Chapter 4
Empirical Investigation

SE Investigation

- What is software engineering investigation?
- Applying "scientific" principles and techniques to investigate properties of software and software related tools and techniques.
- Scientific investigation: is a rigorous, systematic approach, designed to eliminate bias and other subjective influences in the search, identification, and measurement or validation of facts and cause-effect relationships, and from which scientific laws may be deduced.
- Why talking about software engineering investigation?
- Because the standard of empirical software engineering research is quite poor.

SE Investigation: Examples

- Experiment to confirm rules-of-thumb
  - Should the LOC in a module be less than 200?
  - Should the number of branches in any functional decomposition be less than 7?
- Experiment to explore relationships
  - How does the project team experience with the application affect the quality of the code?
  - How does the requirements quality affect the productivity of the designer?
  - How does the design structure affect maintainability of the code?
- Experiment to initiate novel practices
  - Would it be better to start OO design with UML?
  - Would the use of SRE improve software quality?

SE Investigation: Why?

- To improve (process and/or product)
- To evaluate (process and/or product)
- To prove a theory or hypothesis
- To disprove a theory or hypothesis
- To understand (a scenario, a situation)
- To compare (entities, properties, etc.)

SE Investigation: What?

- Person's performance
- Tool's performance
- Person's perceptions
- Tool's usability
- Document's understandability
- Program's complexity etc.

SE Investigation: Where & When?

- In the field
- In the lab
- In the classroom
- Anytime depending on what questions you are asking

SE Investigation: How?

- Hypothesis/question generation
- Data collection
- Data evaluation
- Data interpretation
- Feed back into iterative process

SE Investigation: Characteristics

- Data sources come from industrial settings
  - This may include people, program code, etc.
- Usually
  - Surveys
  - Case-studies (→ hypothesis generation)
  - Experiments (→ hypothesis testing)

Where Data Come From?

- First Degree Contact
  - Direct access to participants
- **Example:**
  - Brainstorming
  - Interviews
  - Questionnaires
  - System illustration
  - Work diaries
  - Think-aloud protocols
  - Participant observation

Where Data Come From?

- Second Degree Contact
  - Access to work environment during work time, but not necessarily participants
- **Example:**
  - Instrumenting systems
  - Real time monitoring

Where Data Come From?

- Third Degree Contact
  - Access to work artifacts, such as source code, documentation
- **Example:**
  - Problem report analysis
  - Documentation analysis
  - Analysis of tool logs
  - Off-line monitoring

Practical Considerations

- Hidden Aspects of Performing Studies
  - Negotiations with industrial partners
  - Obtaining ethics approval and informed consent from participants
  - Adapting “ideal” research designs to fit with reality
  - Dealing with the unexpected
  - Staffing of project

Investigation Principles

There are 4 main principles of investigation:

1. **Selecting investigation technique:** conducting surveys, case studies, formal experiments
2. **Stating the hypothesis:** What should be investigated?
3. **Maintaining control over variables:** dependent and independent variables
4. **Making meaningful investigation:** verification of theories, evaluating accuracy of models, validating measurement results.

SE Investigation Techniques

- Three ways to investigate:
  - **Formal experiment:** A controlled investigation of an activity, by identifying, manipulating and documenting key factors of that activity.
  - **Case study:** Document an activity by identifying key factors (inputs, constraints and resources) that may affect the outcomes of that activity.
  - **Survey:** A retrospective study of a situation to try to document relationships and outcomes.

Guidelines to choose A particular investigation method

- If the activity has already occurred, we can perform survey or case study. If it is yet to occur, then case study or formal experiment may be chosen.
- If we have a high level of control over the variables that can affect the outcome, then we can use an experiment. If we have no control over the variable, then case study will be a preferred technique.
- If replication is not possible at higher levels, then experiment is not possible.
- If the cost of replication is low, then we can consider experiment.

Hypothesis

- The first step is deciding what to investigate.
- The goal for the research can be expressed as a hypothesis in quantifiable terms that is to be tested.
- The test result (the collected data) will confirm or refute the hypothesis.
- **Example:**

Can Software Reliability Engineering (SRE) help us to achieve an overall improvements in software development practice in our company?

