Chapter 4
Empirical Investigation

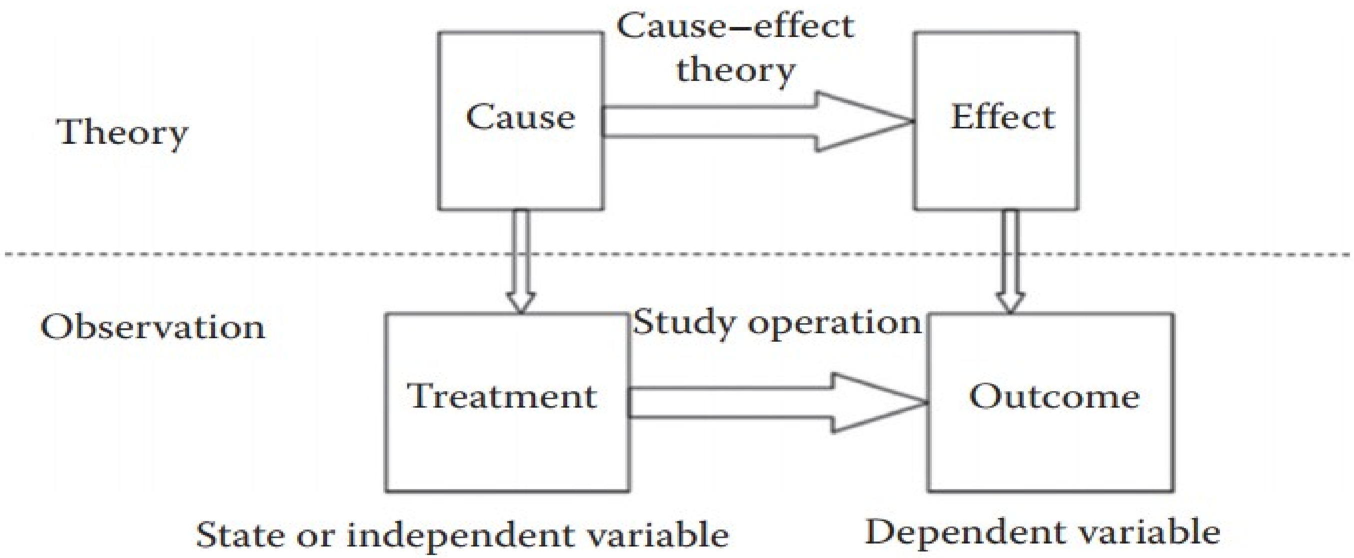
Prepared by: Mintesinot A.

SE Investigation

- What is software engineering investigation?
- Applying "scientific" principles and techniques to investigate properties of software and software related tools and techniques.
- Why talking about software engineering investigation?
- Because the standard of empirical software engineering research is quite poor.
- "the scientific method is merely a formalization of learning by experience"

SE Investigation(cont.)

- An empirical study examines some specific sample or observation of all of the possible values of the variables involved in a ***cause–effect relationship***.



SE Investigation: Examples

- Experiment to confirm rules-of-thumb
  - Should the LOC in a module be less than 200?
- 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.)
- ✍ For better processes, tools, and languages to gather and analyze requirements, model designs, develop, test, and evolve applications.

SE Investigation: What?

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

SE Investigation: How?

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

SE Investigation Principles

There are key principles of investigation:

1. **Selecting investigation technique:** conducting surveys, case studies, formal experiments

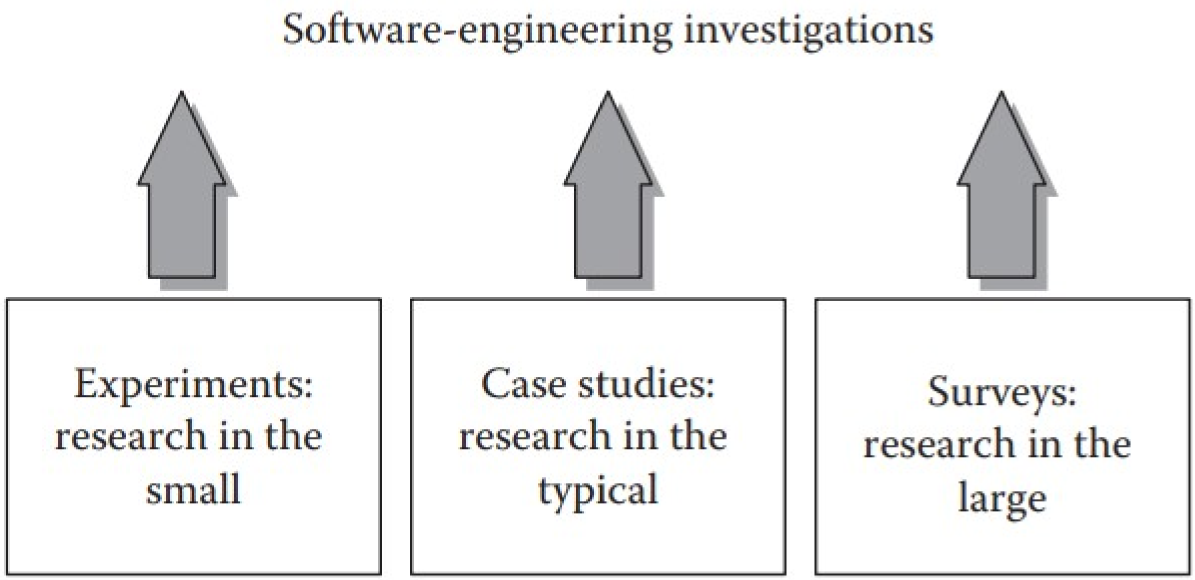
2. **Stating Goal & hypothesis:** What should be investigated?

