# Introduction to Empirical Software Engineering

Idaho State University | Computer Science

## Isaac Griffith

CS 6620 - Empirical Software Engineering
Department of Computer Science
Idaho State University

ROAR

# Outline

- Empirical Strategies
- Replication Studies
- Basics of Research Ethics

ROAR

# Empirical Strategies

# Overview of Empirical Strategies

- **Surveys** - a system for collecting information from or about people to describe, compare or explain their knowledge, attitudes and behavior.
- **Case Studies** - an empirical enquiry that draws on multiple sources of evidence to investigate one instance (or a small number of instances) of a contemporary SE phenomenon within its real-life context, especially when the boundary between phenomenon and context cannot be clearly specified.

ROAR

# Overview of Empirical Strategies

- **Experiments** - an empirical enquiry that manipulates one factor or variable of the studied setting. Based in randomization, different treatments are applied to or by different subjects, while keeping other variables constant, and measuring the effects on outcome variables. In human-oriented experiments, humans apply different treatments to objects, while in technology-oriented experiments, different technical treatments are applied to different objects.

- **Quasi-Experiments** - an empirical enquiry similar to an experiment, where the assignment of treatments to subjects cannot be based on randomization, but emerges from the characteristics of the subjects or objects themselves.

ROAR

# Overview of Empirical Strategies

| Strategy | Design type | Qualitative/quantitative |
|----------|-------------|--------------------------|
| Survey | Fixed | Both |
| Case Study | Flexible | Both |
| Experiment | Fixed | Quantitative |

ROAR

# Surveys

- Conducted before **or** after a technique or tool is used
- Used to gain a general understanding of a population
- General Objectives (one of the following):
  - Descriptive - to understand the distribution of some characteristic or attribute of the target population
  - Explanatory - to make some explanatory claim about the population
  - Exploratory - as a pre-study to a more through investigation
- Data is collected using questionnaires or interviews

ROAR

# Case Studies

- Investigates a single entity or phenomenon in its real-life context
- Case studies can evaluate how or why certain phenomena occur and also can evaluate the difference between approaches
- Useful when investigating SE methods and tools in industry
  - Avoids scale-up problems
- Useful if an effect is widespread

ROAR

# Case Study Arrangements

Comparative Cases studies can be arranged in 3 primary ways:

- Comparison of a new method to a known baseline
- Comparison of a new method to a similar project using a baseline method
- Comparison of product component, where some have had the new method applied (at random) while others have not.

ROAR

# Case Studies Pros/Cons

Pros:

- incorporate qualities that an experiment cannot

Cons:

- A small/simplified case study is usually not adequate to handle discovering SE principles and techniques.
- Researchers are not in full control

# Experiments

- Launched when we want full control over the situation
- Allow us to manipulate behavior directly, precisely and systematically
- Experiments are:
  - human-oriented: humans apply different treatments to objects
  - technology-oriented: different tools are applied to different objects
- Allow us to:
  - Confirm theories
  - Confirm conventional wisdom
  - Explore relationships
  - Validate measures

ROAR

# Research Strategy Factors

- Execution Control – How much control a researcher has over the study.

- Measurement Control – the degree to which the researcher can decide upon which measures to be collected, and to include or exclude during execution of the study

- Investigation Cost

- Ease of Replication

ROAR

# Comparison

| Factor | Survey | Case Study | Experiment |
|---|---|---|---|
| Execution control | No | No | Yes |
| Measurement control | No | Yes | Yes |
| Investigation cost | Low | Medium | High |
| Ease of replication | High | Low | High |

ROAR

# Replication Studies

# Replication Studies

It's not science until it has been replicated.

- Replication of an experiment involves repeating the investigation under similar conditions.

- Replication helps to determine the level of confidence to place in the original results.
  - If the assumptions of the original study are correct
  - and the results are believed to generalize to a larger population
  - then a replication within the population should get the same results

- If we do not get the same results, then we did not capture all aspects of the experiment design affecting the result.

ROAR

# Replication Types

- Strict Replications - follow the original procedures as closely as possible

- Differentiated replications
  - Replications that vary variables intrinsic to the study
  - Replications that vary variables intrinsic to the focus of the study
  - Replications that vary the context variables in the environment in which the solution is evaluated
  - Replications that vary the manner in which the experiment is run
  - Replications that extend the theory

Note: due to the complex nature of SE research, most replications could be called close, but we should always focus on replication original hypotheses rather than experimental designs (differentiated over close).

ROAR

# Factors to Vary

The following factors which should be reported in a replication study:

- **Site** - where the experiment was conducted
- **Experimenters** - those conducting the experiment
- **Design** - experimental design
- **Instrumentation** - forms and other materials
- **Variables** - measured
- **Subjects** - conducting the experiment

# Secondary Studies

Aggregating Evidence from Empirical Studies

ROAR

# Secondary Studies

- As the empirical studies grow in number a need to aggregate their findings arises

- The goal is that research should build upon what we already know

- Multiple studies can provide answers which a single study may be incaple of providing

- There are two approaches to this:
  - Systematic Literature Reviews (SLRs)
  - Systematic Mapping Studies (SMSs)

ROAR

# **Secondary Studies**

SLRs

- Present a means to collect and synthesize empirical evidence from multiple sources
- Research questions focus on the outcomes of reviewed empirical studies to compare different technologies or methods.

