Carnegie Mellon University

# Effective User Survey Design and Data Analysis

Hanan Hibshi and Travis Breaux

Monday, September 12, 2016

Full Day Tutorial
23rd IEEE International Requirements Engineering Conference

institute for SOFTWARE RESEARCH

# Tutorial Outline

- Effective User Survey Design
  - Session 1: Introduction to experimental research and user surveys
  - Session 2: Building user surveys
- Analyzing and Reporting User Survey Data
  - Session 3: Quantitative analysis of survey data
  - Session 4: Qualitative analysis of survey data

# Why Are We Here?

| | | | | |
|---|---|---|---|---|
| Requirements elicitation, analysis, documentation, validation, and verification | Requirements management, viewpoints, prioritization and negotiation | Requirements specification languages, methods, processes and tools | Modeling of requirements, goals and domains | Tool support for requirements engineering |
| Evolution of requirements over time, product families, variability and reuse | Relating requirements to business goals, architecture and testing | Formal analysis and verification | Social, cultural, global and cognitive factors in requirements engineering | Industry and research collaboration, interdisciplinary insights, learning from practice, and technology transfer |
| Requirements traceability | Requirements related to safety, reliability, security, privacy and digital forensics | Requirements in agile, product line and model-driven development | Requirements in service-oriented, virtualization, embedded, cloud and mobile environments | Domain-specific problems, experiences, and solutions, including new and emerging domains |

# Why take this tutorial?

- You're a junior (or senior!) researcher and you want to use surveys in your research
- You're a practitioner and you want to understand…
  - How customers use one of your products
  - An internal process at your company

Session 1:

# INTRODUCTION TO EXPERIMENTAL RESEARCH AND USER SURVEYS

# Example from a Paper!

- Does this paper measure privacy without asking about it (authors' goal)?

**Survey Question:**

*(Keeping in mind that purchase records may Likely contain sensitive information) If you were to leave a hard copy of one of your [content type] on a restaurant table how likely are you to return to retrieve them?*

*Let's say a server went down and you lost access to your [content type] for two weeks.*

*How would this affect you? It would be ….*

# What is Science?

- Static view:
  - An activity contributing systematic activity
  - Output: laws, theories, hypothesis..
  - Example: atomic energy, technology, etc.
- Dynamic view:
  - Also called *heuristic* view
  - The scientific research methodology
  - The base of other theory and research
  - Leads to further theories, hypothesis and investigation

- Lee, H., 2007. Essentials of Behavioral Science Research.
- Conant, J.B., 1951. Science and common sense.

# Scientific Research

*"Scientific research is systematic, controlled, empirical, amoral, public and critical investigation of phenomena. It is guided by theory and hypotheses about the presumed relations among such phenomena [1]."*

- Systematic and controlled:
  - confidence that other explanations are ruled out
- Empirical:
  - A belief must be put to an outside independent test
- Science is peer-reviewed!
- Results are in terms of validity and reliability
  - Not in "bad" and "good" moral evaluations
- The methods are subject to morality

- Lee, H., 2007. Essentials of Behavioral Science Research.

# Experiments and Science

- Common sense is not always good for science [1].
  - We don't want the new ideas to look like old ones!
- Differences between science and common sense [2]:
  - Uses of concepts and theoretical structures.
    - Scientists test man-made concepts.
  - Hypothesis and theory testing
    - systematic vs. selective
  - The notion of control and limitation of influence
  - Relations among phenomena
  - Explanation of phenomena
    - Scientists rule out metaphysical explanations that cannot be tested

[1] Whitehead, A.N., 1911. *An introduction to mathematics.*
[2] Lee, H., 2007. Essentials of Behavioral Science Research.

# Experiments in CS/SE

- Should follow a scientific model rather than common sense.
- Examples of traditional metrics: task completion time, no of clicks…
- Is Tool A better then Tool B [1]?
  - Do: test the scientific theory
  - Don't: test the tool
- Think about the contribution to science

[1] Steve Easterbrook, "You Gotta Have a Theory," FSE Doctoral Symposium, 2006.

# Ask Questions about the Tool

- What is the purpose of the tool?
  - Test a theory
  - Develop a theory
  - Explain a theory

- Steve Easterbrook, “You Gotta Have a Theory,” FSE Doctoral Symposium, 2006.