Examples /1

- **Experiment: research in the small**
  - You have heard about software reliability engineering (SRE) and its advantages and may want to investigate whether to use SRE in your company. You may design a controlled (dummy) project and apply the SRE technique to it. You may want to ***experiment*** with the various phases of application (defining operational profile, developing test-cases and decision upon adequacy of test run) and document the results for further investigation.

Examples /2

- **Case study: research in the typical**
  - You may have used software reliability engineering (SRE) for the first time in a project in your company. After the project is completed, you may perform a ***case-study*** to capture the effort involved (budget, personnel), the number of failures investigated, and the project duration.

Examples /3

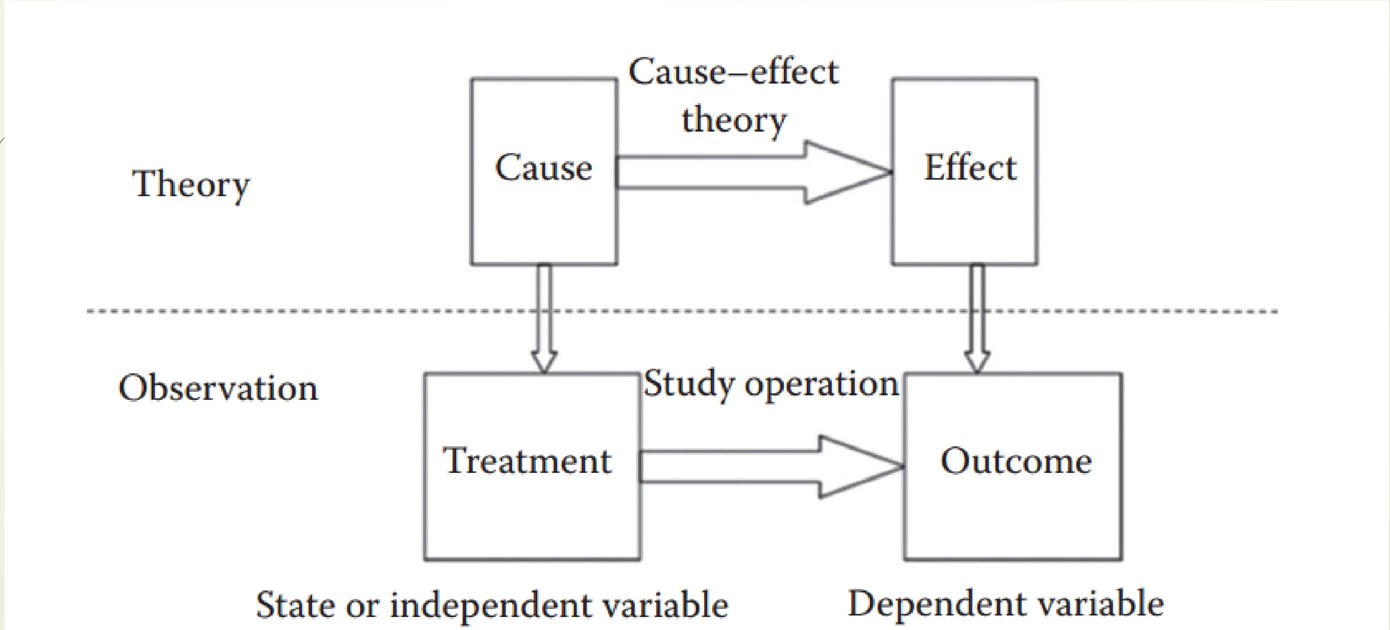
- **Survey: investigate in the large**
  - After you have used SRE in many projects in your company, you may conduct a ***survey*** to capture the effort involved (budget, personnel), the number of failures investigated, and the project duration for all the projects. Then, you may compare these figures with those from projects using conventional software test technique to see if SRE could lead to an overall improvements in practice.

Hypothesis (cont'd)

- **Other Examples:**
  - Can integrated development and testing tools improve our productivity?
  - Does Cleanroom software development produce better-quality software than using the conventional development methods?
  - Does code produced using Agile software development have a lower number of defects per KLOC than code produced using the conventional methods?

Control /1

- What variables may affect truth of a hypothesis? How do they affect it?
- **Variable:**
  - ***Independent*** (values are set by the experiment or initial conditions)
  - ***Dependent*** (values are affected by change of other variables)
- **Example:** Effect of "programming language" on the "quality" of resulting code.
  - Programming language is an independent and quality is a dependent variable.


