, speed
- Accuracy:
  - e.g., error rate
- Subjective satisfaction:
  - e.g., Likert scale ratings
- Ease of learning and retention rate
- Physical or cognitive demand
  - e.g., NASA task load index

Components of experiment

- Treatments, or conditions: the different techniques, devices, or procedures that we want to compare
- Units: the objects to which we apply the experiment treatments. In HCI research, the units are normally human subjects with specific characteristics, such as gender, age, or computing experience
- Assignment method: the way in which the experimental units are assigned different treatments.

Randomization

- Randomization: the random assignment of treatments to the experimental units or participants
- In a totally randomized experiment, no one, including the investigators themselves, is able to predict the condition to which a participant is going to be assigned
- Methods of randomization
  – Preliminary methods
  – Random table
  – Software driven randomization

Significance tests

- Why do we need significance tests?
  - When the values of the members of the comparison groups are all known, you can directly compare them and draw a conclusion. No significance test is needed since there is no uncertainty involved.
  - When the population is large, we can only sample a sub-group of people from the entire population.
  - Significance tests allow us to determine how confident we are that the results observed from the sampling population can be generalized to the entire population.

Type I and Type II errors

- All significance tests are subject to the risk of Type I and Type II errors.
- A Type I error (also called an *α error or a “false positive”) refers to the mistake of* rejecting the null hypothesis when it is true and should not be rejected.
- A Type II error (also called a *β error or a “false negative”) refers to the mistake of not rejecting the null* hypothesis when it is false and should be rejected.

Type I and Type II errors

		Jury decision	
		Not guilty	Guilty
Reality	Not guilty	√	Type I error
	Guilty	Type II error	√

Table 2.3 Type I and Type II errors in the judicial case.

		Study conclusion	
		No difference	Touchscreen ATM is easier to use
Reality	No difference	√	Type I error
	Touchscreen ATM is easier to use	Type II error	√

Table 2.4 Type I and Type II errors in a hypothetical HCI experiment.

Type I and Type II errors

- It is generally believed that Type I errors are worse than Type II errors.
- Statisticians call Type I errors a mistake that involves "gullibility".
  - A Type I error may result in a condition worse than the current state.
- Type II errors are mistakes that involve "blindness"
  - A Type II error can cost the opportunity to improve the current state.

Controlling risks of errors

- In statistics, the probability of making a Type I error is called alpha (or significance level, p value).
- The probability of making a Type II error is called beta.
- The statistical power of a test, defined as $1-\beta$, refers to the probability of successfully rejecting *a* null hypothesis when it is false and should be rejected

Controlling risks of errors

- Alpha and beta are interrelated. Under the same conditions, decreasing alpha reduces the chance of making Type I errors but increases the chance of making Type II errors.
- In experimental research, it is generally believed that Type I errors are worse than Type II errors.
- So a very low p value (0.05) is widely adopted to control the occurrence of Type I errors.

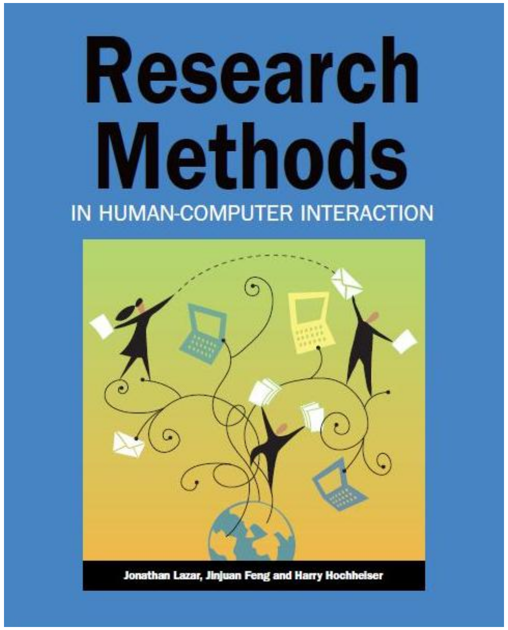
Limitations of Experimental Research

- Experimental research requires well-defined, testable hypotheses that consist of a limited number of dependent and independent variables.
- Experimental research requires strict control of factors that may influence the dependent variables.
- Lab-based experiments may not be a good representation of users' typical interaction behavior.

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



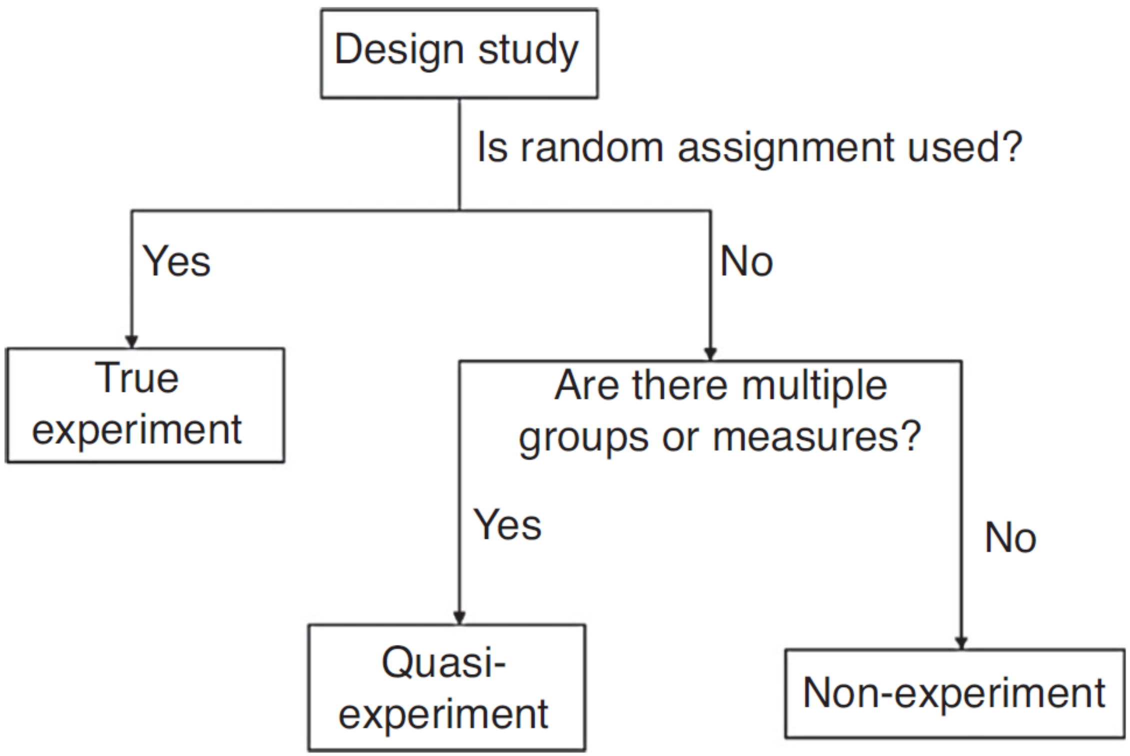
Chapter 3- Experimental Design

Overview

- What needs to be considered when designing experiments?
- Determining the basic design structure
- Investigating a single independent variable
- Investigating more than one independent variable
- Reliability of experimental results
- Experimental procedures

Three groups of studies

- Experiments, quasi-experiments, and non-experiments:



Characteristics of true experiments

- A true experiment
  - is normally based on at least on hypothesis
  - have multiple conditions
  - The dependent variable can be quantitatively measured
  - uses statistical significance tests
  - thrives to remove biases
  - is replicable

Factors to consider

- Research hypothesis
  - Clearly defined
  - Appropriate scope
- Dependent variables
  - Easy to measure
- Independent variables and conditions
  - Easy to control

Basic design structure

- Two basic questions:
  - How many independent variables do we want to investigate in the experiment?
  - How many different values does each independent variable have?

Basic design structure

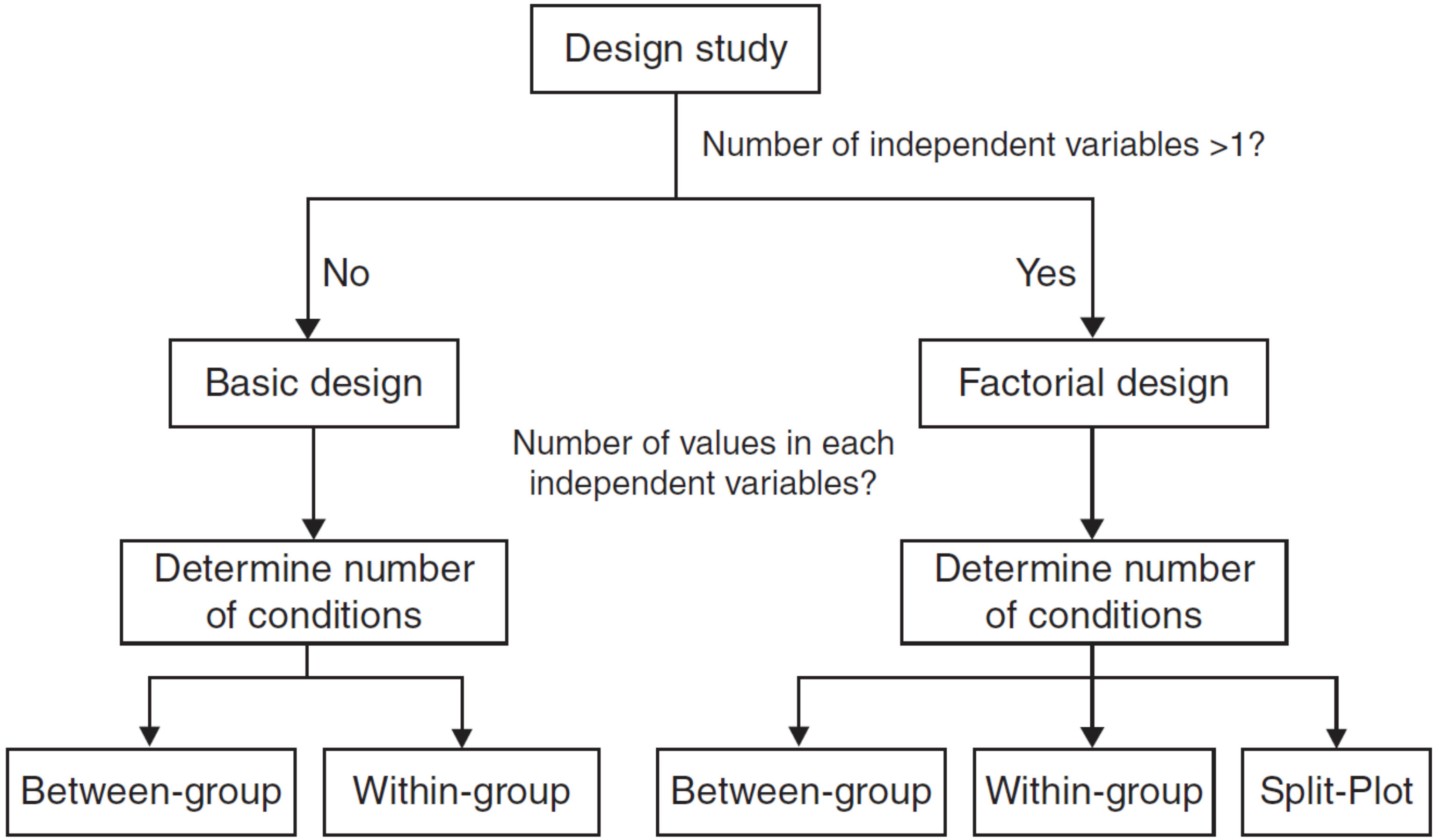


Figure 3.2 Determining the experiment structure.

Investigating one independent variable - Between group design

- Also called 'between subject design'
- One participant only experience one condition

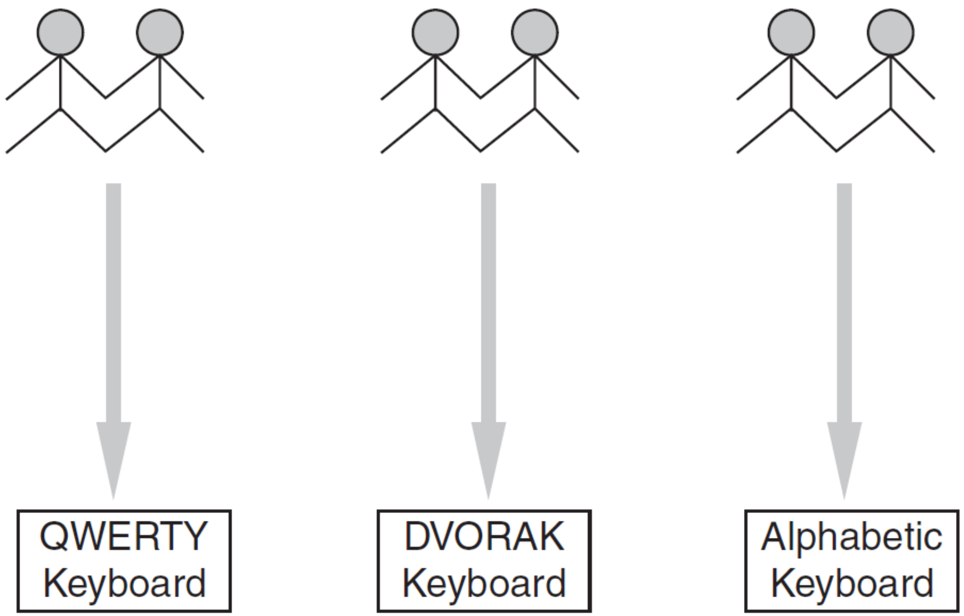


Figure 3.3 Between-group design.

Between group design

- Advantages
  - Cleaner, better control of learning effect
  - Requires shorter time for participants
  - less impact of fatigue and frustration
- Disadvantages
  - Impact of individuals difference
  - Harder to detect difference between conditions
  - Require larger sample size

Investigating one independent variable - Between group design

- Also called 'within subject design'
- One participant experience multiple conditions

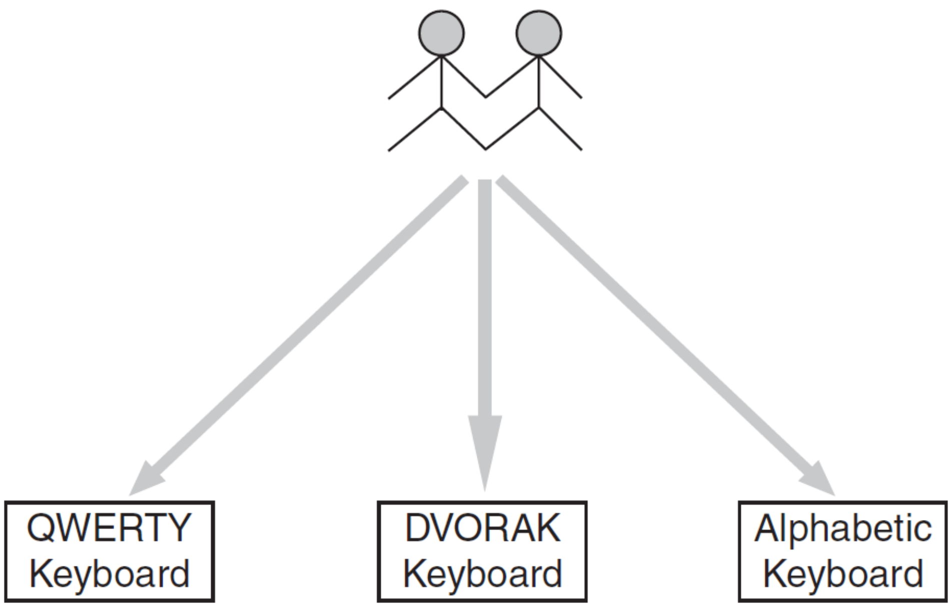


Figure 3.4 Within-group design.