# What is a Theory?

- *A theory is a set of interrelated constructs(concepts), definitions, and propositions that present a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting the phenomena*

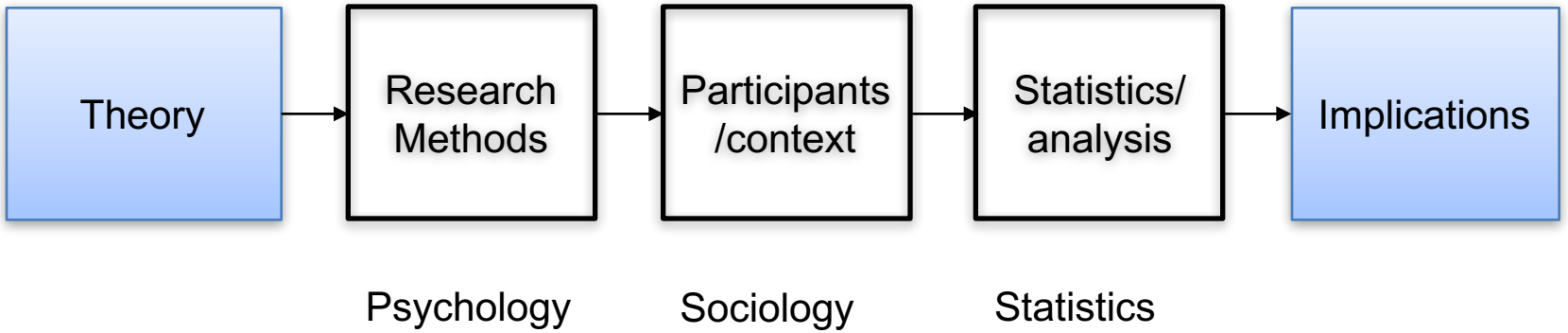
- Lee, H., 2007. Essentials of Behavioral Science Research.

# Concepts, Constructs, and Variables

- A **Concept** is an abstraction formed by generalization from particulars
  - Can be real phenomena (pain) or agreed-upon phenomena (justice, truth)
- A **Construct** is a concept with added meaning created for scientific purposes (satisfaction, IQ)
- A **Variable** is a symbol to which numerals or values can be assigned

- Lee, H., 2007. Essentials of Behavioral Science Research.

# Disciplines in the Research Process




- Lazar, J., Feng, J.H. and Hochheiser, H., 2010. Research methods in human-computer interaction. John Wiley & Sons.

# Theory and Research Questions

## Descriptive / Structural Theories

- What is X?
- In what category is X?
- What are the parts of X, or what is X comprised of?
- What is X similar to or different from within it's category?
- What purpose does X serve?
- How does X affect change and its environment?

## Explanative / Causal Theories

- When and how often does X occur?
- Does X cause or prevent Y from occurring?
- What effects does X have on Y and under what specific conditions?

*Steve Easterbrook, "You Gotta Have a Theory," FSE Doctoral Symposium, 2006.*

# Types of Experimental Research

| Type | Claim | Typical Methods |
|---|---|---|
| Descriptive | X is happening | Observations, field studies, focus groups and interviews |
| Relational | X is related to Y | Observations, field studies, surveys |
| Causal | X is responsible for Y | Controlled experiments |

- 8 out of 10 students who play a game can touch type vs 2 out of 10 who don't play the game can touch type
- Number of hours of game play and touch typing speed
- Causal is difficult to demonstrate theoretically and methodologically


- Rosenthal, R. and Rosnow, R.L., 1991. Essentials of behavioral research: Methods and data analysis. McGraw-Hill Humanities Social.
- Lazar, J., Feng, J.H. and Hochheiser, H., 2010. Research methods in human-computer interaction. John Wiley & Sons.

# Deciding on Surveys

- Advantages:
  - Collect data from a large population
  - Helpful in getting overview of a user population
  - Low cost
  - Easier to obtain Institution Review Board approvals