Cause–effect theory
Theory
Cause
Effect
Observation
Study operation
Treatment
Outcome
State or independent variable
Dependent variable

Control /2

- A common mistake: ignoring other variables that may affect the values of a dependent variable.
- **Example:**
- Suppose you want to determine whether a change in ***programming language*** (independent variable) can affect the ***productivity*** (dependent variable) of your project. For instance, you currently use FORTRAN and you want to investigate the effects of changing to Ada. The values of all other variables should stay the same (e.g., application experience, programming environment, type of problem, etc.)
- Without this you cannot be sure that the difference in productivity is attributable to the change in language.
- But list of other variables may grow beyond control!

Control /3

- How to identify the dependent and independent variables?
- Example:

$$A \rightarrow D$$

$$F \,\&\, B \rightarrow Z \quad D \,\&\, C \rightarrow F$$

$$Given: \{A, B, C\}$$

Using causal ordering:

$$\{A, B, C\} \Rightarrow \{D\} \Rightarrow \{F\} \Rightarrow \{Z\}$$

Planning Experiments

A Process Model for Performing Experiments: six-phase process

1. **Conception**
   - Defining the goal of investigation
2. **Design**
   - Generating *quantifiable* (and *manageable*) hypotheses to be tested
   - Defining experimental objects or units
   - Identifying experimental subject
   - Identifying the response variable(s)

Planning Experiments

3. **Preparation**
   - Getting ready to start, e.g., purchasing tools, hardware, training personnel, etc.
4. **Execution**
5. **Review and analysis**
   - Review the results for soundness and validity
6. **Dissemination & decision making**
   - Documenting conclusions

Formal Experiments: Principles

1. **Replication**
   - Experiment under identical conditions should be repeatable.
   - Confounded results (unable to separate the results of two or more variables) should be avoided.
2. **Randomization**
   - The experimental trials must be organized in a way that the effects of uncontrolled variables are minimized
   - Types of experimental design

     crossing and nesting

Formal Experiments: Principles

3. Local control

- **Blocking:** allocating experimental units to blocks or groups so the units within a block are relatively homogeneous. The blocks are designed so that the experimental design captures the anticipated variation in the blocks by grouping like varieties, so that the variation does not contribute to the experimental error.
- **Balancing:** is the blocking and assigning of treatments so that an equal number of subjects is assigned to each treatment. Balancing is desirable because it simplifies the statistical analysis.