Within-group design

- Advantages
  - Requires smaller sample size
  - Easy to detect difference between conditions
- Disadvantages
  - Learning effect
  - Takes longer time
  - Larger impact of fatigue and frustration

Investigating one independent variable - Between group vs. Within group

- Between-group design should be taken when:
  - Simple tasks
  - Learning effect has large impact
  - Within-group design is impossible

- Within-group design should be taken when:
  - Learning effect has small impact
  - Small participant pool

More than one independent variable

- Factorial design divides the experiment groups or conditions into multiple subsets according to the independent variables
- Can study interaction effects
- Number of conditions:

$$C = \prod_{a=1}^{n} Va$$

More than one independent variable

- Three options of factorial design
  - Between group design
  - Within group design
  - Split-plot design

- Split-plot design
  - Has both a between-group and a within-group component

Interaction effect

- The differing effect of one independent variable on the dependent variable, depending on the particular level of another independent variable

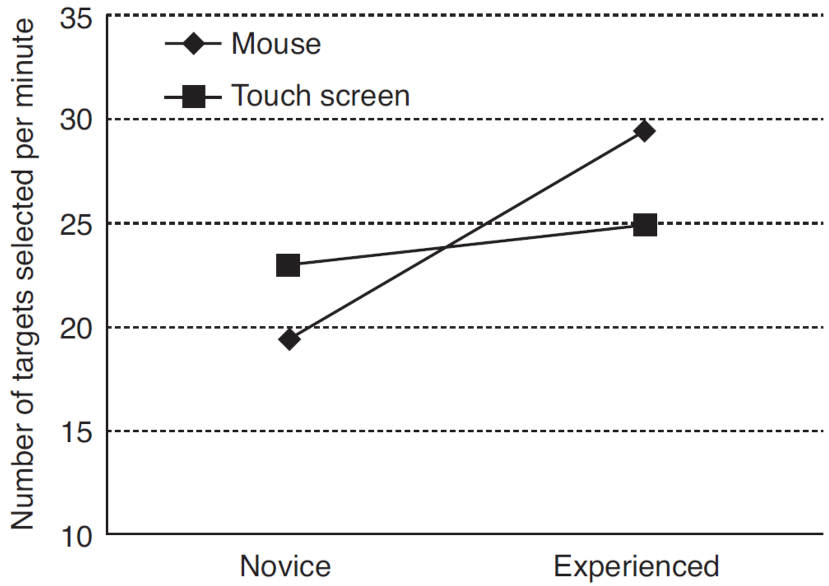


Figure 3.6 Interaction effects.

Reliability of experiments

- Random errors
  - Also called ‘chance errors’ or ‘noises’
  - Cause variations in both direction
  - Occur by chance
  - Can be controlled by a large sample size
- Systematic errors
  - Also called ‘biases’
  - Always push actual value in the same direction
  - Can never be offset no matter how large the sample is

Reliability of experiment results

- Five major sources of system errors
  - measurement instruments
  - experimental procedures
  - participants
  - experimenter behavior
  - the experimental environment

Lifecycle of an experiment

- Identify a research hypothesis
- Specify the design of the study
- Run a pilot study to test the design, the system, and the study instruments
- Recruit participants
- Run the actual data collection sessions
- Analyze the data
- Report the results

Experiment session procedure

- Ensure the systems or devices being evaluated and the related instruments are ready for the experiment
- Greet the participants
- Introduce the purpose of the study and the procedures
- Get the consent of the participants
- Assign the participants to a specific experiment condition according to the pre-defined randomization method

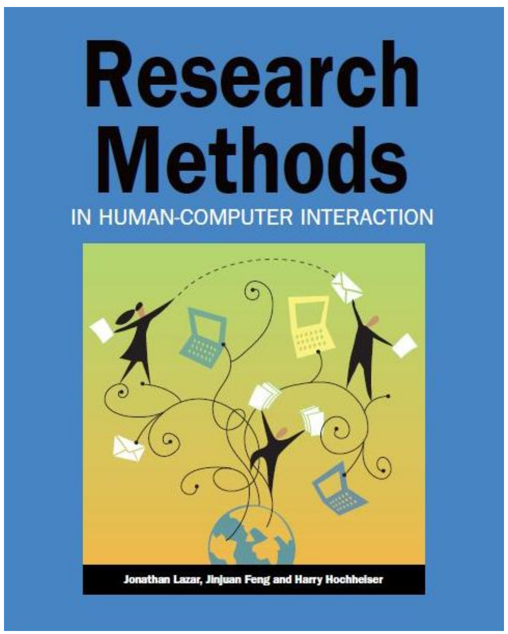
Experiment session procedure

- Participants complete training task
- Participants complete actual tasks
- Participants answer questionnaires (if any)
- Debriefing session
- Payment (if any)

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 4-
Statistical Analysis

Overview

- Preparing data for statistical analysis
- Descriptive statistics
- Comparing means
- *T tests*
- Analysis of variance (ANOVA)
- Assumptions of *t tests and F tests*
- Identifying relationships
- Regression
- Nonparametric statistical tests

Preparing data for analysis

- Cleaning up data
  - Detect errors
  - Formatting
- Coding
  - Types of data that need to be coded
  - Be consistent
- Organizing the data
  - Accommodate to the requirements of statistical software

Descriptive statistics

- Measures of central tendency
  - Mean
  - Median
  - Mode
- Measures of spread
  - Range
  - Variance
  - Standard deviations

Comparing means

- Summary of methods

Experiment design	Independent variables (IV)	Conditions for each IV	Types of test
Between-group	1	2	Independent-samples *t* test
	1	3 or more	One-way ANOVA
	2 or more	2 or more	Factorial ANOVA
Within-group	1	2	Paired-samples *t* test
	1	3 or more	Repeated measures ANOVA
	2 or more	2 or more	Repeated measures ANOVA
Between- and within-group	2 or more	2 or more	Split-plot ANOVA

Table 4.3 Commonly used significance tests for comparing means and their application context.

Comparing 2 means: T tests

- Independent-samples t test: between-group design

Group	Participants	Task completion time	Coding
No prediction	Participant 1	245	0
No prediction	Participant 2	236	0
No prediction	Participant 3	321	0
No prediction	Participant 4	212	0
No prediction	Participant 5	267	0
No prediction	Participant 6	334	0
No prediction	Participant 7	287	0
No prediction	Participant 8	259	0
With prediction	Participant 1	246	1
With prediction	Participant 2	213	1
With prediction	Participant 3	265	1
With prediction	Participant 4	189	1
With prediction	Participant 5	201	1
With prediction	Participant 6	197	1
With prediction	Participant 7	289	1
With prediction	Participant 8	224	1

Table 4.4 Sample data for independent-samples *t* test.

Comparing 2 means: T tests

- Paired-sample t test: within-group design

Participants	No prediction	With prediction
Participant 1	245	246
Participant 2	236	213
Participant 3	321	265
Participant 4	212	189
Participant 5	267	201
Participant 6	334	197
Participant 7	287	289
Participant 8	259	224

Table 4.5 Sample data for paired-samples *t* test.

Comparing 2 or more means: Analysis of variance (ANOVA)

- Also called F tests
- One-way ANOVA: for between-group design
- Data layout: Table 4.6
- Results summary:

Source	Sum of squares	df	Mean square	*F*	Significance
Between-group	7842.250	2	3921.125	2.174	0.139
Within-group	37880.375	21	1803.827		

Table 4.7 Result of the one-way ANOVA test.

Group	Participants	Task completion time	Coding
Standard	Participant 1	245	0
Standard	Participant 2	236	0
Standard	Participant 3	321	0
Standard	Participant 4	212	0
Standard	Participant 5	267	0
Standard	Participant 6	334	0
Standard	Participant 7	287	0
Standard	Participant 8	259	0
Prediction	Participant 1	246	1
Prediction	Participant 2	213	1
Prediction	Participant 3	265	1
Prediction	Participant 4	189	1
Prediction	Participant 5	201	1
Prediction	Participant 6	197	1
Prediction	Participant 7	289	1
Prediction	Participant 8	224	1
Speech-based dictation	Participant 1	178	2
Speech-based dictation	Participant 2	289	2
Speech-based dictation	Participant 3	222	2
Speech-based dictation	Participant 4	189	2
Speech-based dictation	Participant 5	245	2
Speech-based dictation	Participant 6	311	2
Speech-based dictation	Participant 7	267	2
Speech-based dictation	Participant 8	197	2

Table 4.6 Sample data for one-way ANOVA test.

Factorial ANOVA

- For between-group design
- 2 or more independent variables involved
- Data layout: table 4.9

	Standard	Prediction	Speech
Transcription	Group 1	Group 2	Group 3
Composition	Group 4	Group 5	Group 6

Table 4.8 A between-group factorial design with two independent variables.

Task type	Entry method	Participant number	Task time	Task type coding	Entry method coding
Transcription	Standard	Participant 1	245	0	0
Transcription	Standard	Participant 2	236	0	0
Transcription	Standard	Participant 3	321	0	0
...	...	...	...	...	...
Transcription	Prediction	Participant 9	246	0	1
Transcription	Prediction	Participant 10	213	0	1
Transcription	Prediction	Participant 11	265	0	1
...	...	...	...	...	...
Transcription	Speech-based dictation	Participant 17	178	0	2
Transcription	Speech-based dictation	Participant 18	289	0	2
Transcription	Speech-based dictation	Participant 19	222	0	2
...	...	...	...	...	...
Composition	Standard	Participant 25	256	1	0
Composition	Standard	Participant 26	269	1	0
Composition	Standard	Participant 27	333	1	0
...	...	...	...	...	...
Composition	Prediction	Participant 33	265	1	1
Composition	Prediction	Participant 34	232	1	1
Composition	Prediction	Participant 35	254	1	1
...	...	...	...	...	...
Composition	Speech-based dictation	Participant 41	189	1	2
Composition	Speech-based dictation	Participant 42	321	1	2
Composition	Speech-based dictation	Participant 43	202	1	2
...	...	...	...	...	...

Table 4.9 Sample data for the factorial ANOVA test.

Factorial ANOVA

- Summary results

Source	Sum of square	Df	Mean square	*F*	Significance
Task type	2745.188	1	2745.188	1.410	0.242
Entry method	17564.625	2	8782.313	4.512	0.017
Task*entry	114.875	2	57.437	0.030	0.971
Error	81751.625	42	1946.467		

Table 4.10 Result of the factorial ANOVA test.

Repeated measures ANOVA

- For within-group design
- Can investigate one or more variables
- One-way ANOVA

	Standard	Prediction	Speech
Participant 1	245	246	178
Participant 2	236	213	289
Participant 3	321	265	222
Participant 4	212	189	189
Participant 5	267	201	245
Participant 6	334	197	311
Participant 7	287	289	267
Participant 8	259	224	197

Table 4.11 Sample data for one-way repeated measures ANOVA.

Repeated measures ANOVA

- One way ANOVA summary report:

Source	Sum of square	Df	Mean square	*F*	Significance
Entry method	7842.25	2	3921.125	2.925	0.087
Error	18767.083	14	1340.506		

Table 4.12 Result of the one way repeated measures ANOVA test.

Repeated measures ANOVA

- Two way ANOVA experiment design:

	Standard	Prediction	Speech
Transcription	Group 1	Group 1	Group 1
Composition	Group 1	Group 1	Group 1

Table 4.13 Experiment design of a two-way, repeated measures ANOVA.

Repeated measures ANOVA

Two way ANOVA data layout

	Transcription			Composition		
	Standard	Prediction	Speech	Standard	Prediction	Speech
Participant 1	245	246	178	256	265	189
Participant 2	236	213	289	269	232	321
Participant 3	321	265	222	333	254	202
Participant 4	212	189	189	246	199	198
Participant 5	267	201	245	259	194	278
Participant 6	334	197	311	357	221	341
Participant 7	287	289	267	301	302	279
Participant 8	259	224	197	278	243	229

Table 4.14 Sample data for two-way, repeated measures ANOVA test.

Repeated measures ANOVA

- Two way ANOVA summary report:

Source	Sum of square	df	Mean square	*F*	Significance
Task type	2745.187	1	2745.187	14.217	0.007
Error (task type)	1351.646	7	193.092		
Entry method	17564.625	2	8782.313	2.923	0.087
Error (entry method)	42067.708	14	3004.836		
Task type * entry method	114.875	2	57.438	0.759	0.486
Error (task type * entry method)	1058.792	14	75.628		

Table 4.15 Result of the two-way, repeated measures ANOVA test.

Split-plot ANOVA

- Involves both between-group and within-group factors
- Experiment design

	Keyboard	Prediction	Speech
Transcription	Group 1	Group 1	Group 1
Composition	Group 2	Group 2	Group 2

Table 4.16 Split-plot experiment design.

Split-plot ANOVA data layout

Task type	Participant number	Task type coding	Standard	Prediction	Speech
Transcription	Participant 1	0	245	246	178
Transcription	Participant 2	0	236	213	289
Transcription	Participant 3	0	321	265	222
Transcription	Participant 4	0	212	189	189
Transcription	Participant 5	0	267	201	245
Transcription	Participant 6	0	334	197	311
Transcription	Participant 7	0	287	289	267
Transcription	Participant 8	0	259	224	197
Composition	Participant 9	1	256	265	189
Composition	Participant 10	1	269	232	321
Composition	Participant 11	1	333	254	202
Composition	Participant 12	1	246	199	198
Composition	Participant 13	1	259	194	278
Composition	Participant 14	1	357	221	341
Composition	Participant 15	1	301	302	279
Composition	Participant 16	1	278	243	229

Table 4.17 Sample data for the split-plot ANOVA test.

Split-plot ANOVA summary report

Source	Sum of square	df	Mean square	*F*	Significance
Task type	2745.187	1	2745.187	0.995	0.335
Error	38625.125	14	2758.937		

Table 4.18 Results of the split-plot test for the between-group variable.

Source	Sum of square	df	Mean square	*F*	Significance
Entry method	17564.625	2	8782.313	5.702	0.008
Entry method * task type	114.875	2	57.437	0.037	0.963
Error (entry method)	43126.5	28	1540.232		

Table 4.19 Results of the split-plot test for the within-group variable.

Assumptions of t tests and F tests

- Errors should be independent of each other
- Errors should be identically distributed
- Errors should be normally distributed

Identify relationships

- Correlation: Two factors are correlated if there is a relationship between them
- Most commonly used test for correlation is the Pearson's product moment correlation coefficient test
- Pearson's r: ranges between -1 to 1
- Pearson's r square represents the proportion of the variance shared by the two variables

Identify relationships

- Correlation does not imply causal relationship

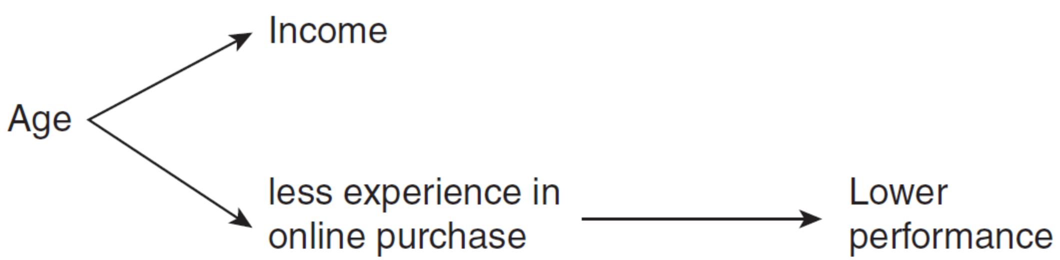


Figure 4.2 Relationship between correlated variables and an intervening variable.