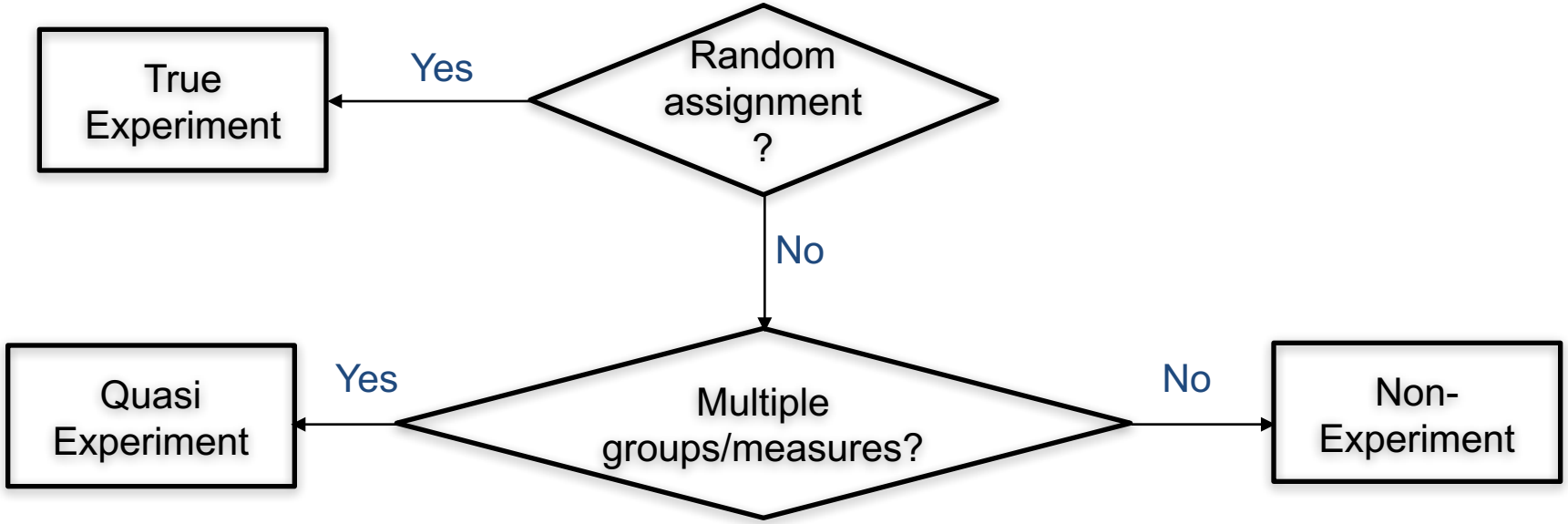
Practical!

- Limitations
  - Good for shallow data, not deep!
  - Follow-up questions is less convenient
  - Information bias when related to non-factual phenomena
    - e.g. age vs. feeling
  - Human factor: fatigue, preference of oral expression…


- Kitchenham, B. and Pfleeger, S.L., 2003. Principles of survey research part 6: data analysis. ACM SIGSOFT Software Engineering Notes, 28(2), pp.24-27.
- Lazar, J., Feng, J.H. and Hochheiser, H., 2010. Research methods in human-computer interaction. John Wiley & Sons.

# Surveys as an Experiment Tool

- With online tools, surveys can be used to conduct online experiments



# PRE-SURVEY PREPARATION

# Pre-survey Checklist

- Research question(s)
- Experiment design
  - Variables and metrics
  - Metrics used
  - Method (online vs. paper)
  - Tools
- Identifying target population
- Care and handling of participants
  - Recruitment methods
- Paper work and approvals
  - IRB
  - Deception
- Demographics & participants' screening

# Research Question

- We need to start with a research question
- The research question be further broken into smaller questions
- We test each hypothesis using empirical investigation.
- Finding the question is critical yet challenging

# RQ Examples

- What are the the concerns of analysts writing use-cases?
- Does background culture affect distributed software teams practices?
- Is developers productivity affected by type of programming language?

# Research Question Exercise

# Exploring the Experimental Options

- *What are the concerns of analysts writing use-cases?*
  - List of survey questions, descriptive data
- *Does distributed software teams practices vary between different cultures?*
  - Ask about cultures, check responses...
- *Is developers' productivity affected by type of programming language?*
  - Ask about languages known, ask questions related to productivity, or
  - Experimental approach:
    - Pick languages: A, B, ….
    - Decide on productivity measure, compare the two groups

# Identifying the Variables and Metrics

- We need to think of independent/dependent variables
  - **Independent variable:** that is being changed or controlled
  - **Dependent variable:** that is being tested
  - **Control variable:** holding a dependent variable constant (e.g. age)

*Is developers' productivity affected by type of programming language?*

- How to measure productivity:
  - Time!
  - Number of lines
  - Number of hours
  - Number of tasks, submissions, commits, results etc.