SMSs

- Used when the research questions is more general or the area of research less explored.
- SMSs have much broader research questions than SLRs
- More qualitative than SLRs
- Less stringent classification of documents than SLRs
- Typically goals are:
    - Identify current state of practice

ROAR

# Secondary Studies

1. Identify research questions
2. Develop search string and identify research databases to search
3. Identify inclusion/exclusion criteria
4. Execute search to find initial studies
5. Apply inclusion/exclusion criteria to whittle down to the list of primary studies
6. Use snowball sampling to identify missed studies
7. Extract data from primary studies to answer research questions
8. Analyze extracted data

ROAR

# Analytical Methods

- Meta-analysis - application of statistical methods to synthesize data.
- Descriptive synthesits - visualization and tabulation of data and descriptive statistics of data.

ROAR

# Ethics in Experimentation

# Singer & Vinson

- Paper based on a series of vignettes
- Key ideas:
  - Informed Consent (participants' perceptions matter!)
  - Scientific Value (is the study worth doing?)
  - Beneficience (Do the benefits outweigh the risks?)
  - Confidentiality
  - Consent from creators of artifacts?
- Role of Ethics Review Board (or IRB)
- Ethical considerations needed before the study starts!

ROAR

# Discussion Questions

- What other codes of ethics apply (e.g. ACM?)
- How much detail do participants need about the study?
- How can we ensure confidentiality in the era of data mining?
- Is it ever ethical for course instructors to run controlled experiments on different pedagogical techniques?
- How important is it to share the results of your study with the participants?

ROAR

# Research Ethics

- Reasons to take ethics seriously:
  - Funding depends on it
  - Relationship with research subjects/organizations depends on it
  - Legal issues (e.g. liability for harm to subjects/organizations)
  - Compliance with privacy and data protection laws
  - …and it's the right thing to do!

- Institutional Review Boards (IRB)
  - Approval usually needed for all studies involving human subjects
  - Every IRB has it's own rules…
    - A study approved at one university may be disallowed at another!
    - Design of the study might have to be altered
  - Institutional research funding may depend on this process!
  - Note: guidelines from other fields may not apply to Software Engineering
    - E.g. use/ownership of source code

ROAR

# Informed Consent

- Elements
  - Disclosure - participants have full information about purpose, risks, benefits
  - Comprehension - jargon-free explanation, so participants can understand
  - Competence - participants must be able to make rational informed choice
  - Voluntariness - no coercion or undue influence to participate
  - Consent - usually indicated by signing a form
  - Right to withdraw
    - participant can withdraw from study at any point without having to give reasons
    - Participants can request their data to be excluded (might not be possible)
- Challenges:
  - Student participants
    - Perception that their grade will be affected if they don't participate
    - Perception that it will please the course instructor if they participate
  - Industrial participants

ROAR

# An Ethical Dilemma...

You are dong a study of how junior analysts use new requirements tool at a leading consultancy company. As part of informed consent, staff are informed that they will remain anonymous. During the study, you notice that many of the analysts are making data entry errors when logging time spent with clients. These errors are causing the company to lose revenue. Company policy states clearly that workers salaries will be docked for clear mistakes leading to loss of revenue.

Questions:

- Would you alter the results of your study to protect the people who help you in the study?

- How can you report results without causing harm to the participants?

- Would you cancel the study as soon as this conflict of interest is detected?

ROAR

# Should you pay your participants?

- Arguments in favor
  - Can help with recruitment
  - Compensate participants for their time

- Arguments against
  - May induce participants to take risks they otherwise would not take
  - May get expensive (especially if rates are to be more than a token)

- Issues
  - IRB might have standard rate; might be too low for professional SE

- Alternatives:
  - All participants entered into draw for some new gadget

ROAR

# Beneficence

- Risk of harm to Participants
  - Disrupts participant's work
  - Results of the research may devalue participants' work
  - Publication of study may harm the company's business

- Benefits of study
  - Scientific value: useful to society?
  - Depends on importance of the research topic!
  - Note: validity is crucial - invalid results means the study has no benefits
  - May also be specific benefits to participants
    - e.g. training, exposure to state-of-the art techniques, etc.

- Beneficence: Benefits should outweigh the risks
  - Understand and justify any tradeoffs in the design of the study

ROAR

# Confidentiality

- Protecting Anonymity
  - Do not collect any data (e.g. names) that allow participants to be identified
  - But you need a signed consent form, so…
  - Sever participants' identify from their data before it is stored and analyzed
  - Researcher-subject interactions should be held in private
- Protecting the data
  - Consent form states who will have access to the data, and for what purpose
    - Do not stray from this!
  - Raw data should be kept in a secure location
  - Reports should only include aggregate data
- Exceptions:
  - When it is impossible to identify individuals from the raw data
  - When more harm results from maintaining confidentiality than breaching it

# Are there any questions?

ROAR