Identify relationships

- Regression: can investigate the relationship between one DV and multiple IVs
- Regression is used for 2 purposes:
  - Model construction
  - Prediction
- Different regression procedures
  - Simultaneous
  - Hierarchical

Non-parametric tests

- Non-parametric tests are used when:
  - The error is not normally distributed
  - The distances between any two data units are not equal
  - The variance of error is not equal

Non-parametric tests

- CHI-square test
  - Used to analyze categorical data
  - Table of counts (contingency table)
  - Assumptions of the test
    - Data points need to be independent
    - The sample size should not be too small

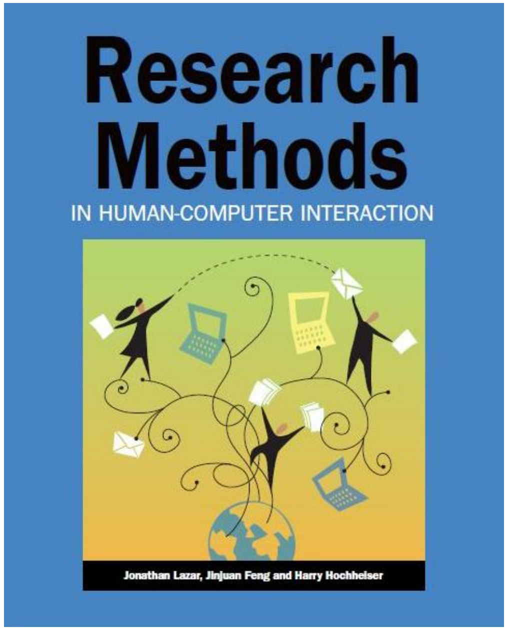
Non-parametric tests

- Two groups of data
  - For between-group design: Mann–Whitney U test or the Wald–Wolfowitz runs test
  - For within-group design: Wilcoxon signed ranks test
- Three or more groups of data
  - For between-group design: Kruskal–Wallis one-way analysis of variance by ranks
  - For within-group design: Friedman’s two-way analysis of variance test

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 5- Surveys

Introduction

- Surveys are a very commonly used research method
- Surveys are also often-maligned because they are not done in the proper manner
- A survey is a well-defined and well-written set of questions to which an individual is asked to respond
- Surveys are typically self-administered

Introduction

- Surveys are good at:
  - getting a large number of responses
  - Quickly
  - From a geographically dispersed population
- You can capture the “big picture” relatively quickly
- If structured random sampling is used, you might be able to make accurate population estimates

Introduction

- Doing a survey research study is easy, but doing a well-planned, valid survey study takes a lot of work
  - Surveys must reach the actual population of interest
  - Appropriate sampling methods must be used
  - Questions must be well-worded and tested
- Researchers are not present when users fill out a survey, so it must be easy to use, easy to understand, and easy to return

What is a survey?

- Is a survey the same thing as a questionnaire?
- A questionnaire is the actual list of questions
- A survey is the complete methodological approach
- But the two terms are often used interchangibly

Benefits of Surveys

- You can collect data from a large number of people, at a relatively low cost
- You can get an overview of a population of users in a short amount of time
- Surveys do not require any special equipment
- Surveys are generally approved by institutional review boards because they are usually non-intrusive

Drawbacks of surveys

- Surveys are good at getting shallow data from a large number of people, but not good at getting deep data
- Since surveys are usually self-administered, it is usually not possible to ask follow-up questions
- Surveys can lead to biased data when the questions are related to patterns of usage, rather than clear factual phenomena

Drawbacks of Surveys

- User's age or gender is a factual statement
- If you are asking users to recall usage patterns or mood, there may be recall bias
  - E.g. how many times did you use this software application over 6 months?
  - What was your mood when you used the software application?
- If recall bias may occur, use time diaries or data logging (or a combination) instead

Target user population

- The target population, or the population of interest, are the computer users who you want to study
- Is it a well-defined population? Do you know approximately how many people are in the target user population
- Set some parameters:
  - Age, gender, education, computer experience
  - Users of certain web sites, applications, OS

Target user population

- How will you contact the members of the target user population?
  - Is there a directory of targeted users?
  - An e-mail distribution list?
  - A postal mailing list?
  - A web site they all visit?
  - A social networking group?
  - Face-to-face meetings?
  - Membership in a certain organization
  - Job licensing or certification?

How to sample?

- Two major types of sampling methods:
- Probabilistic sampling
  - Where there is a known probability of someone being chosen
- Non-Probabilistic sampling
  - It is not exactly known what the likelihood of being chosen is
- Note that non-probabilistic sampling is accepted in HCI research, although some social sciences do not believe in it

Probabilistic sampling

- Two types of probabilistic sampling:
- A census
  - Where every single person in the targeted user population is chosen to take part in the survey
- A random sample
  - Where not all people are chosen to participate, but it is known what the likelihood is of being chosen to participate
- This can be targeted users, or targeted organizations (such as libraries)

Stratified sample

- A stratified sample is when you have an appropriate number of responses from each subset of your user population
- Example: a random sample of college students would not have an equal number of freshman, sophomores, juniors, and seniors
- A stratified random sample would have an equal number from each class year
- But It doesn’t need to be equal. It would still be stratified if you took 40% seniors, 40% juniors, 10% sophomores, and 10% junior. The researcher decides what is the appropriate breakdown

Response size

- What response is considered to be sufficient for a random sample?
- It depends on the confidence level and margin of error you consider acceptable
- For instance, to get a 95% confidence level and +-5% margin of error, you need 384 responses
- If the sample is large (5-10%) compared to the population size, the margin of error is smaller
- Only relevant for random sampling

Sources of error and bias

- Sampling error (not enough responses)
- Coverage error (not all members of the population of interest have an equal likelihood of being sampled)
- Measurement error (questions are poorly worded)
- Non-response error (major differences in the people who were sampled and the people who actually responded)

Non-probabilistic sampling

- Non-probabilistic sampling is used when:
  - You do not use a strict random sample
  - You do not know the likelihood of an individual being selected
  - You are not interested in a population estimate
  - There may not be a clearly defined population of interest
- Non-probabilistic sampling is considered acceptable in HCI research, because often, the HCI researcher must collect their own data (different from many other fields)

Demographic data

- Collecting demographic information is always important in survey data, but becomes more so when using non-probabilistic sampling
- Goal should be to demonstrate either:
  - Diverse, cross-section of respondents
  - A response that is somewhat representative of already-established, baseline data
- Often, non-probabilistic responses are fully anonymous, with no identifying data

Oversampling

- When there isn’t a well-defined list of users, no exact knowledge of population size, and random sampling isn’t possible, the number of responses becomes more important
- When the number of survey responses in large in proportion to the estimated or *perceived* population size, this is known as oversampling
- Helps establish informal validity

Oversampling

- Having a large number of responses can reduce the likelihood of excluding any segment of the population
- Not all researchers agree that oversampling increases validity
- 30 survey responses might be a baseline minimum for small estimated populations, while 500 or more might be considered minimum for larger estimated populations
- Oversampling would mean many more responses than 30 or 500, in those cases

Other techniques

- Random sampling of usage, not users
  - If a survey appears every 10th time that a web page is loaded, this is a random sampling of usage, not users. Users who visit the web page often will be over-represented (unless instructions or IP addresses are used to limit)
- Self-selected surveys
  - If a web page always has a link to a survey, everyone is invited to fill out the survey
  - Often, non-probabilistic surveys are self-selected, not random

Establishing informal validity

- If non-probabilistic surveys are used, both demographic information and response size both become important in establishing informal validity
- Self-selected, non-probabilistic surveys are often used as a first step in researching unknown research phenomena or user groups, where little is known about the user group or

Uninvestigated populations

- If researching an uninvestigated population, it may be challenging to find and recruit users who meet the criteria
- You may need to:
  - Partner with leaders of that community
  - Use snowball sampling (where one respondent recruits others that they know)
- While snowball sampling may lead to bias in who responds, it increases sample size

Developing survey questions

- The overall goal is to develop well-written, non-biased questions
- Since most surveys are self-administered, the questions need to stand alone, without any explanations
- You need to focus on both:
  - The overall structure of the entire survey
  - The wording of specific questions

Types of questions

- Open-ended questions
  – Respondents may provide more information, but it can be harder to do data analysis
  – Make sure to ask specific, not general questions. The answer should be open-ended, but the question should not be
- Closed-ended questions
  – Ordered response (e.g. ranking or likert scale)
  – Unordered response (e.g. multiple choice)

Common problems w/questions

- Asking two separate, and possibly related questions in one question (respondents often don’t answer both questions)
- The use of negative words in questions can cause confusion
- Biased wording in questions
- Identifying the position of a well-respected person or organization
- The use of “hot-button” words

Overall survey structure

- All surveys must begin with instructions
  - On paper, should checkboxes and ovals be filled in, checked, an “X” placed in them?
  - Should all respondents fill out all questions?
  - A reminder of who qualifies to participate, and who does not
- Each section of the survey should have a heading
- What path through the survey should the respondent take?

Overall survey structure

- If the survey is electronic, are help links provided?
- If a paper survey:
  - Is there enough white space?
  - Is white paper used? Are standard fonts used?
  - Do folds and staples interfere with any text?
- Is contact info (e-mail, phone, web site, etc.) provided if the respondent has any questions?
- If some questions are “contingent” make that clear using arrows, boxes, and indenting

Overall survey structure

- Questions related to a similar topic should be grouped together
- It’s generally NOT a good idea to randomize the order of the questions
- Provide interesting questions at the beginning of the survey
- Leave demographic questions until the end of the survey
- If there are any sensitive or objectionable questions, leave them until the end, when the respondent has become interested!
- Be reasonable about the length of the survey

Existing surveys

- There are some existing surveys that have been tested and validated in the HCI literature, primarily for usability testing and evaluation:
  - Computer System Usability Questionnaire
  - Interface Consistency Testing Questionnaire
  - Questionnaire for User Interaction Satisfaction
  - Website Analysis and Measurement Inventory
- See book and web site for a list of more surveys

Paper or online surveys?

- How could you contact the potential respondents?
  - Phone?
  - Postal mailing addresses?
  - E-mail addresses?
  - A social networking site?
- Find out if all potential respondents have internet/e-mail access
  - If not, you must use either paper surveys, or a hybrid approach

Paper or online surveys?

- Choose the most practical approach
- Using a hybrid design, with both paper and electronic surveys, can improve the number of responses and insure representation from different portions of the target user population
  - But make sure the different forms of the survey are exactly the same!

Paper or online surveys?

- Copying costs, mailing, postage, and data entry costs can be eliminated with electronic surveys
- Initial set-up costs for web-based surveys are higher, but data entry can be automated, and error in data entry can be lower
- Paper surveys are often preferred by people, and response rates can be higher
- Responses themselves can be faster with electronic surveys

Paper or online surveys?

- Responses from paper and electronic surveys are considered to be equally valid
- In sensitive topical areas, people may feel more open about disclosing personal information using electronic surveys
- Either paper or electronic, you may need to make sure that respondents/participants are made aware of their rights in the research (informed consent)

Testing the survey tool

- You must test both:
  - The survey interface/structure
  - The survey questions themselves
- In an ideal world, you should:
  - Have experts review the survey tool
  - Have interviews with potential respondents to evaluate content/motivational qualities
  - Have a pilot study of the survey tool and implementation procedures

Response rate

- One of the main challenges of a survey is getting a high response rate
- Incentives for survey respondents are typically lower, with a higher number of participants needed, than with other forms of research
- Because they are often self-administered, motivation is a factor

Response rate

- Techniques for improving the response rate:
  - Send an introductory letter from a respected member of the respondent community
  - Increase the ease of returning a survey (include a self-addressed envelope with postage paid)
  - Send out reminders
  - Send a replacement survey 2-4 weeks after the initial one was sent out
  - Make a final contact using a different form of communication

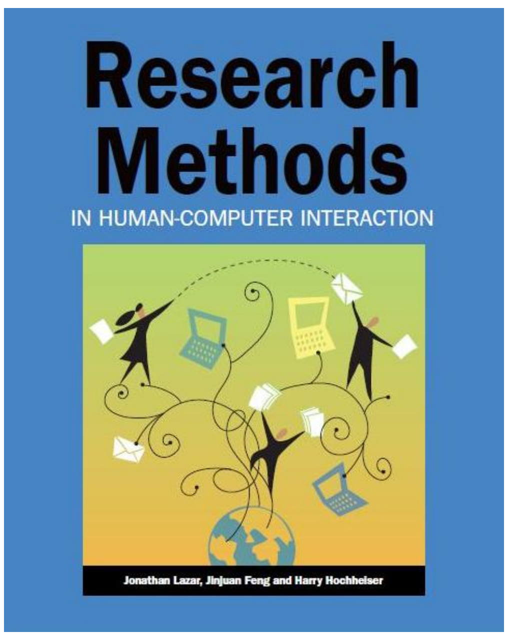
Data analysis

- Separate the quantitative and qualitative data
- “Clean” the data, looking for:
  - Invalid responses
  - Repeats (the same person submitted the same survey twice)
  - Incomplete responses
- Descriptive and inferential statistics

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 6- Diaries

What is a Diary?

- A diary is a document created by an individual who maintains regular recordings about events in their life, at the time that those events occur
- Diaries are useful for recording information that is fluid and changes over time, such as user mood
- If recall after-the-fact will be hard, inaccurate, or subject to biases, diaries might be a useful data collection method

What is a Diary?

- Many people keep blogs, status updates using Facebook, or “tweets” on Twitter
- These are all informal diaries, because entries are made as they occur, in real-time (although there are no stated research purposes)
- Diaries have been adopted, from sociology and history, for use in human-computer interaction research

What is a diary?

- User mood and feelings, such as frustration, anger, and pain, can be best understood using a diary
- Diaries can be time-focused or not
- A time diary is when there is a focus on how time is used, and entries need to be made on a regular basis (with a specified dimension of time)
- Time diaries are useful, because people often remember time inaccurately

Why use diaries in HCI?

- Diaries fill the gap between observation in naturalistic settings and fixed laboratory settings, and surveys
- Users may have different reactions when being observed, and observers may not always understand what is going on
- If interested in collecting data that is fluid and changes over time (rather than factual data), surveys can lead to biased data due to biases in recall
- Multi-method research is often the best approach

Why use diaries in HCI?

- Diaries allow for more collecting more detailed research than surveys
- Diaries are good at understanding not only what users are doing, but WHY they are doing it
- Diaries ask users about themselves, their perceptions, and their mood?
- Time is an important dimension, because asking users to recall after-the-fact how much time they spent or wasted will lead to inaccurate answers

Why use diaries in HCI?