# Concepts, Constructs, and Variables

- A **Concept** is an abstraction formed by generalization from particulars
  - Can be real phenomena (pain) or agreed-upon phenomena (justice, truth)
- A **Construct** is a concept with added meaning created for scientific purposes (satisfaction, IQ)
- A **Variable** is a symbol to which numerals or values can be assigned

- Lee, H., 2007. Essentials of Behavioral Science Research.

# Measurements

- Can be:
  - Direct observables: height, weight, color, etc.
  - Indirect observables: questions about gender, age, income
  - Constructs: theoretical creations based on observed concepts that cannot be measured directly or indirectly
- Before measurements, variables need to be defined
  - Theoretical definitions: based on theory, definition, or common use
  - Operational definition: assigning the meaning by specifying needed operation
    - Direct: weight, IQ
    - Experimental: frustration: what happens if the Internet goes down while performing a task.

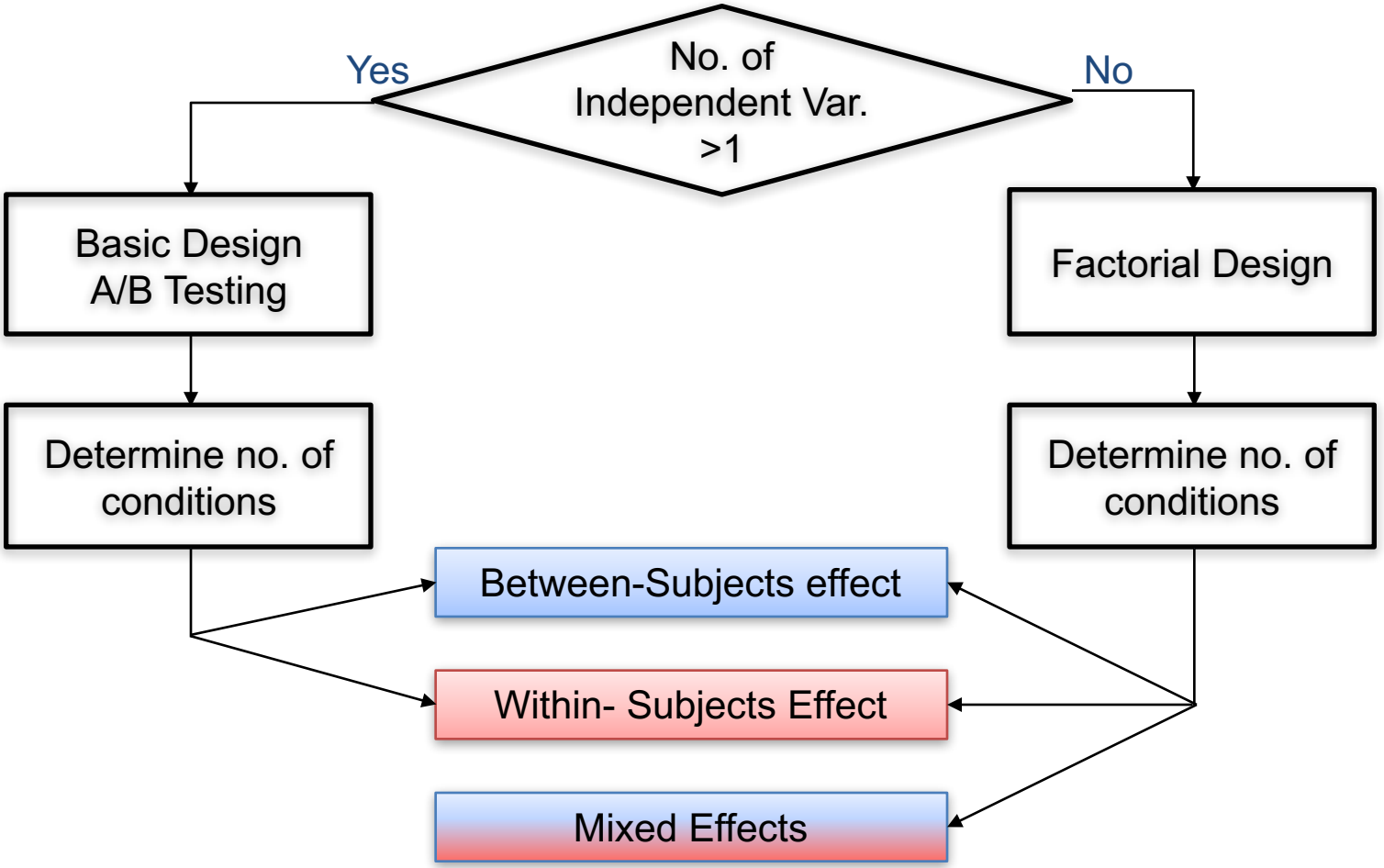
*Is developers' productivity affected by type of programming language?*

- Lee, H., 2007. Essentials of Behavioral Science Research.