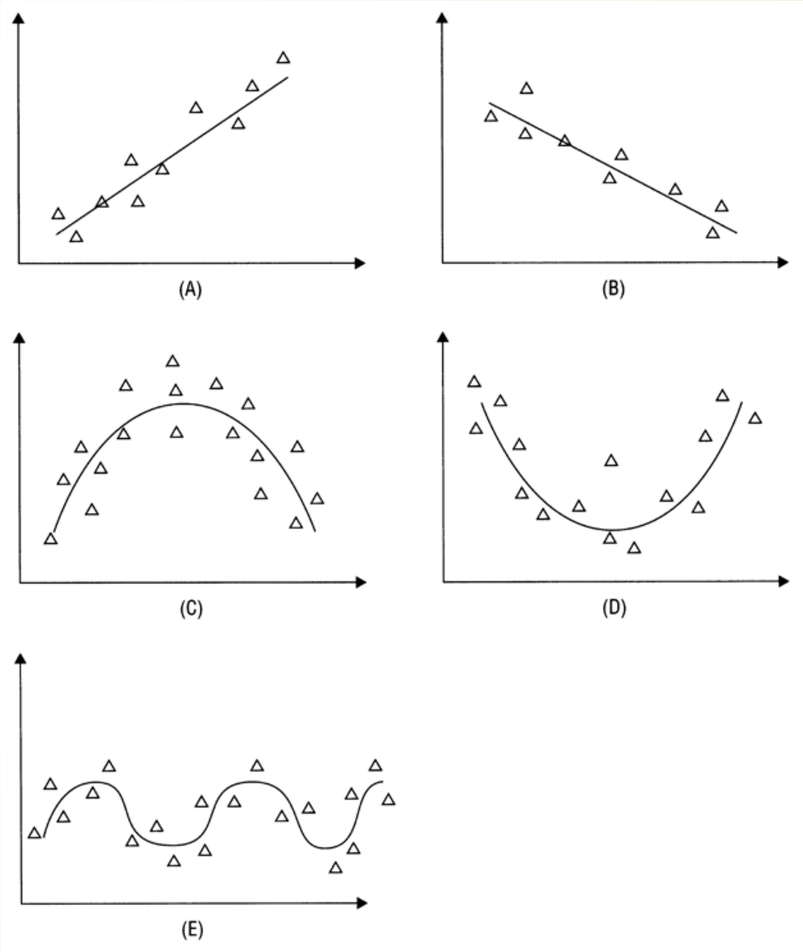
Example: Blocking & Balancing

- You are investigating the comparative effects of three design techniques on the quality of the resulting code.
- The experiment involves teaching the techniques to 12 developers and measuring the number of defects found per 1000 LOC to assess the code quality.
- It may be the case that the twelve developers graduated from three universities. It is possible that the universities trained the developers in very different ways, so that being from a particular university can affect the way in which the design technique is understood or used.
- To eliminate this possibility, three blocks can be defined so that the first block contains all developers from university X, the second block from university Y, and the third block from university Z. Then, the treatments are assigned at random to the developers from each block. If the first block has six developers, two are assigned to design method A, two to B, and two to C.

Formal Experiments: Principles

3. Local control

- **Correlation:** the most popular technique to assess relationships among observational data
- Linear and nonlinear correlation.
- Nonlinear correlation is hard to be measured and may stay hidden.



Formal Experiments: Types

Factorial design:

- **Crossing** (each level of each factor appears with each level of the other factor)
- **Nesting** (each level of one occurs entirely in conjunction with one level of another)
- Proper nested or crossed design may reduce the number of cases to be tested.

		Factor *B*		
		Level 1	**Level 2**	**Level 3**
Factor *A*	**Level 1**	$a_1\ b_1$	$a_1\ b_2$	$a_1\ b_3$
	Level 2	$a_2\ b_1$	$a_2\ b_2$	$a_2\ b_3$

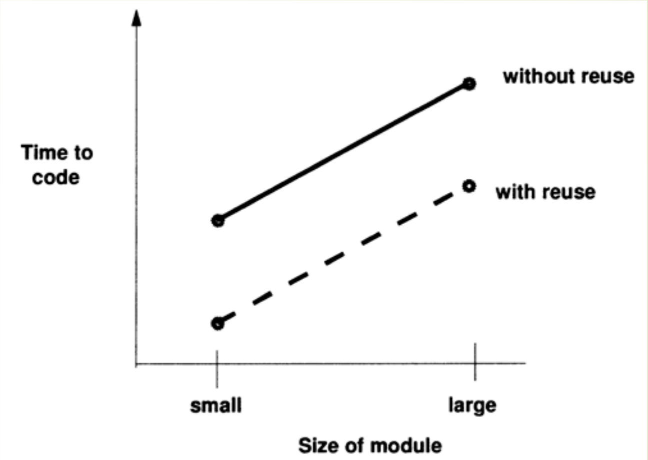
Factor *A*					
Level 1			**Level 2**		
Factor *B*			Factor *B*		
Level 1	**Level 2**	**Level 3**	**Level 1**	**Level 2**	**Level 3**
$a_1\ b_1$	$a_1\ b_2$	$a_1\ b_3$	$a_2\ b_1$	$a_2\ b_2$	$a_2\ b_3$

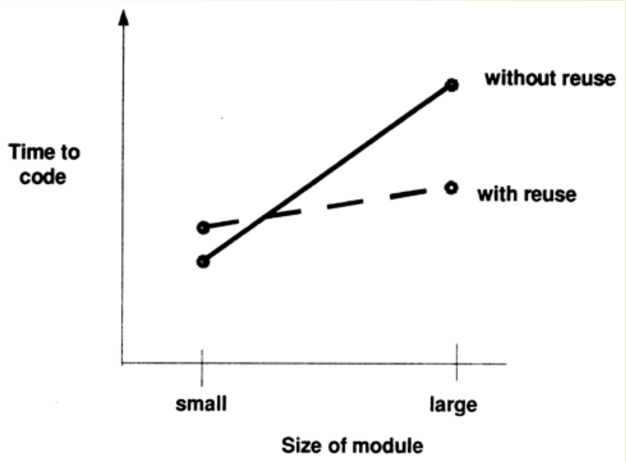
Formal Experiments: Types

- Advantages of factorial design
  - Resources can be used more efficiently
  - Coverage (completeness) of the target variables’ range of variation
  - Implicit replication
- Disadvantages of factorial design
  - Higher costs of preparation, administration and analysis
  - Number of combinations will grow rapidly
  - Some of the combinations may be worthless