- Diaries are good for recording user-defined incidents
  - When users intended to perform a task, but decided not to do so
  - When users feel that they have learned something new
- Diaries are also good at researching situations where users move around and don’t stay in one place
  - Mobile phones, GPS devices, hand-held tech

Challenges with diaries

- Users sometimes are not introspective and are not even aware of the specifics of what they are doing and therefore may have trouble recording it in a diary entry
- Users may not follow through and record (via paper or electronic) a sufficient number of entries
- Time recording may still be less accurate for time diaries than for controlled laboratory setting or automated data collection

Challenges with diaries

- Generally harder to recruit users for a diary study than for something less intrusive, like a survey
- Since data is both qualitative and quantitative, data analysis may take a long time
- Hard to strike a balance between a frequent-enough series of diary entries, and infringement on daily activities (user participation may then trail off)

Diaries for future technology

- Diaries can be used to investigate the use of technology that exists at multiple stages:
  - Technology that doesn’t exist yet but could (where researchers investigate communication or information usage patterns, separate from technology)
  - Technology that exists but needs to be improved (how people use existing technology)
  - Prototypes of new technology that need to be evaluated

Participants for a diary study

- Determine in advance who appropriate participants/users are
  - Demographic, education, computer experience
- Try to get a representative group of participants, but it's more important to have users who can provide useful insight
- Potential diarists must understand the purpose, be motivated and use any required technology for diary entries

Participants for a diary study

- The diary study would be structured so that it yields useful data, without imposing an unreasonable burden on users
- The diary study should not negatively impact on employment, health, or relationships
- Participants should be paid for taking part in the diary study
- Participants need to be informed of their rights, including to remain anonymous

Types of Diaries

- Two main types of Diaries:
- Feedback diary- the data recorded in the diary is itself the purpose of the research
  - Users make entries when a certain event or threshold occurs, or on a stated time basis
- Elicitation diary- the users record only basic information about important events occurring in their day
  - These data points are used as prompts for expansion at a later time

Feedback diary

- Feedback diaries can be structured or unstructured
  - Likert scales, checkboxes, and time recording
  - How did you respond? How did you feel?
  - How did you think X could be improved?
  - OR personal reflection

Elicitation diary

- Data points recorded in an elicitation diary are quick and simple
  - Short snippets of text
  - Short audio records
  - Short video clips or pictures
  - The goal is not interrupt what users were doing
- Then, later (in an interview, on a web site, or other format), users expand on their entries, describing in more detail

Data collection

- How will the diaries be recorded?
  - Paper?
  - Electronic?
  - Voice?
  - Smart phones?
- Often now, technology is being used to record diary entries
  - But you may not want to use the technology that is focus of the diary study as the only method for recording data

When to record an entry?

- Participants should be given clear guidance on when to perform an entry in the diary
  - What activities are of interest?
  - What events, incidents, or feelings should result in a diary entry?
  - How often should diary entries be made?
  - How detailed should the entries be?
- Make sure NOT to pay participants based on the number of diary entries

When to record an entry?

- Two weeks is often an appropriate length of time for a diary study
- If diary reports are turned in during the study period, researchers should monitor the incoming reports, check on who is not reporting diary entries, or if the entries are not providing useful data
- Reminders and feedback can be sent during the period of the diary study

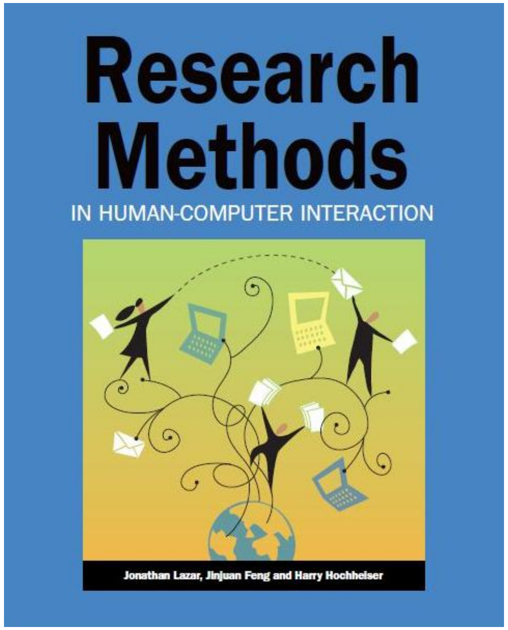
Analysis of diaries

- Transfer all records to an easy-to-analyze electronic format
- Do statistical analysis on quantitative data
- Prepare and examine qualitative data, potentially do a content analysis
- With qualitative data, you can contact the participants after the fact to see if your interpretations are correct

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 7
Case Studies

One or Many?

- More often seen as better
  - Especially for controlled studies looking for statistically significant results, but..
- Large groups not necessarily available, and
- Going “in-deep” with small fewer can provide more insight

Case Study

- in-depth study of a specific instance (or a small number of instances) within a specific real-life context.

- Build understanding, generate hypotheses, document behavior that would otherwise be hard to document
- Time and labor intensive, but potentially informative

Observing Sara: a case study of a case study

- Understand case studies by looking at a published study in the research literature
- Shinohara, K. and Tenenberg, J. (2007) Observing Sara: A case study of a blind person's interactions with technology. Proceedings of the ACM Conference on Assistive Technology (ASSETS), 171–178.
-

Observing Sara: Goals/Methods

- Goal: develop in-depth understanding of how a blind person uses assistive technologies
- Method: In-depth and in-context interviews and observations with one person
  - Digital and physical objects
  - Specific tasks
- Results: insights that might be used to improve designs.

What is a case study?

- Four key aspects:

1. In-depth examination of a small number of cases
2. Examination in context
3. Multiple Data Sources
4. Emphasis on Qualitative Data and analysis

Small number of cases

- Sara: 12 hours of interviews, transcriptions, etc.
  - Couldn't easily do this for 20-30 people
- Case study: small numbers 1,2,3
- Don't worry about statistical significance

In Context

- Lab studies don't reflect how people work and live
  - They aren't “ecologically valid”
- Real environments
- Real users
- Real tasks
- .. provide more realistic understanding of how technology is actually used.

Multiple Data Sources

- Combine observations from
  - Artifacts
  - Observations
  - Interviews
  - Documents, etc.
- Data Triangulation: Using supporting evidence from multiple sources to increase confidence in observations
- But, they might contradict…dig deeper!

Focus Qualitative Data

- Questions are usually broad and open-ended.
  - “How does Sara use technology?”, not
  - “How fast can she complete as specific task”
- Answers (data) often (but not always) descriptive, not quantitative.
- See Chapter 11 for Qualitative Data Analysis

Goals of Case Studies

- **Exploration** of novel problems or situations, possibly to inform new designs
- **Explanation:** develop models to understand technology use
- **Description:** document a context that led to a design
- **Demonstration:** show how a tool was successfully used.

Types of Case Study

- Intrinsic: Specific to a given context
- Instrumental: try to generate general insights

- Caveat: be cautious when generalizing
  - Case studies can provide insight that might apply more broadly...
  - But, they might not...

Single or Multiple Cases?

- Single case ,
- Multiple case (n>1) possible

- Similarities between cases might support generalization

- But,can't generalize complete based on 2 or 3 cases

Literal vs. Theoretical Replication

- **Literal Replication:** Repetition of two or more similar cases designed to demonstrate consistency
  - Another blind college student alongside Sara
- **Theoretical Replication:**
  - Cases that differ in some crucial way
  - Differences between cases might explain differences in observations
  - Blind college student vs. blind executive..

Multiple cases as confidence boosters

- What if your single case is idiosyncratic?
- Less likely that multiple cases will be unrepresentative all in the same way

- Less chance of cherry-picking
- But, generalization still might not be appropriate.

Embedded vs. Holistic

- **Unit of analysis:** granularity of discussion
- Embedded: multiple units of analysis
  - Sara - study of 12 separate tasks.
- Holistic: one unit
  - Sara as an individual
- Holistic study of a large organization vs. embedded study - multiple departments

Components of a Case Study Design

- Questions
  - What are you interested in understanding?
- Hypotheses or propositions
  - Statements of what you expect to find
- Units of analysis
  - Granularity of what you expect to focus on
- Data analysis plan
  - How will you interpret data?

Observing Sara: Goals

- Goals
  - understanding how a blind person might use a variety of assistive technologies to accomplish tasks and to recover from task failures using workarounds.

Propositions

- Expected to see...
- Common types of failure and workaround strategy.
- Influence of choice of implementing features in hardware or software on user interaction, including failures and response to those failures.

Choosing Cases

- Often no choice
  - Mandate of project
  - Convenience
- When making a choice
  - Identify participants who are committed
  - Maximize convenience for participants
  - Try to be representative of complete class of users, if possible
    - Greater External Validity

Alternative Strategies

- “Edge” cases
  - Situations that are innovative, highly capable, or otherwise not representative
  - But, in some way interesting
- Critical Cases
  - Particularly distinctive wit respect to problem under consideration

Choosing between cases

- Multiple possibilities – which to pick?
- Conduct a screening survey
  - See Chapter 5 on surveys
- Questions designed to assess alignment with your goals and commitment to project
- Screening surveys can stand on their own to provide valuable data

Collecting Data

- Use multiple sources
  - Documentation
  - Archival Records
  - Interviews
  - Observations
  - Artifacts
  - Computational data (See Chapter 12)

Choosing Data Sources

- Documentation and archives good for studying past history
- Interviews – perceptions, concerns, needs, reactions
- Observation – what do people do?
- Artifacts – How do people bridge gap between physical and computational worlds

More on Data Collection

- Goals guide data sources and specific questions
- Be clear about what you want to learn, and how data sources will help you learn it
- If only one data source will address a specific question, add other data sources

Logistics of data collection

- Each data sources is a mini-experiment
  - Develop questions, procedures, etc.
- Protocol
  - Guide to reproduction establishes reliability
- Draft of write up
  - Describe what you can, and fill in observation, analysis, and conclusion later.

Case Study Protocol

- Introduction – questions and hypotheses
- Details of data collection
- Case selection criteria
- Contact information for relevant individuals
- Plans for each data source
- Specific questions and methods
- Outline of report

Pilot Case Study

- As with other research methods, a dry run can help debug
- Not always possible, particularly if there are no other suitable cases
- Pilot selected materials if complete pilot is not possible

Analysis and Interpretation

- Plan analysis early
- Qualitative techniques – Chapter 11
- Triangulate
- Chains of analysis – link conclusions and inferences back to original observations
- Go from units of analyses to whole case, or to all cases
- Triangulate

Matrix Displays

- Units of analysis in rows, aspects of analysis in columns

Object/Task	Description	Intentions/ Goals	Limitation (what exactly is going on?)	Explanation (why does limitation happen?)	Workaround (how is limitation overcome?)
Navigating with JAWS	Incorrect key strokes may cause her to lose her bearings.	Execute an action through specific hotkeys.	JAWS is doing something other than the intended action.	Other keys may have been hit by mistake.	Keeps trying different key combinations to execute intended action.
Searching for a CD to play.	Linearly searches all CDs.	To select a specific CD to listen to.	She cannot quickly read CD covers.	CD jewel cases not easily identifiable. Labels do not fit on case spines.	Labeled CDs, mentally organized by preference, read one at a time.

Pattern Matching

- Match observations to predictions
- Researchers believed that Sara would use a range of approaches and workarounds
- Description of tasks in terms of situations that led to difficulties & characteristics of workarounds lead to matches between observations and theory

Developing Models

- Combine observations to develop model or framework
- Higher-level patterns
- Common concerns
- Recurring ideas
- Sara: criteria for technologies
  - Efficiency, portability, distinguishability, suitability for socially appropriate use
- Quantitative data when available

Writing up the Study

- Descriptive – constructive a narrative
- Start early
- Present theories, data, methodologies, analytic steps, and models
- Summaries followed by analysis, or interspersed
- Thematically or chronologically organized
- Tell the story – quotes, specific incidents

More on the writeup

- Rival explanations: show why your observations support your model better than alternatives
- Be cautious in making overly general claims
- Share draft with participants
  - They will tell you if you have details wrong
- Protect participant concerns
  - Privacy/anonymity when possible

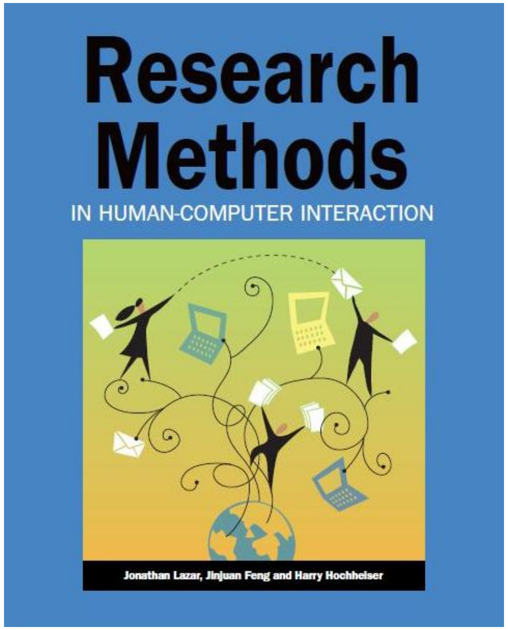
Informal Case Studies

- Need "quick-and-dirty" feedback for needs assessment or usability evaluation
- Document successful use of a tool
- Full-blown case study may be overly complicated
- Might forego theoretical background or analytic framework
- But, try to be as rigorous as possible.
- Intermediate steps in larger studies

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 8- Interviews & Focus Groups

Ask the users

- Direct conversations as tools for data collection
  – Understand requirements, needs, problems

- Interviews – one at a time
- Focus groups – many

Pros & Cons of Interviews/Focus Groups

- Pros
  - Go deep
  - Flexible
- Cons
  - Skill to manage
  - Time and resource intensive
  - Recall problems

Applications of Interviews

- Initial exploration
- Requirements elicitation
- Evaluation and Subjective Reactions

Who to Interview

- Beyond users – Stakeholders
  - Anyone who may be affected by the use of a system
- Interview representatives of different groups from different perspectives.
- *Key informants:* particularly useful/ informative individuals

Types of Interviews

Fully Structured

Semi-Structured

Unstructured

Less structure: greater difficulty in conducting and interpreting interview

But

More opportunity for insight

Comparing the types

- Fully structured – Orally-presented survey
  - Stick with the script.
  - Good for comparison across individuals

- Semi-structured – pre-specified questions serve as starting point for discussion. Digression is ok.

Comparing the types

- Unstructured – initial question, possible list of topics, but no pre-defined script
- Less structure good for open ended exploration

Focused & Contextual Interviews

- Go beyond asking questions
- Ask participant to
  - Demonstrate use of technology
  - Show artifacts (papers, photos, etc.)
  - React to “probes” - props or prototypes designed to elicit reaction

Interviews vs. Focus Groups

- Interviews take time –
  - 1 hour or more/response
  - Several hours for analyzing notes
- Focus groups
  - More people in less time
  - Up to 8-12 people at once.

Focus Groups: Pros & Cons

- Pros
  - Broad range of viewpoints and insights
  - Each group will likely have at least one person who will stimulate others to talk
- Cons
  - Hard to manage group dynamics
  - Generally can't be fully structured
  - May need to ask fewer questions
  - Selection can be challenging

Closed-ended Questions

- Closed-ended
  - “On a scale of 1-10, 10 being best, how did you like the web page?”