# Problems with Measurements

- Measurement error
  - Due to sampling
  - Due to subject or experiment effects
- Reliability
  - = consistency
- Validity
  - The degree to which the measuring instrument is actually measuring the concept

# Experiment Design: Structure




- Lazar, J., Feng, J.H. and Hochheiser, H., 2010. Research methods in human-computer interaction. John Wiley & Sons.

# Instrument: Online vs. Paper

- Online surveys are:
  - Easier to store data electronically
  - Has lower deployment cost
  - Environment friendly
  - Things to think about:
    - assuring a participant takes survey once
    - security, privacy, media, clear instructions
- Paper surveys are:
  - Good for semi-interview style and follow up questions
  - Good for lab experiments (before and after)
  - A better way to recruit participants

# Tools and Technology for Online Surveys

- Depends on:
  - The complexity of survey design
  - The target population and their recruitment method.
- Some scripting might be needed
- Ideas to help with recruitment and verification:
  - Unique URLs with variables
  - Codes (random, unique)

# Examples of Tools

- Survey Gizmo:
  - user friendly interface
  - special scripting language
  - basic plan has limited flexibility


- Qualtrics:
  - special plans for academia
  - JavaScript
  - research oriented


- Amazon's Mechanical Turk:
  - great to advertise surveys to a more general population
  - less helpful with targeted population

# Sources of Bias

- Possible causes:
  - Measurement instrument: stop watch, timer on a page
  - Experimental procedures
    - Within-subjects: learning effect and fatigue → Randomization
    - Instructions received → clear instructions, use visuals if needed
    - Trivial and unforeseen details → write down the procedure
    - Pilot!
  - Participants: domain knowledge, background, self esteem...
  - Experimenter behavior or attitude
  - Environment factors
- We can only be careful, but we cannot fully eliminate bias
  - Careful reporting is a must

- Lazar, J., Feng, J.H. and Hochheiser, H., 2010. Research methods in human-computer interaction. John Wiley & Sons.

# Identifying Target Population

- Why?
  - Decide on recruitment strategy
  - Design questions accordingly
  - Think about demographics
  - Most importantly: data quality!
  - Participants pool need to represent target population [1,2]

*Is developers' productivity affected by type of programming language?*

[1] Broome, J., 1984. Selecting people randomly. Ethics, 95(1), pp.38-55.
[2] Smart, R.G., 1966. Subject selection bias in psychological research. Canadian Psychologist/Psychologie canadienne, 7(2), p.115.

# Target Population Exercise

*Is developers' productivity affected by type of programming language?*

- Who is your target population here?
  - Developers
  - Software engineers
  - IT Managers
  - Help desk personnel
  - Students

# Recruitment Methods

- Online advertisements
  - Amazon MTurk, forums, classified adds (craigslist)
- E-mail
- Personal networking
- Social networks
- Ads in high traffic public spaces
  - Coffee shops, waiting areas, conferences
- Ask in person (beg! use puppy face!)
- Need to have a mechanism for verification
- Add details about the study in advertisement materials

# Care and Handling of Participants

- Reduce participant's bias
  - Try to eliminate stressful environment/situations.
  - Reassure participants that you are testing the tool not them
  - Eliminate timers when not needed
  - Provide clear test material
  - Thank participants and show gratitude for their commitment
  - Offer fair compensation for their time
  - Allow them to opt out
  - Maintain privacy and confidentiality
    - Storage of data, access to data, handling names and PII…

# Ethical Research Guidelines

- A researcher should always conduct ethical research
  - Whether enforced by the country or not
  - Ensure no harm to subjects: physical or emotional
- The Belmont Report (1979) defined 3 principles:
  - Respect for persons to make their own judgments
  - Beneficence: minimize harm and maximize benefits
  - Justice: participation is not limited to easily manipulated groups.
- Need to give participant an informed consent
  - Due to harms carried historically by medical experiments
- The Stanford Prison Experiment

- National Commission, 1979. The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research. National Commission for the Protection of Human Subjects of Biomedical and Behavioral research.
- Lazar, J., Feng, J.H. and Hochheiser, H., 2010. Research methods in human-computer interaction. John Wiley & Sons.