Formal Experiments: Selection

- Selecting the number of variables:
  - Single variable
  - Multiple variables
- **Example:** Measuring time to code a program module with or without using a reusable repository
  - Without considering the effects of experience of programmers
  - With considering the effects of experience of programmers





Formal Experiments: Baselines

- Baseline is an “average” treatment of a variable in a number of experiments (or case studies).
- It provides a measure to identify whether the value is within an acceptable range.
- It may help checking the validity of measurement.

Empirical Research Guidelines

Contents

1. Experimental Context

Goals:

- Ensure that the *objectives* of the experiment have been properly defined
- Ensure that the *description* of the experiment provides enough details for the practitioners

1. Experimental Context

- **C1:** Be sure to specify as much of the context as possible. In particular, clearly define the entities, attributes and measures that are capturing the contextual information.
- **C2:** If a specific hypothesis is being tested, state it clearly prior to performing the study, and discuss the theory from which it is derived, so that its implications are apparent.
- **C3:** If the target is exploratory, state clearly and, prior to data analysis, what questions the investigation is intended to address, and how it will address them.

2. Experimental Design

Goal:

- Ensure that the design is appropriate for the objectives of the experiment
- Ensure that the objective of the experiment can be reached using the techniques specified in the design

2. Experimental Design /1

- **D1:** Identify the population from which the subjects and objects are drawn.
- **D2:** Define the process by which the subjects and objects were selected (inclusion/exclusion criteria).
- **D3:** Define the process by which subjects and objects are assigned to treatments.
- **D4:** Restrict yourself to simple study designs or, at least, to designs that are fully analyzed in the literature.
- **D5**: Define the experimental unit.

2. Experimental Design /2

- **D6:** For formal experiments, perform a pre-experiment or pre-calculation to identify or estimate the minimum required sample size.
- **D7:** Use appropriate levels of blinding.
- **D8:** Avoid the use of controls unless you are sure the control situation can be unambiguously defined.
- **D9**: Fully define all treatments (interventions).
- **D10:** Justify the choice of outcome measures in terms of their relevance to objectives of the empirical study.

3. Data Collection

Goal

- Ensure that the data collection process is well defined
- Monitor the data collection and watch for deviations from the experiment design

3. Data Collection

- **DC1:** Define all software measures fully, including the entity, attribute, unit and counting rules.
- **DC2**: Describe any quality control method used to ensure completeness and accuracy of data collection.
- **DC3:** For observational studies and experiments, record data about subjects who drop out from the studies.
- **DC4:** For observational studies and experiments, record data about other performance measures that may be adversely affected by the treatment, even if they are not the main focus of the study.

4. Analysis

Goal

- Ensure that the collected data from the experiment is analyzed correctly
- Monitor the data analysis and watch for deviations from the experiment design

4. Analysis

- **A1:** Specify any procedures used to control for multiple testing.
- **A2:** Consider using blind analysis (avoid “fishing for results”).
- **A3:** Perform sensitivity analysis.
- **A4:** Ensure that the data do not violate the assumptions of the tests used on them.
- **A5:** Apply appropriate quality control procedures to verify the results.

5. Presentation of Results

Goal

- Ensure that the reader of the results can understand the objective, the process and the results of experiment

5. Presentation of Results

- **P1:** Describe or cite a reference for all procedures used. Report or cite the statistical package used.
- **P2:** Present quantitative results as well as significance levels. Quantitative results should show the magnitude of effects and the confidence limits.
- **P3:** Present the raw data whenever possible. Otherwise, confirm that they are available for review by the reviewers and independent auditors.
- **P4:** Provide appropriate descriptive statistics.
- **P5:** Make appropriate use of graphics.

6. Interpretation of Results

Goal

- Ensure that the conclusions are derived merely from the results of the experiment

6. Interpretation of Results

- **I1:** Define the population to which inferential statistics and predictive models apply.
- **I2:** Differentiate between statistical significance and practical importance.
- **I3:** Specify any limitations of the study.