- Easy to analyze, but may not be informative.

Yes/no questions

- Remember, the goal is to get interviewees to give you useful insight
- Simple questions get simple answers
- "Did you like the home page?"
  - You'll get a one-word answer

Open-ended questions

- “What did you think about the web page?”
- Invite elaboration, discussion.
- Ask users to complete a sentence
  - “My favorite web browser feature is...”
- Conceptual mapping
  - Draw pictures or layouts to describe understanding of a situation or problem

Other Guidelines

- Simple questions – no jargon
- Avoid compound questions with multiple parts
  – Not "“What were the strengths and weaknesses of the menu layout and the toolbar?”
  – Ask two separate questions instead.
- Avoid judgmental phrasing or tone
  – Possible bias

Questions in un- or semi-structured interviews

- Often, questions are generated in response to participant comments
- Can be hard to do this well.
- Start with structured interviews
  - Get a few under your belt before moving to unstructured.

Preparing for Interviews

- Pilot test – w/ colleagues and participants
  - Good for logistics and for confidence
- Write an interview guide listing what to do and when
  - Particularly good if multiple researchers are involved
- Logistical backups
  - Batteries for laptop, audio recorder, extra paper, etc.

Notes

- Audio and video recordings are fine, but
- Paper notes are still important
  - Record insights, non-verbal responses, etc.
  - Try to record what you can, but
  - Don't do so at the expense of listening
- Summarize written notes as soon as possible after the interview

  before you forget...

Recordings

- Complete, but expensive
- Transcription can take many hours.
- Video is tricky, but gets useful information
- Consider audio + still pictures
- Respect privacy and anonymity
- Have a consistent policy for comments made after the notebook is away and the recorder is off.
  – Ok to restart, but be consistent about it.

During the Interview

- You're the Host: Build Rapport
  - Be friendly, respectful, nonjudgmental
  - Listen carefully
- Outline
  - Briefly introduce research goals
  - Complete paperwork (informed consent)
  - Simple questions first, hard questions later

During the Interview, cont.

- Be flexible
  - If your interview is not fully structured
- But, try to keep things on track
- Explain why you are asking each question
- Define terms, avoid jargon
- Ask for clarification

Read between the lines…

- Is the interviewee telling you what they think you want to hear?

- If so, make a note of it
- Might want to downplay in interpretation

Challenges of focus groups

- Manage the room. Be prepared to deal with
  - Digressions
  - Arguments
- Give everyone a chance to talk
  - Address them directly
    - “Joan, what do you think about...?”

Promoting Discussion

- What if they won't talk?
- Fully-structured – not much to do
- Otherwise
  - Rephrase questions
  - Dig deeper into specifics
- Use props and probes to stimulate feedback
- Focus groups – ask for dissenting or concurring feedback

Closing it out: Debriefing

- Ask for any final comments
- Provide more detail about research goals
- Brief summary of findings
- Turn off recording devices
  - Interviewees might make additional useful comments
- Say “thanks”!
- Reflect and summarize notes immediately

Telephone or online

- Phone, web chat, email, conference calls
- Pros
  - Easy, inexpensive
  - Reach more people with less effort
- Cons
  - Lack of face-to-face contact
  - Fewer non-verbal cues
  - Pacing can be harder

Data Analysis

- Do it as soon as possible
- Avoid “cherry-picking”
- Fully-structured, closed-ended: tabulate answers
- Open-ended questions require coding
  - Transcribe audio?
  - Written notes?

Qualitative Analysis

- Content analysis – frequency of terms, patterns in the notes
- Categorization
  - Affinity Diagrams
- Critical-incident analysis

- Multiple analyses can increase validity

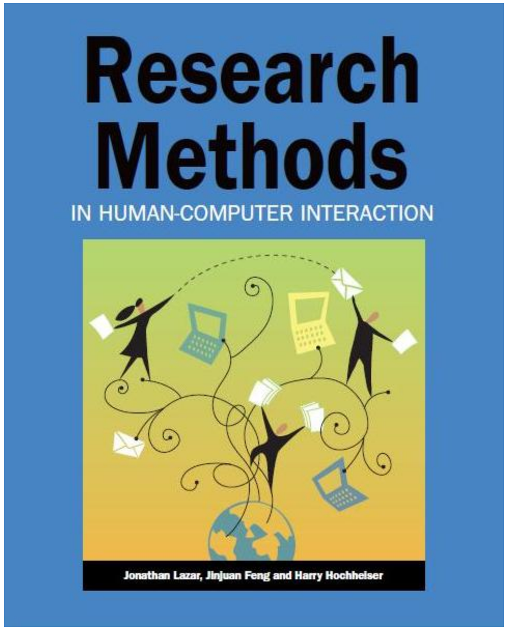
Reporting Results

- Be as clear as possible
  - "7 out of 10", instead of "most"
- Use quotes or paraphrases from respondents
  - But don't use participant name
  - use identifiers (Subject 3) or pseudonyms

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 9
Ethnography

Challenge: Research in Unfamiliar Places

- How do you design a tool for use in situations that are completely new to you?

- Build an information systems for intensive care units …
  - But you're not a health-care professional
- … in a foreign country
  - That you've never been to

Lack of familiarity presents a challenge

- You don’t know the work
  - How do people communicate?
  - What information do they need?
  - Who’s in charge?
- You don’t know the culture
  - Broader societal/social context
- You don’t know the language
  - Idioms and subtexts

Shortcomings of other approaches

- Surveys, interviews require some idea of
  - what you’re trying to learn
  - which questions to ask
  - How to ask them

- Results may be hard to interpret
  - Another disadvantage of not knowing the culture

In-depth, In context observation

- Go to site
- Work closely with someone who can show you around
- Ask questions for basic understanding
- Observe workers in action, talk with some in detail
  - “Shadow” – follow them around as they work
- Build understanding

In-depth, In context observation

- Build models of work
- Refine in discussion with users
- Build requirements and elements of design
- Refine with users
- Travel to another hospital to determine whether or not findings generalize

Ethnography

- Combination of multiple forms of data collection
  - Observations
  - interviews
- Participation is key
- Deep immersion to develop deep understanding

- Richer and more detailed than other methods, but ..
- Expensive and challenging

Ethnography: Background

- "The art and science of describing a human group – its institutions, interpersonal behaviors, material productions, and beliefs" (Angrosino, 2007)
- Roots in anthropological studies of non-Western cultures
  - Limited interactions and interviews insufficient
  - Step out of role of scientific observer, engage directly with people in daily lives
  - Years living in traditional villages

Conceptual Basis for Ethnography

- True understanding of complex human practices requires in-depth, engaged study
- “to gain an understanding of a world that you know little about, you must encounter it firsthand” (Blomberg and Burrell, 2007, p. 967).
- Participation is key

Ethnography vs. Hypothesis-driven Research

- Ethnography is inductive
  - Data -> Patterns -> Theories
- Not hypothesis driven
- No controls – every study is unique
- Similar to case studies (Chapter 7)
  - Multiple sources of data, triangulation

Ethnography vs. Case Studies

- Similarities
  - Multiple sources of data
  - Triangulation
  - Time-intensive
  - Personal
  - In-context
- Difference
  - Ethnography is generally not theory-driven
    - (but this is a topic of debate)

Ethnography as Fluid Research

- Deep engagement with subjects
  - More so than with case studies
- Data collection “in the movement”
  - Ordinary conversations and events can become data collection opportunities

How does this relate to HCI?

- HCI researchers rarely, if ever, spend long periods of time living in traditional villages
- But, we do build technologies for communication and collaboration in complex environments with deep cultures
  - Workplaces, hospitals, schools, homes
- Need to develop deep understanding of how participants work and communicate

Ethnography in HCI

- Suchman did some of the earlier work on ethnography in HCI
- Situated action—all action is a product of the context from which it is taken
- The human, social and organizational aspects of information systems development are the ones most critical to ensuring the success of a project
- But ethnography doesn't need to just be focused on building systems—ethnography is focused on understanding the context!

Generalization? No.

- Many research methods focus on reducing research to understand portions of it, with the hope of being able to generalize
- Ethnography is the opposite—understanding a context of individuals in groups, their processes and norms, at a specific point in time, without every hoping to generalize

Example: Communication and Multitasking

(Su and Mark, 2008)

- “Shadow” 19 workers at a large US corporation
- Note all activities at desk
- Follow them around
- 550 hours of data, 13,000 events
- Analyze and code to understand who workers talk to, how they switch tasks
- Find: coordinating with multiple people is stressful and difficult
- Systems might be designed to reduce overhead

Participatory Design

- Participatory design (PD) is the process of using ethnographic approaches with the end goal of building an information system
- PD- participation of users at all stages of design: requirements, prototyping, and eventual system design, used when:
  - User tasks are not well-understood
  - Users themselves are not well-understood (often people with impairments)
  - High-risk life-critical information systems

Conducting ethnographic research

- Working "in the wild"
  - Homes, workplaces, schools, etc.
- Extended time periods
- Juggle goals
  - Understanding how to navigate unfamiliar world
  - Collecting data
- Is this for you?
- Be sensitive to background and bias

Selecting a site or group?

- There may be no alternative – if you are building a system for a specific customer
- Do you want to find a “representative” situation? Or an extreme one?
  - But all ethnographies are different
- Convenience is often a consideration
- Preliminary interactions to evaluate suitability of site and build trust
- Consider impact on subjects
  - Benefits should outweigh the costs

Barriers to your involvement

- Are there privacy laws that limit access?
  - Hospitals and patient data
  - Schools and pupil performance data
  - Financial information
  - Military information
- You may need to undergo background and security checks, fingerprinting, or sexual harassment training
- You may need to sign confidentiality or other legal agreements

Choosing a role in ethnography

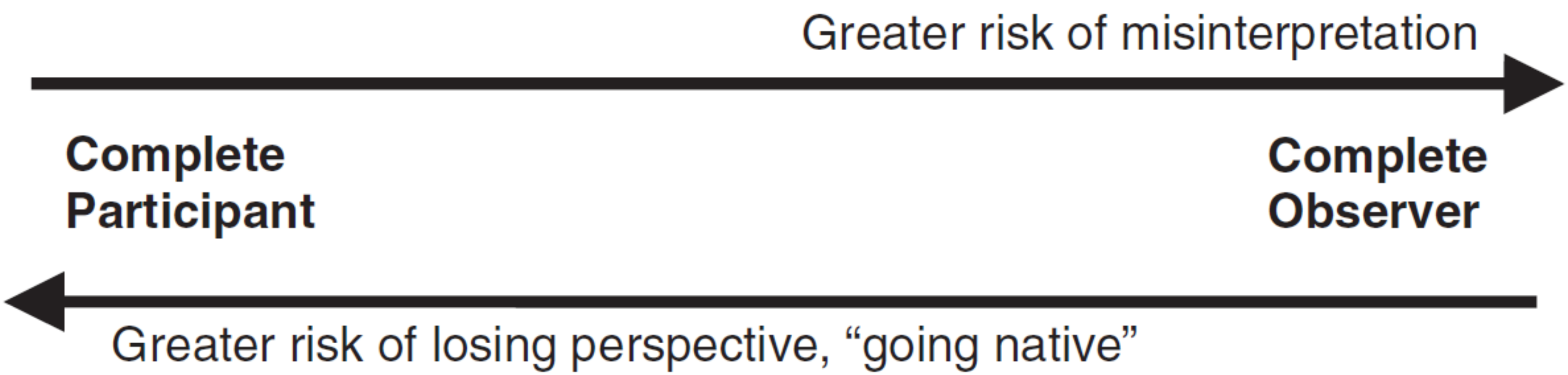
- What does “participate” mean?
  - Complete participant – hides role as researcher
  - Very challenging
  - Rare in HCI work
- Participant-Observer
- Observer-Participant
- Complete Observer

Is observation like usability testing?

- Usability testing (chap. 10) uses primarily observation, not participation, in finding and fixing flaws in a specific interface
- Usability testing is short-term, focuses only on a few individuals, and does NOT focus on groups, context, or human dynamics
- Ethnography=understanding people, context, organizations, and problems
- Usability testing=evaluating diff. solutions

Tradeoffs in Choice of roles

- Fidelity of data vs. risk of losing perspective
  - “going native”



Working with a group you know

- Should you do an ethnography of a group that you are already a member of?
- You would already have access, relationships, trust, and knowledge
- BUT, you may already have pre-informed opinions, biases, and existing relationships that would limit your ability to be neutral and truly understand
- The reality is, your membership in a group may lead to interesting research opportunities

Building Relationships

- Successful research will require trust of the group
- Group members may vary in interest and enthusiasm
  - Workplace studies – users may fear for their jobs
- Work to show that you can be trusted
  - Be helpful
  - Explain research
  - Respect needs and goals of group members

Making Contact

- Start with a small number of well-chosen group members, who can introduce you to others
- Beware of *stranger-handlers* and *deviants*
  - They may mislead you
- Good initial contacts are well liked, respected, observant
- Don't be too closely-tied to any individuals
  - May bias your observations and complicate communications with others
  - It's like you have already "taken sides"

Interacting with Group Members

- Informants are likely to give you the story as they see it—it's "their truth"
  - Or as they want you to see it
- Use discussions to derive questions, build theories, plan further investigations
  - Confirm or refute comments
- Work to maintain relationships
  - Present yourself as non-threatening

Goals of ethnography

- Ethnographic goal: create interpretation of
  - potentially biased
  - incomplete
  - and contradictory
- data points gathered through interacting with group members
- Your interpretation should help with understanding how the group functions, and triangulation improves accuracy

Ethnographic Interviews

- Part of longer, ongoing relationships
- Start with establishing trust and understanding broad parameters
- May not feel like interviews at all
  - Ask questions as they show you around
- Goal – get people talking
- Opportunistically ask for more detail about interesting comments or situations
- Follow your curiosity
- More formal interviews are also possible

Observation

- Not just “stand back and watch”
- Goal – see situations with “new” eyes
- Record only what you see, don’t interpret
  - Less biased
- Try to broaden scope of observations
  - “Is there anything I’m missing?
- Skill that takes time and practice

Note-Taking

- Record details
  - Time, place, participants (anonymized), context, behaviors, interactions
- In the wild, you may not be able to take notes in real time
  - Can’t write while observing ongoing events
  - Write notes as soon as possible afterwards
- Recordings might be possible, but
  - Obtrusive
  - Hard to transcribe/analyze

More on Notes

- Start by recording lots of detail
- As you build understanding and develop patterns
  - Record observations in terms of patterns
  - Make special note of unfamiliar events for further consideration

Other Data Sources

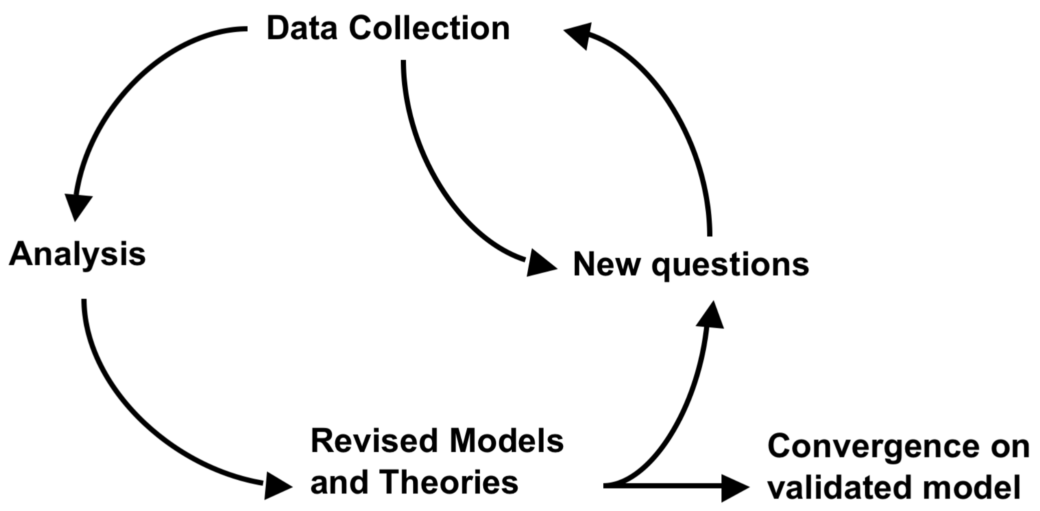
- Documents, artifacts, and archives
- Pictures, letters, e-mails, documents, reports, tools
- You can take your time, re-read them, but understand that they only tell a portion of the story!