# Informed Consent

- Should be done before participation
- Informed: participants need to know
  - procedure
  - risks
  - benefits
  - reasons
  - contact information of researchers
- Consent: participation is:
  - voluntary
  - free from any form of pressure
    - e.g. students class credit, medical treatment

- Lazar, J., Feng, J.H. and Hochheiser, H., 2010. Research methods in human-computer interaction. John Wiley & Sons.

# Sections of Informed Consent

- Title and Purpose
- Description of Procedures
- Duration
- Risks
- Benefits
- Alternatives to participation
- Confidentiality
- Costs/additional Expenses
- Participant's rights
- Contact Information
- Supplemental information
- Signature

- National Cancer Institute (2001). A guide to understanding informed consent,

# Participants' Confidentiality

- Especially important with online surveys
- Be extra careful about Personally Identifiable Information (PII)
  - Emails
  - Telephone numbers
  - Names
  - Addresses
- Rule of thumb: **If you don't need it, do NOT collect it**
- Some guidelines:
  - Use passwords
  - Drawers with locks
  - Store emails and PII separate from responses
  - Anonymize the data before analyses
  - Only share Anonymized data
  - Remember to list the risk of data breach in the informed consent!

# Institutional Review Boards

- IRB: A committee to review human subjects research
- Conduct risk-benefit analysis
- Regulations (by FDA and HHS) in the US require IRB reviews
  - National Research Act of 1974
  - Belmont Report
- Informed consent is a must
  - Can be waved under very specific circumstances when the risk is very low
- Audio and video studies complicate the review process
- International researchers who conduct research in the US still need IRB approvals
- There are similar boards in other countries

# IRB Review Types

- Exception reviews:
  - Existing data that is publicly available
  - Data collection is not for research purposes
    - e.g. improve classes, courses…
  - Data to evaluate things not subjects behavior or attitude
  - Exception decision is made by a board member
- Type of reviews:
  - Exempt
  - Expedited
  - Full

# IRB Paper Work

- Provide details of:
  - Study procedures
  - Compensation
  - How confidentiality is handled
  - Risks and benefits
  - Recruitment method and venues
  - Population
  - Whether deception is used
- Provide copies of:
  - Informed consent
  - Study script
  - Study questions

# Deception

- If the study requires concealing the real purpose
- The participant is asked about a certain task while the goal is measuring something else.
- Deception must be approved by IRB
- Deception is tricky:
  - Metrics for the construct asked need to be related to the intended construct

# Example from a Paper!

- Does this paper measure privacy without asking about it (authors' goal)?

**Survey Question:**

*(Keeping in mind that purchase records may Likely contain sensitive information) If you were to leave a hard copy of one of your [content type] on a restaurant table how likely are you to return to retrieve them?*

*Let's say a server went down and you lost access to your [content type] for two weeks.*

*How would this affect you? It would be ….*

# Deception Exercise

*Is developers' productivity affected by type of programming language?*

# Collecting Demographics

- Demographics data help describe your sample
- Standard questions are available online
- Its better to tweak questions according to the research needs

*Is developers' productivity affected by type of programming language?*

- Age
  - Below 18
  - 18 – 35
  - 35 -50
  - 50 – 60
  - Above 60

# More Demographics Examples

- What is your occupation?
- What is your income range?
- Which country do you reside in ?
- Gender?
- Highest degree?
- Nationality?
- No of computers owned?
- Type of Gadgets?

# Background Questions

- Questions that aim to evaluate one's expertise
- Can be:
  - direct:
    - How many years of experience?
    - What programming language(s) do you use?
  - Indirect:
    - The following line of code is used in the which PL?
- Self-assessment ratings could be biased
  - Experts tend to underestimate their ability while novices tend to be over confident
- Background questions can be used for screening

# Participants' Screening

- Very useful when targeting a specific population
  - Experts
  - Novices
  - Certain skills
- Help to save time and resources
- Especially useful when using crowdsourcing platforms (Mturk)
- Make sure to still compensate particiaptns for their time taking the screening test.
- When to use screening: when data from other groups is useless even for comparison.

# Sampling

- The process of selecting subjects from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen.