

# Empirical studies

**Definition: Evaluate**

To judge the quality of

## 1 Empirical vs Experimental

**Definition: Empirical**

Relying on observation and experiment rather than theory

**Definition: Experiment**

A study in which an intervention is deliberately controlled to observe its effects

## 2 Forms of measure

**Definition: Quantitative evaluation**

Used to determine whether a cause effect relationship exists

- May test the effect of some intervention
- Uses measures based on "counting" scales
- Can employ statistical forms to aid analysis

**Definition: Qualitative evaluation**

Studies entities in their natural setting, usually through observation:

- Analysis involves interpretation based on explanations
- Recognises that there may be different interpretations

## 3 Primary or secondary

**Definition: Primary study**

Directly study the entity of interest by making observations and measurements

**Definition: Secondary study**

Seek to aggregate the outcomes of many different primary studies

## 4 The research protocol

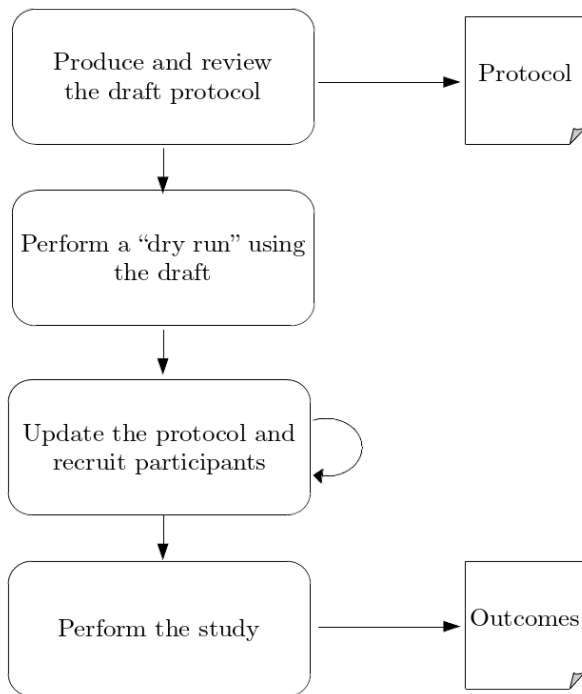
A good evaluation process needs to be:

- Objective
- Unbiased

And should avoid “fishing” for results from the outcomes. So we first draw up a plan for conducting the study, called the research protocol

- Usually perform some form of dry run to test the protocol in a controlled situation
- When reporting the study, we also need to describe any divergences from the plan that occurred
- The protocol also identifies likely threats to validity, factors that we can’t control that might reduce our confidence in the outcomes

## 5 The dry run



## 6 Primary studies in SE

Major forms of primary studies used in software engineering research:

- Controlled experiments
- Quasi experiments
- Surveys
- Case studies

## 7 Randomised controlled trial (RCT)

The ideal RCT includes:

- Participants being unaware of whether they are receiving the treatment or a placebo (blinding)
- Those running the trial being unaware of who is receiving the treatment (double blinding)
- The analyst working out the result not knowing who was in the trial group (triple blinding)

In SE the only element we can blind is analysis

### 7.1 Between subjects

Participants are divided into two groups:

- The "control", a group who performs their task using "standard" forms
- The "experimental group", who perform their task using the treatment under investigation

## 8 Quasi-Experiments

- Quasi-experimental forms are used when it is impossible or impractical to perform random allocation of participants to a group
- Often used when participants are required to have specific skills or knowledge
- There are many ways to organise a quasi-experiment, including
  - Cross-over forms
  - Before-after forms
  - Time interval forms

## 9 Analysis

- There will always be variation in the measured outcomes because we are using people and hence other factors may have an influence upon the outcomes
- This means that deciding between the hypothesis and the null hypothesis becomes a statistical task. By convention, aim for a 95% confidence level
- Experimenters are expected to report the confidence level when they state their results

## 10 Surveys

Used to collect information from a large group of people in a standard and systematic way, so that we can seek patterns in the data and generalise what these imply for a wider population than our sample

Typically used for two purposes:

- **Experimental:** To assess the impact of some intervention
- **Descriptive:** So that we can make assertions about some phenomenon of interest and where we are less interested in why this occurs as to how much it does

### 10.1 Key concepts

We seek a sample of respondents who are suitably representative of a larger sample frame

Selection is made through use of a sampling technique, this can be:

- **Probabilistic** - random, systematic, stratified, cluster
- **Non-probabilistic** - self selection, snowballing, convenience

## 10.2 Data collection

**Questionnaires** provide consistency of data collection but

- Limited as to type of question can use
- Response rates may be poor

**Interviews** can be either structured or semi-structured

**Observations** requiring no direct involvement with the participants

**Literature** perhaps using data mining

## 10.3 Problems

- We rarely know the size of our sampling frame
- Difficult to identify and access a sample of participants
- Small sample sizes make it difficult to get results with a high confidence level
- It is quite challenging to design questions that are unbiased and that allow users to answer them reliably
- In a well designed survey the questions will reinforce each other so that during analysis we can ensure that participants are giving consistent answers

## 11 Case studies

### Definition: Case study

A controlled form of observational field study, involving planned data collection

A case study typically involves:

- More variables of interest than data points
- Use of triangulation between multiple sources of evidence
- Prior development of propositions

### 11.1 Types of case study

#### Definition: Explanatory study

Used to answer questions about how some phenomenon works and why it works

#### Definition: Descriptive study

Used to produce a rich and detailed analysis of a phenomenon and its context. Involves less detail about mechanisms than an explanatory study

#### Definition: Exploratory study

Used to lay the groundwork for a later fuller study, perhaps by helping identify the questions or help understand a problem

## 11.2 Use in software engineering

Particularly useful when investigating how software engineering practices are adopted or used in an industry setting, where we have:

- Limited control of the situation, so can only observe
- Relatively few "cases"
- Many diverse sources of data, project logs, minutes of meetings, interviews with the team
- A need to study an effect "in the field"