Analysis

- Combine qualitative (Chapter 11) and quantitative analysis techniques
- Goal: Informs development of models & description
- Group data into categories and frameworks
- Analysis may identify questions to be pursued in subsequent observations or interviews
- No “cherry-picking”- account for all of your data!

Iterative Process

- Continue until you are
  - Not learning anymore or
  - Out of time/money



Increasing validity

- Ethnography is inherently interpretive
- Why is your model better than any other?
- Compare to alternatives
- Include more viewpoints
  - Multiple informants
  - Multiple observers

Reporting Results

- Describe goals and methods
- Justify choice of groups
- Describe analyses
- Matrices, charts and figures to display data
- Consider rival explanations
- Show participants your report – get their feedback

Examples of HCI research

- Ethnographic methods are often used in HCI in 4 types of settings:
  - Homes
  - Workplaces
  - Educational settings
  - Virtual settings
- Ethnography is also increasingly used for studying mobile devices

Home settings

- County, culture, and religion have a great impact on how technology is used in homes around the world
- Sometimes, researchers discover a relationship between gender issues and technology usage in the home
- Often, ethnography can help identify technologies that are needed in home settings

Workplace settings

- Ethnographic methods for HCI are used most often in workplace settings
  - Examples: Insurance claims handling, department of highways, healthcare settings
- Workplaces which need to be understood can also be outside (e.g. vineyards)
- Important to understand the context of work, the constraints of technology, where technology needs to be introduced, who the potential users are

Educational settings

- School settings can be complex
  - Many activities occurring at once
  - Many constraints from government policy and administrators
- There are major differences in how technology is used for different aged children, different educational approaches, and in different countries
- Often, the school building itself may pose constraints to technology use

Mobile and ubiquitous systems

- On-the-go ethnography may be necessary for technology used outside of a static setting, such as:
  - Automobile drivers using GPS
  - Smartphone users
  - Technology used by firefighters or other emergency responders

Virtual ethnography

- Used when the groups and communities of interest are either primarily or completely online
- Since the group, the community IS online, you are inherently participating as others do, even if you only lurk
- Online identity is easier to change and control than in real-world settings
  - Researchers can define themselves as participants however they want
  - Objectivity can be easier to maintain

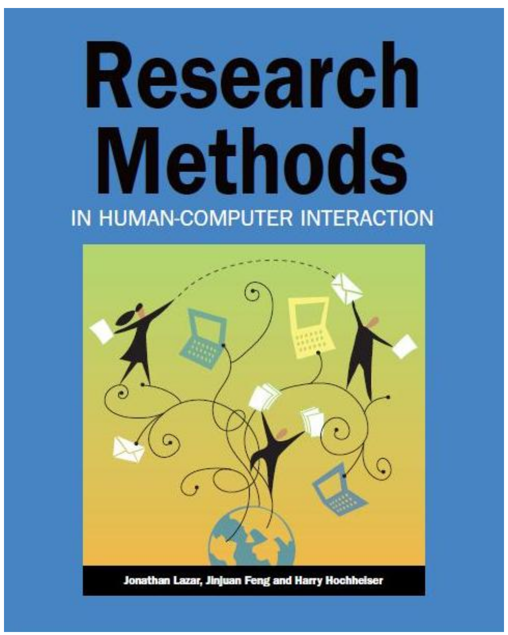
Virtual ethnography

- You can conduct an ethnographic study online without revealing your identity or the fact that you are a researcher
  - But in what situations should you identify yourself as a researcher?
- You might even create multiple identities online, to see how people react to an argument between these identities
- What happens when there is a face-to-face component of the group? How do researchers handle that? Do they let people know their goal?

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 10-
Usability Testing

What is a Usability Testing?

- Representative users performing representative tasks
- The goal is to identify and fix interface-related flaws as early as possible
- This can include:
  - Fully-functional interfaces
  - Paper prototypes
  - Screen mockups that look complete but aren't
  - Somewhat-functional software

What is a Usability Testing?

- We want to learn what problems users have with the interface, and also what is working well
- Identify interface features that are problematic, unclear, could be better
- It’s not about general style preferences, it’s about interface components that impact on performance and frustration
- We are researching the interface, not the user

Scope of usability testing

- Usability testing is a general term that could mean:
  - Having hundreds of users test interfaces, in different treatment groups
  - Sitting next to a total of 3 users and watching as they attempt different tasks
  - The reality is usually somewhere in between
- The best usability testing is the one that actually takes place
  - Be flexible, be reasonable, be practical

Is it traditional research?

- Is usability testing considered to be "research"?
- The historical roots of usability testing and experimental research and ethnography are similar
- Observation, key logging, click-stream analysis, quantitative measurement such as task and time performance, anonymity of participants are all part of usability testing

Is it traditional research?

- The end goals of usability testing and classical research are different!
- Usability testing:
  - An industry-based approach to improving interfaces
  - Little concern for using only one approach or a theoretical point of view
  - Building a successful product using the fewest resources, the fewest risks, in the shortest amount of time

Is it traditional research?

- Practices that would be unacceptable in experimental design, such as modifying the interface after every user, are acceptance in usability testing
- Usability testing is not used to generalize to other interfaces, situations, or user groups
- Many companies keep their usability testing results confidential

A comparison

Classical Research	Usability Testing
Experimental design- isolate and understand specific phenomena, with the goal of generalization to other problems	Find and fix flaws in a specific interface, no goal of generalization
Experimental design- a larger number of participants is required	A small number of participants can be utilized
Ethnography- observe to understand the context of people, groups, and organizations	Observe to understand where in the interface users are having problems.
Ethnography- researcher participation is encouraged	Researcher participation is not encouraged in any way

A comparison

Classical Research	**Usability Testing**
Ethnography-longer-term research method	Short-term research method
Ethnography or experimental design-used to understand problems and/or answer research questions	Used in systems and interface development
Ethnography or experimental design-used at earlier stages, often separate (or only quasi-related) from the interface development process	Typically takes place in later stages, after interfaces (or prototype versions of interfaces) have already been developed
Used for understanding problems	Used for evaluating solutions

Research ABOUT usability testing

- So, usability testing IS different from traditional research
- But, there is also research ABOUT usability testing, with research questions such as:
  - Which methods are most effective
  - Which methods are most cost-effective
  - How many users are optimal

Different approaches

- There are three general categories of usability testing or engineering methods:
  - User-based testing
  - Expert-based testing (inspections)
  - Automated testing (software tools)
- All of these focus on usability, but when we talk about “usability testing” we generally are referring to representative users performing representative tasks

Expert-based testing

- An expert test is a structured inspection of an interface by one or more experts
- Expert-based tests should always come before user-based tests, as experts can identify the obvious flaws first, and get those fixed before users test the interface
- Heuristic review and consistency inspections are very popular
- Guidelines review and cognitive walkthrough also take place

Automated usability testing

- Software applications, or web-based applications, examine a series of interfaces, and compare them to a known set of interface guidelines (e.g. Web Content Accessibility Guidelines)
- The guidelines are often the same as those used in a guidelines review
- Automated tools often require manual checks (where human interpretation is needed), and may offer misleading feedback

User-based testing

- A set of representative users attempting to perform a set of representative tasks
- The earlier that these can take place, the better (because it is more likely to influence actual design)
- User-based testing can take place early on with paper prototypes, or with non-functional screen mock-ups
- The later in development that user-based testing takes place, the less likely that it will impact the actual design

Types of user-based testing

- Formative testing
  - Takes place at an early stage of development, may include paper or other low-fidelity prototypes
  - More of a focus on how the user perceives an interface component, rather than how the user completes a tasks
  - Low-fidelity prototypes are low-cost, can be quickly development, and multiple options tested
  - The focus is on design concepts

Types of user-based testing

- Summative testing
  - Takes place when there is a more formal prototype ready
  - High-level design choices have already been made
  - The goal is to evaluate the effectiveness of specific design choices
  - Also known as high-fidelity prototypes

Types of user-based testing

- Validation testing
  - An interface is compared to a set of benchmarks for other interfaces
  - Can users complete specific tasks within a given timeframe (say, 30 seconds)?
  - Occurs much less frequently than formative or summative testing

Stages of usability testing

- Select and recruit representative users
- Select the setting (home, workplace, usability lab, etc.)
- Decide what tasks users should perform
- Decide what type of data to collect and how to measure
- Before the test session
- During the test session
- After the test session

How many users do you need?

- It’s still a hotly debated question
- Often, people will say, “5 users” even though there is no research basis for this
- You don’t know in advance how many you need, because it depends on how complex an interface is, how many interface flaws exist
- The answer: get as many users as you afford to have, in the timeframe given

How many users do you need?

- Remember, usability testing is about being practical, so it's not about choosing the "right number" of users
- The goal: find major flaws in the interface
- In industry, the key factors are:
  - How much time is left in development
  - How much money is available for testing
  - How many users are available and willing
- Only some of the user groups may be involved

Where does it take place?

- A fixed usability laboratory
  – Two separate rooms, one for the participant, one for the observer, separated by a one-way mirror, with a video/audio feed
- In the user’s workplace or home
  – Observe them by sitting near them, or using portable video equipment
- Remote usability testing, where you observe them over the web or through videoconferencing

The task list

- Unless the usability testing is very early stage where you are looking for impressions on general design choices, you will need to create a task list
- Task lists make users goal-directed
- Tasks need to be clear, unambiguous, and not require further explanation
- Provide a background scenario at the beginning
- Make sure tasks can only be completed by using the interface (not something that is commonly known)
- Do not require the use of user’s personal data

The task list

- Be clear in your instructions:
  - Must the tasks be completed in a certain order? Can users skip around?
  - Do certain tasks require that others are completed first?
  - How should participants note if they are giving up on a task and moving on?
  - Is there a time limit for the entire session?

Measurement

- Often, you are interested in quantitative measurements
  - Task performance
  - Time performance
  - User satisfaction (those three are most common)
  - Time spent recovering from an error
  - Time spent using the help function
  - Number of search engine responses viewed
- Make sure to decide in advance how to collect this data…using keylogging or a stopwatch?

Measurement

- Qualitative observations can also be helpful
- Encourage users to “think aloud”
- More common in formative testing than in summative testing
- People may feel uncomfortable criticizing an interface out loud
- Note that the more users talk, the more their task and time performance may be impacted
- You can also run a “reflection session” where users give feedback AFTER their session

The testing session

- Confirm the time and location with participants beforehand
- Make sure to leave extra time in your schedule
- Test any equipment that you will need
- Participants must be informed of their rights and sign an informed consent form (ch. 14)
- Clarify how participants will be paid for their time

During the testing session

- Decide in advance how to handle it, if participants get stuck and can't move on
- Make sure to remind participants that they are not being tested
- Remind participants that their feedback is important!
- Make sure to get video/audio recordings if possible, and take good news
- Feel free to make changes on the fly (it's not traditional research, it's industrial!)

Organizing the data

- Often, inferential statistics are not the goal
- Just descriptive statistics, and “telling the story”
- Write up a report, short, clear, to the point
- Your goal is have the testing results influence the interface design
- Consider who will be reading the report
- Not all flaws will be fixed…so, summarize and prioritize which are most important!
- Suggest fixes for each flaw
- Never include data that identifies the participants

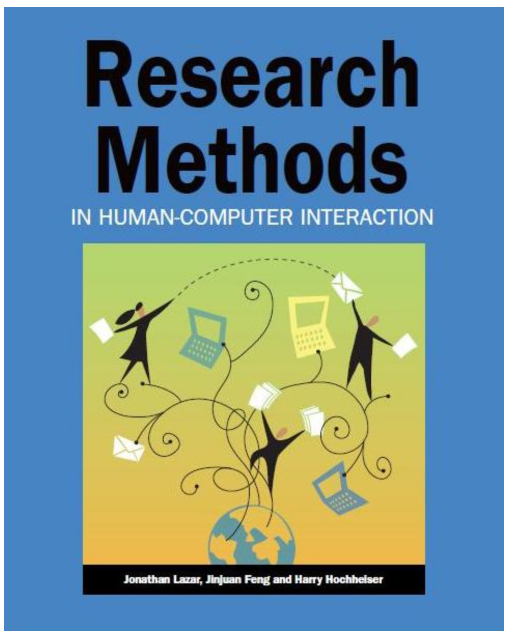
Other variations of user testing

- Technology probes
  - Place a new technology into a real-world setting to see how it’s used and how people react and interact
  - Sort of a hybrid of usability testing and ethnography
- Wizard-of-oz method
  - A non-functional prototype is used for user-based testing, but a human is providing the responses to the user, as if it was the computer providing the responses

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 11- Analyzing Qualitative Data

Overview

- Introduction
- Stages of qualitative analysis
- Grounded theory
- Content analysis
- Analyzing text content
- Analyzing multimedia content

Stages of qualitative analysis

- Identify components of the substance
- Study properties and dimensions of each component
- Understand and make inference about the substance

Grounded theory

- An inductive research method

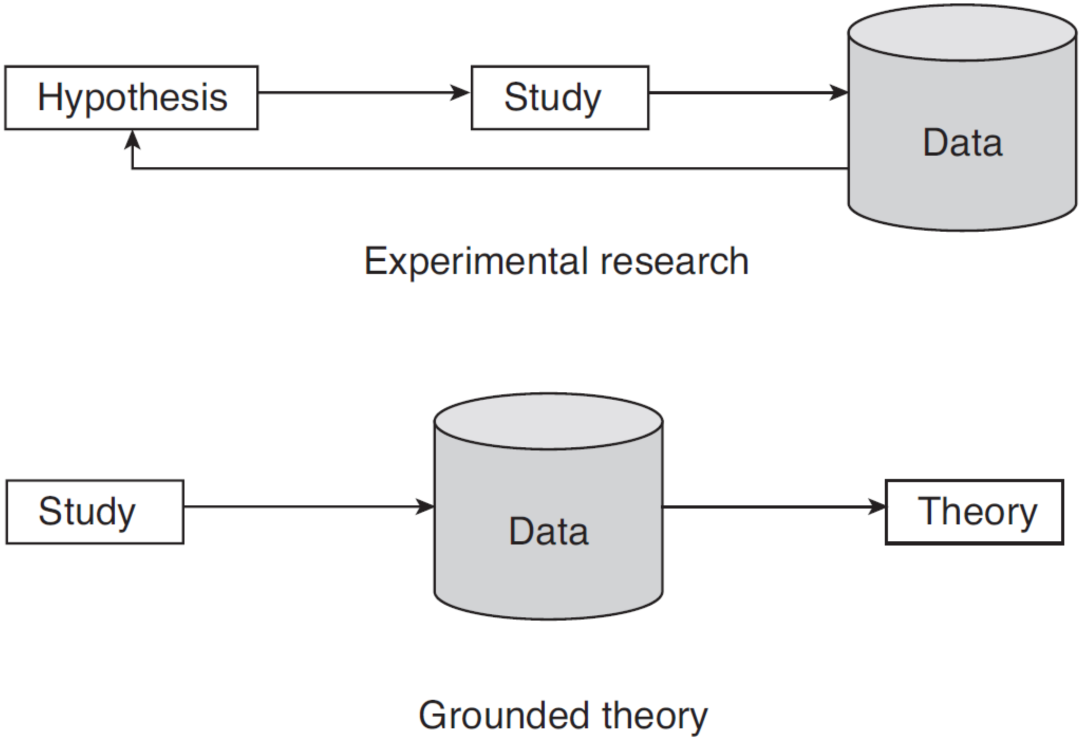


Figure 11.1 Experimental research compared with grounded theory.

Procedures of grounded theory

- open coding
- development of concepts
- grouping concepts into categories
- formation of a theory

Grounded theory

- Advantages
  – a systematic approach to analyzing qualitative, mostly text-based, data,
  – generating theory out of qualitative data that can be backed up by ample evidence of the coding
  – Interplay between data collection and analysis

Grounded theory

- Disadvantages
  - Researcher can be overwhelmed by the details of the data
  - The theory generated is hard to evaluate
  - Findings may be subject to bias

Content analysis

- A more specific view: a systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding
- A broader view: any technique for making inferences by objectively and systematically identifying specified characteristics of messages

Content categories

Category	Sub-category	Examples
Media content	Publications	Books, journals, newspapers, brochures
	Broadcasting	TV programs, radio programs
	Websites	News, web portals, organizational websites, blogs
	Others	Films, music, photos
Audience content	Text	Notes from interviews, focus groups, or observations or diaries or surveys
	Multimedia	Video- or audio-recording of interviews, focus groups, observations, or user studies

Table 11.1 Major categories of content.

Preparing for content analysis

- Define the data set
- Define the population
- Clean up the data
- Understand the context of the data

Analyzing text data

- A priori coding
  - Identify coding categories
  - Coding
  - Reliability check
- Emergent coding
  - Multiple coders identify coding categories based on subset of data
  - Consolidate category list
  - Code a subset of data
  - Reliability check
  - Repeat the process until satisfactory result is met
  - Code the rest of the data

Identify coding categories

- Theoretical framework
- Researcher-denoted concepts
- In-vivo codes
- Building a code structure (nomenclature)

Coding the text

- Look for key items

Statement	Examples
Objectives	Use computers for educational purposes
Actions	Enter a password, chat online
Outcomes	Success or failure, whether the objective is achieved
Consequences	Files unintentionally deleted, a specific application abandoned
Causes	Limited memory, dated equipment
Contexts	User is computer savvy, user works with classified information
Strategies	Avoid specific tasks, multimodal interaction

Table 11.2 What to look for while coding.

Coding the text

- Ask questions about the data
- Making comparisons
  - Between different coding category
  - Between different participant group
  - Between existing data and previous literature
- Using computer software

Ensure high quality analysis

- Validity
  - constructing a multi-faceted argument in favor of your interpretation of the data
  - Constructing a database
  - Data source triangulation
  - Interpretation should account for as much as possible of the data
  - Alternative interpretations may also help

Ensure high quality analysis

- Reliability check
  - Stability
    - also called *intra-coder reliability*
    - examines whether the same coder rates the data in the same way throughout the coding process
  - Reproducibility
    - also called *inter-coder reliability or investigator triangulation*
    - examines whether different coders code the same data in a consistent way

Reliability measures

- Percent agreement

$$\% \text{ Agreement} = \frac{\text{Number of cases coded the same way}}{\text{Total number of cases}}$$

- Cohen's Kappa:

$$K = (Pa - Pc)/(1 - Pc)$$

Reliability check

- Agreement matrix

		Coder 2			
		Physical	Cognitive	Perceptual	Marginal total
Coder 1	Physical	0.26 (0.14)	0.07 (0.08)	0.04 (0.15)	0.37
	Cognitive	0.04 (0.07)	0.12 (0.04)	0.01 (0.07)	0.17
	Perceptual	0.09 (0.18)	0.02 (0.10)	0.35 (0.18)	0.46
	Marginal total	0.39	0.21	0.40	1.00

Table 11.3 The distribution of coded items under each category by two coders (agreement matrix).

Reliability check

- Interpretation of Cohen's Kappa

Interpretation	Kappa range
Poor or slight agreement	$K \leq 0.20$
Fair agreement	$0.20 < K \leq 0.40$
Moderate agreement	$0.40 < K \leq 0.60$
Satisfactory agreement	$0.60 < K \leq 0.80$
Near-perfect agreement	$K > 0.80$

Table 11.4 Interpretation of Cohen's Kappa.

Subjective vs. objective coder

- Subjective coders
  - Knowledge and experience can help interpret the data
  - Less training required
  - May cause inflated reliability
- Objective coders
  - Less likely to cause inflated reliability
  - Lack of knowledge affect the ability to understand the data
  - More training required

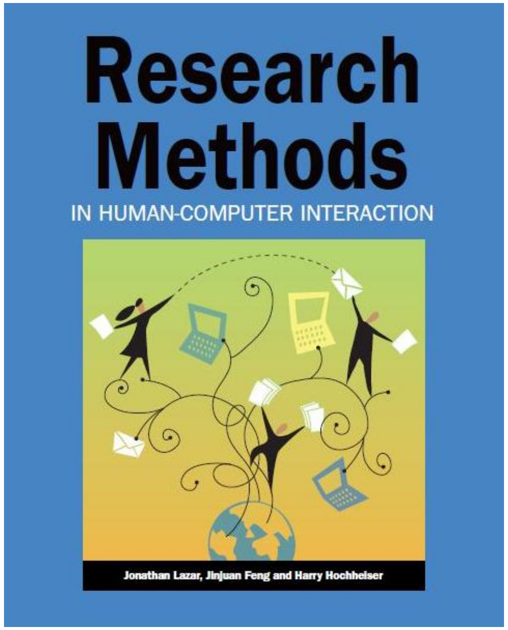
Analyzing multimedia content

- The supporting techniques and methods are less mature compared to text data analysis
- Approaches:
  - Manual analysis:
    - highly labor intensive and time consuming
    - More accurate
  - Completely automated analysis:
    - Faster, less amount of work
    - Highly inaccurate
  - Partially automated approach
    - Combines the advantages of the manual process and the completely automated process

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



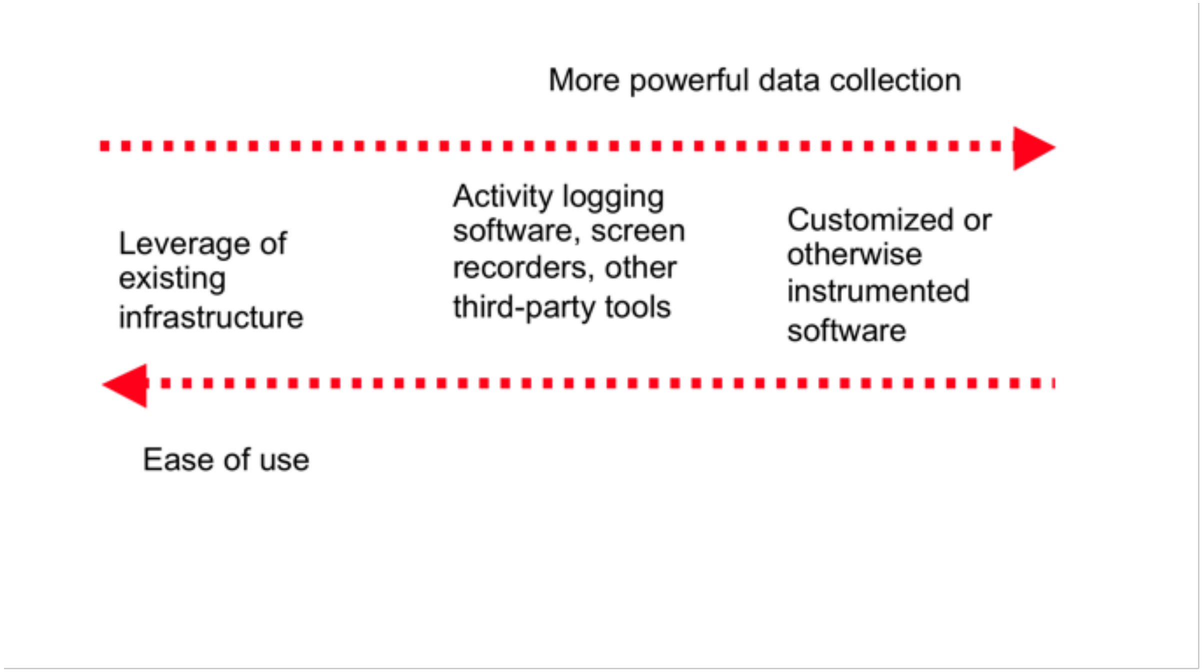
Chapter 12- Automated Data Collection Methods

Data Collection

- Use the Computer!
- Existing software
- Activity-logging tools
- Custom or instrumented tools

Tradeoff

- Power vs. ease of use



Web logs

- Log files indicate
- Which pages were requested
- When
- IP address of request
- Status
- Size (# of bytes)
- Referer – where they came from
- User-Agent.. and more

Examples

- 10.55.10.14 - - [13/Jul/2007:13:42:10 -0400] "GET /homepage/classes/spring07/686/index.html HTTP/1.1" 200 8623

- 10.55.10.14 - - [13/Jul/2007:13:48:32 -0400] "GET /homepage/classes/spring07/686/schedule.html HTTP/1.1" 200 16095

- 10.55.10.14 - - [13/Jul/2007:13:48:33 -0400] "GET /homepage/classes/spring07/686/readings.html HTTP/1.1" 200 14652

Web Usability & Design

- Use Web logs to understand how your site is being used
- Counts of which pages are accessed
- Referers tell which links (internal or external) are being followed
- Infer paths through the site
  - May need to use web cookies
- Use insights to drive design

Web Logs & Empirical Studies

- Time stamps in logs can be used to track time between events
- Present experimental tasks as links on web pages
- Run server locally on machine used to administer tests
  - Avoid network delays
- Turn of caching in browser
  - So each page generates a new request

Capturing event timing

- Selection of some starting link indicates beginning of task
- Selection of final target is end of task
- Elapsed time between those two events is the task time.
- Read server logs manually or use custom programs to extract timing information.

Stored Application Data

- Ongoing computer use creates metadata
- Useful for understanding and analyzing interaction patterns

- What do users do?
- How do they organize data?
- With whom do they communicate, and when?

Stored Data Examples

- File systems
- GUI Desktops
- Email
- Web bookmarks
- Social networking tools

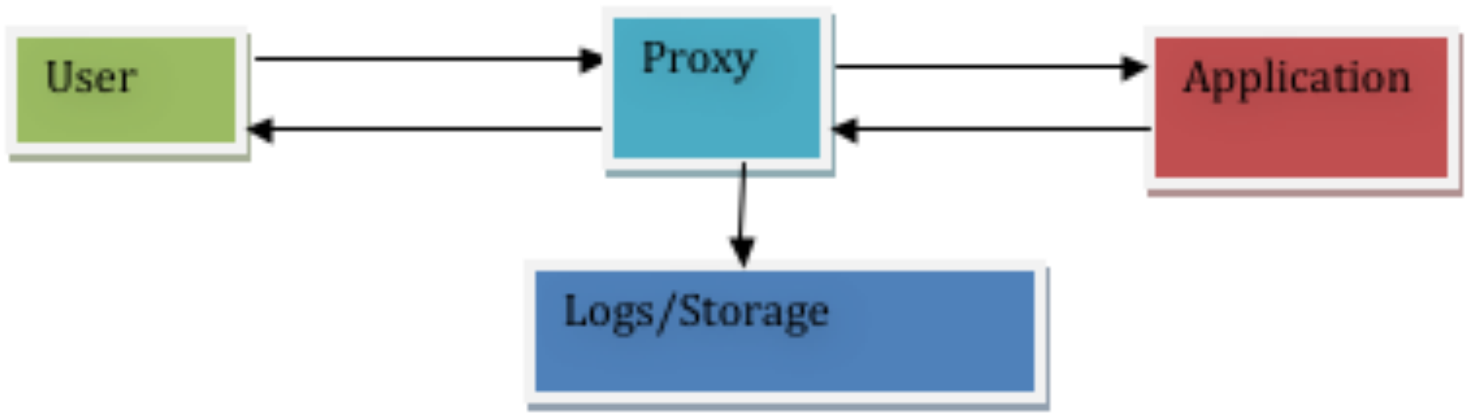
Stored Data – Pros and Cons

- Pros
  - "Ecologically valid" information about real computer use
  - No need to define tasks
- Cons
  - Data extraction may be challenging
  - Potential privacy concerns
  - Not fine-grained

Observation/Recording Software

- Software tools specifically used to collect data
- **Proxies:** programs that intercept and record user actions before passing them on to end programs.
- Store info in log file
- Can be more fine-grained than “built-in” data collection and web logs.

Proxies



Web Proxies

- Widely used for efficiency
- Also for research
- Handle requests from groups of users
- Add additional code for understanding interactions
  - Javascript for mouse movements
- Squid – open source web proxy tool
  - Need appropriate capacity (bandwidth & servers)

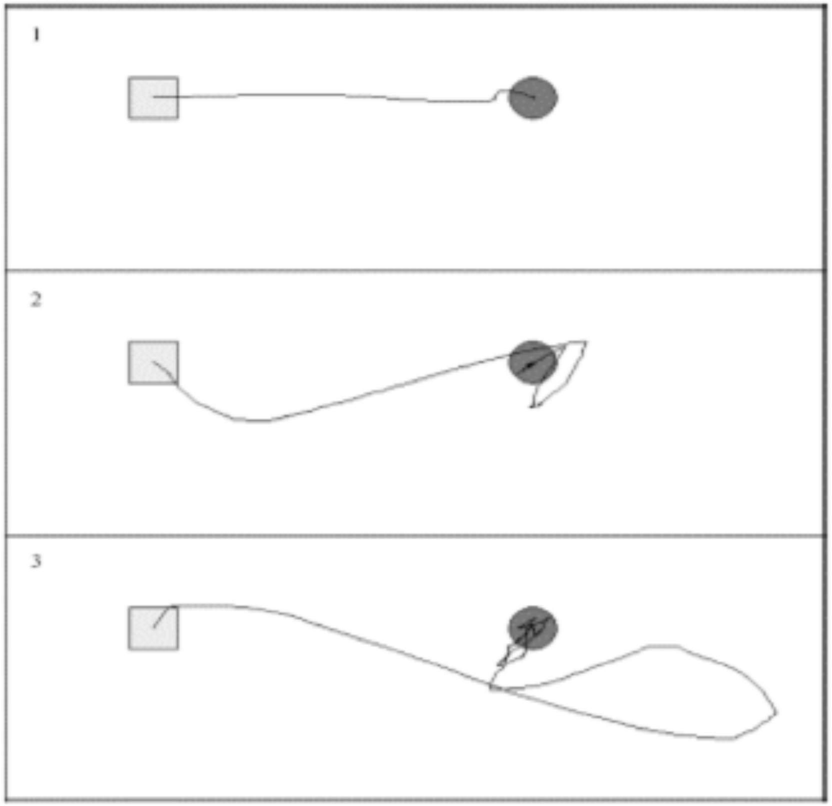
Instrumented Software

- Modifiy software to collect data on its own usage
- Log each mouse movement and menu selection
- Modify existing code
  - Open source? Ingimp (Terry, et al.)
  - Macro packages?
  - Office (Windows 2007 redesign)

Custom-built software

- Write your own software to present tasks and collect relevant data
  - Task completion time
  - errors,etc

Custom-built software: Fitts' Law and Children



- Hourcade, et al. 2004

Keystroke & Activity Loggers

- Local proxies
- Run in the background, recording
  - Mouse movements
  - Keyboard input
  - Window operations
- Spyware? Or legitimate user by employers?
- Can invade privacy
- *Very* fine-grain

Storing &Analyzing Data

- Log files
  - May require custom parsing
  - Relatively easy to use
- Databases
  - Up-front design and population challenges
  - Flexible querying via SQL
- Either way, expect to do some data cleaning

Analysis Goals

- Frequency of access of various resources
- Frequency of actions?
- Patterns?
  - Clicking “save” before “print”
  - Infer “sessions”
  - Data mining
- Visualization?

Hybrid

- Multiple forms of automated collection
  - Proxies and instrumented software
- Automated capture + other approaches
  - Observation
  - Qualitative

Automated Interface Evaluation

- Automated inspection tools
  - Assess compliance with guidelines
  - Frequently used for web accessibility
  - Combine multiple methods?
- Modeling and simulation?

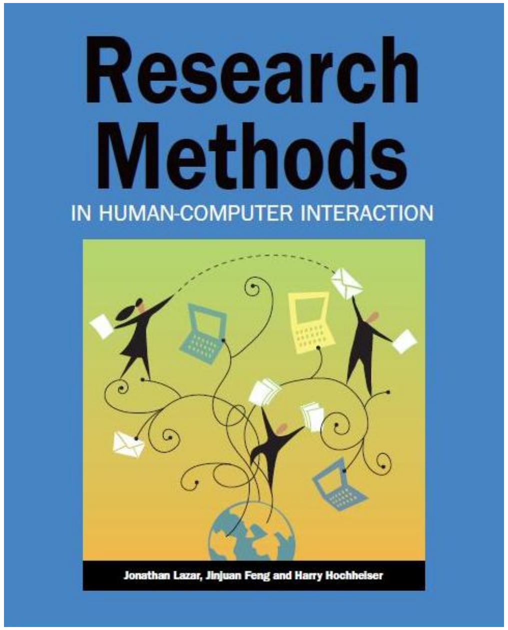
Challenges

- Time scales – from milliseconds to years
  - Individual actions to long-term projects
- Amount of data – granularity
  - Related to specificity of questions
  - How to infer higher-level task from low-level interaction?
- Don't collect data just “because you can”..

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 14- Working with Human Subjects

Working with human subjects

- Finding them – who, how many?
- Treating them right
- Doing it online

Identifying Potential Participants

- Find participants who meet your goals
  - If the tool is for professionals, should you use students as your participants?
  - If your tool should be useful for people of a broad ranges of age, socio-economic status, and education, should you rely solely on undergraduates?

Factors influencing suitability

- Goals
- Background
- Motivation
- Expertise
- Gender
- Age

Many differences may not be important for many studies, but it’s important to ask the questions.

How Many?

- More is better, right?
  - Increased statistical power helps find subtle effects, but
  - Large studies are expensive and long

Controlled Experiments and Empirical Studies

- Need enough participants to produce statistically significant results
- Between-subjects -> more participants
- Use statistical methods to determine how many
- Usually, try for 20 or more.

Fewer Participants for…

- Case Studies

- Ethnographies
  – 1-3 participants

- Usability studies
  – 5 users for heuristic evaluation? (Nielsen ,1994)

Other Factors

- Difficulty of finding participants
  - Some can be hard to find
    - Experts
    - Users with disabilities – see chapter 15
- Time – can’t do large subjects if each one takes too long

- Need more participants for pilot studies

Recruiting Participants

- Experiments, empirical work…
  - Advertisements, community groups, professional organizations
  - Email lists, online groups, social networks
- Ethnographies & long-term case studies
  - Build relationships
- Incentives: cash, pizza, etc..
- Build a database for future consideration
  - With consent

Care and Handling

- Research studies should be designed to protect participants

- Informed consent:
  - Research participants must be provided with the information needed to make an informed consent as to whether or not they will participants

Risks

- Some scientific research can have substantial risk
  - Drug trials, medical devices
- Famous medical and psychological experiments led to current models
- Most HCI experiments are low risk
  - Fatigue? Eye or muscle strain?

Protecting Participants

- Respect for persons
  - Allow individuals to make independent and autonomous decisions
  - Provide necessary information
- Beneficence
  - minimize possible harm
  - maximize potential benefits
- Justice
  - Neither burden nor benefits should limited to certain populations.

Privacy Protection

- Obtain consent for the collection and storage of personally identifying information (PII)
- Limit information collected – only what is necessary
- Identify uses of information
- Limit use, disclosure, and retention
- Securely protect information
- Disclose policies and procedures
- Provide means of addressing concerns
- Be accountable.
- Be careful with photography, audio, and video recording

Informed Consent

- Informed: Participants should understand
  - Reason for conducting study
  - Procedures involved
  - Potential Risks
  - How they can get more information
- Present this information clearly, accessibly, and without jargon

Informed Consent

- Consent - Participation should be
  – Voluntary
  – Free from any implied or implicit coercion
- Decision not to participate should not lead to repercussions or retaliation
  – Withholding of medication
  – Disapproval of researcher
  – Punishment by employer

Participant Rights

- Participation is voluntary
- Participants can choose to stop participating at any time, without penalty
- Participants have the right to be informed of any new information that will affect their participation in the study

Informed Consent Disclosure

- Title and Purpose of study
- Description of Procedures
- Duration
- Risks
- Benefits
- Alternatives to Participation
- Confidentiality
- Costs/Additional Expenses
- Participant’s Rights
- Contact Information
- Supplemental Information

Procedures

- Participants read disclosure (or have it read to them)
- Discuss and answer questions
- Participant (or representative) signs statement

Participant Statement

- Participant states that they
  - Have volunteered to participate
  - Have been informed about tasks and procedures
  - Have had a chance to ask questions and had questions answered
  - Are aware that they can withdraw at any time
  - Consented prior to participation in the study
- Provide copy (with contact information) to participants.

Writing Informed Consent Forms

- Write clearly and carefully
- Pilot test
- Consent must be given by an individual who is capable of understanding
  - Parents, guardians
- Be aware of local or national regulations or laws

IRBS: Institutional Review Boards

- Committees that examine human subjects research for appropriateness
- Protect
  - Participants
  - Researchers
  - Institutions
- You may need to educate IRBs about how HCI research is done

Deceptive Research

- What if informing subjects about your study may bias them?
- Online banking: do people pay attention to security images? (Shechter, et al., 2007)
  - If you tell participants what you're looking at, they'll be sensitized, and results will be biased
  - Instead, use deceit – tell them that your study is intended to "help make online banking better"
- Use this technique carefully and sparingly
- Harder to get IRB approval
- Debrief afterwards

Online Research

- Well-suited for web applications or online tools
- Recruiting
  - May get access to broader pool of participants
  - But, anonymity may make demographics and skills hard to verify
- Surveys, usability evaluations, ethnographies have all been done online

Ethical Concerns

- Studies of online communities – what is the expectation of privacy?
- Informed consent and debriefing are tricky

Closing Thoughts

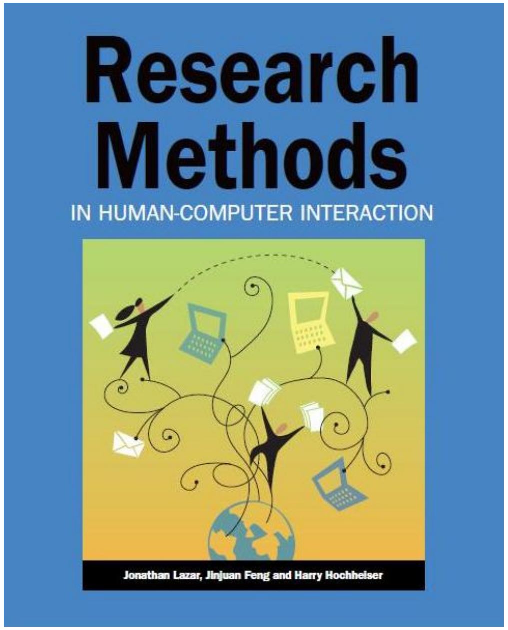
- Participants are crucial – treat them well
- Allow plenty of time
- Food and drink
- Breaks as needed

- Make participation fun!

End-of-chapter

- Summary
- Discussion questions
- Research design exercise

Research Methods in Human-Computer Interaction



Chapter 15- Working with research participants with impairments

Introduction

- It is important to understand the specific concepts, issues, and challenges of doing HCI research with users with impairments
- The label “users with impairments” is artificial, as these are each different groups of individuals, with different challenges, and all they may have in common is that “impairment” label
- Plus, some users may have multiple impairments

Introduction

- We have a much longer history of HCI research for users with perceptual or motor impairments, than we do for users with cognitive impairments
- You can’t just use design guidelines for people with impairments, and you can’t just use proxy users. You must go out and work with people!
- The logistics may be more challenging!
- Do the research anyway!

How many participants?

- One of the biggest challenges is access to users with impairments
- How can you find participants with certain impairments?
- Will they want to take part in your research?
- How far will you need to travel to visit them?

Approaches

- Smaller sample sizes are accepted in research on people with impairments
  - 5-10 users is sufficient
- Distributed research, where the users do the research in their home or office, without researchers present, and send the data collected
- In-depth case studies, with fewer users, taking part in a more intensive way

Proxy users

- Is it OK to use “proxy users” (without impairments) to represent actual users with impairments? In general, NO
  - Do NOT blindfold users, to simulate blindness, or tie hands behind their back, to simulate motor impairment
- Proxy users are only acceptance when the users with impairments are either:
  - Unable to communicate, even w/technology
  - Unable to process information due to their impairment

Multi-population studies

- For users with perceptual or motor impairment, they have the same task goals as users without impairments
  - They just use different input or output devices
- These popular interfaces (news web sites, shopping web sites), must not only work for the general public, but for multiple groups of users with impairment
  - Often, users with visual and hearing impairment, as well as spinal cord injuries and other motor impairments

Multi-population studies

- Because multiple user groups must be able to successfully use these interfaces, they must be evaluated with multiple groups
- Companies will not adopt interface features that improve the interface for people with impairments but degrade the experience for users without impairments
  – This rarely happens, but there is a misperception that it happens often, so you must document that users without impairments are also successful with the interface

Community partners

- Participants won't come to you, and typical recruitment strategies (e.g. putting up signs on campus) won't work
- It's a good idea to form a collaboration with a community-based group
- The group may be interested in your research, and may be able to help you recruit participants

Community partners

- Note, that it must be a partnership, a two-way street
- The community group must also benefit
- You can't just do "drive-by research"
- You should become a part of the community group, go to their activities, and get involved for the long-term
- Make sure to keep the group involved, informed, and consult with them regularly

Contacting participants

- Understand their preferred method of contact
  - Users with SCI may not prefer e-mail
  - Deaf people may not prefer phone calls
  - Blind people have high spam filtering and often won't read e-mails that are sent their e-mail address as a BCC (blind carbon copy)
  - For users with Alzheimer's or Dementia, it may be necessary to contact caregivers

Pilot studies

- Pilot studies are necessary
- Your expectations and perceptions are likely to be very different from reality
- Since you will have access to a limited number of users, you want to make sure that you get your research design right before you start
- One or two pilot users is sufficient

Pilot studies

- Is the documentation accessible for the participants?
  - Users with spinal cord injuries may not be able to handle paper documentation
  - Blind users cannot use printed material (and many of them cannot use Braille, either)
- Physical settings
  - Will the building/office/bathroom be accessible for someone in a wheelchair?

Pilot studies

- If you are sending electronic files to participants beforehand, or if they will be using them on their own computer, be aware that these documents must work on:
  - Different OS (OS X, Windows 7, Win XP)
  - Different text editors (Word, WordPerfect, etc.)
  - Different screen readers (JAWS, Window-Eyes) and other assistive technology
- What personal technology aids will the participants expect to use?

Scheduling

- Many users with impairments do not drive a car
  - They take public transportation, taxis, or use scheduled services
- It often is not possible to make last minute schedule changes
- You should offer to go visit their workplace or home—it should be an option
- Workplaces or homes give a glimpse of the user in their own environment—with the technology that they use, setup to maximize their usage

Scheduling

- If users are coming to your university or lab, find out if:
  - They are bringing a guide dog
  - Your building has Braille labels on doors and elevators
  - The physical facilities are accessible to those in wheelchairs
- Be aware that employment is a point of pride for someone with an impairment, and they are not likely to miss work for your research
  - Offer to do the research on nights or weekends

Documentation

- Make sure that the documentation is accessible for whichever participants are taking part in your research
  - Informed consent forms, task lists, instructions, may be problematic for people with print-related impairments (blindness, low vision, dyslexia, or motor impairment that keeps them from handling printed materials)
- First, offer to read any documents out loud, or maybe provide audio recordings

Documentation

- Unfortunately, many universities and IRBs require signed informed consent forms
  - For users with motor impairments, ask if you can get an audio or videorecording giving consent to participate
  - For users with cognitive impairment, ask if you can have a caregiver sign on behalf of the participant
  - For blind users, send them electronic copies of the forms beforehand, and then use either Braille labels or a signature guide, giving tactile information to the participant on where to sign the document

Documentation

- While you can make the informed consent form available to participants long before the research begins, you should not send any other research materials beforehand
  - Participants may attempt tasks, or do background research that might change the actual data collection
- If you are reading forms out loud, make sure that you are consistent in what you say

Differing levels of ability

- Be aware that there are varying levels of severity of an impairment
- There are different underlying causes, different levels
- Your assumptions will almost always be wrong
- Other factors that influence performance:
  - Confidence, self-efficacy, and previous experience
- Performance is not always what it seems!

Please remember…

- To leave extra time in your schedule
  - don’t schedule users back-to-back (as they may refuse to give up and therefore take longer to perform tasks)
- To bring extra computer parts if you are going to a participant’s workplace or home
  - if you need monitors or speakers to help with data collection, just remember that if the users don’t need it, they may not have it
- To bring extra cables and keyboards

Please remember…

- To pay participants in a format that is useful for them
  - Gift cards for a certain store may not be appropriate
  - Cash or cash cards are always a good choice
  - Payment formats for university students (e.g. iPod) may not be appropriate!
  - If for a store, make sure that the store has an accessible web site and lots of merchandise choices

End-of-chapter

- Summary
- Discussion questions
- Research design exercise