3. **Maintaining control over variables:** dependent and independent variables

4. **Treats to validity:** No study is perfect; there are many ways that a study can provide misleading results.

5. **Human subjects:** government and university regulations require that research involving human subjects meet specified standards.

Selecting investigation technique(cont.)



Three types of investigation.

Selecting investigation technique(cont.)

Factors Relating to Choice of Research Technique

Factor	Experiments	Case Studies
Level of control	High	Low
Difficulty of control	Low	High
Level of replication	High	Low
Cost of replication	Low	High

2. Stating Goal & 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.
- The goal for your research can be expressed as a hypothesis that you want to test.
- The *hypothesis* is the tentative idea that you think explains the behavior you want to explore.

Stating Goal & hypothesis(cont.)

Examples:

- Eg1. Can integrated development and testing tools improve our productivity?
- Eg2. Does Cleanroom software development produce better-quality software than using the conventional development methods?
- Eg3. Does code produced using Agile software development have a lower number of defects per KLOC than code produced using the conventional methods?
- Eg4.“Using Scrum produces better quality software than using the Extreme Programming method.”

Stating Goal & hypothesis(cont.)

- The method used is to first evaluate a *null hypothesis*, which states that the proposed relationship does not hold.
    - The null hypothesis relevant to Example 4
    - Hyp0: "*There is no difference between the quality of software produced by the Scrum method and the quality of the software produced by using the XP method as indicated by defects per thousand lines of code*".
- Reject the null hypothesis only if there is less than a 5% chance that there is no difference between the two groups.

Stating Goal & hypothesis(cont.)

- Only after your analysis shows conclusively that you can reject the null hypothesis, you can evaluate ***alternative hypotheses***:
  - *HypA1: “The code produced using Scrum will have fewer defects per thousand lines of code than the code produced using the XP method.”*
  - *HypA2: “The code produced using XP will have fewer defects per thousand lines of code than the code produced using the Scrum method.”*

3. Maintaining Control over Variables

- What variables may affect truth of a hypothesis? How do they affect it?
- **Variable:**
  - ***Independent/ State variables***(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.

3. Maintaining Control over Variables(cont.)

- A state variable is used to distinguish the *control* situation from the *treatment* in a controlled experiment.
- When you cannot differentiate control from treatment, you must do a case study instead of a controlled experiment.

4. Threats to Validity

- There are many ways that a study can provide misleading results.
- Potential problems with empirical studies are classified as categories of *threats to validity*.
- There are four categories of threats to validity:

i.Conclusion validity- Using the wrong statistical tests, having too small a sample, Searching for relationships between too many variables

ii.Construct validity- use meaningful measures that have been validated in the narrow sense

4. Threats to Validity(cont.)

iii. ***Internal validity-*** Internal validity refers to the cause–effect relationship between independent and dependent variables.
   - A study has internal validity if the treatment actually caused the effect shown in the dependent variables.
   - Specific threats include the effects of other, possibly unidentified, variables.

iv. ***External validity-*** External validity refers to how well you can generalize from the results of one study to the wider world.
   - The ability to generalize depends on how similar the study environment is to the environment used in actual practice.

5. Human Subjects

- **Human subject** means a living individual about whom an investigator (whether professional or student) conducting research obtains

i. Data through intervention or interaction with the individual

ii. Identifiable private information

- Software engineering studies rarely involve risks of physical harm to human subjects.
  - The major risks are due to privacy issues.

Planning Experiments

- A Process Model for Performing Experiments: six-phase process

1. **Conception**

- Defining the goal of experiment
- To ensure that a controlled experiment, case study, and/or survey are appropriate.

2. **Design**

- Generating *quantifiable* (and *manageable*) hypotheses to be tested
- Defining experimental objects or units
- To plan how the application of these conditions test your hypothesis and answer your objective question

Planning Experiments(cont.)

3. **Preparation**
   - Getting ready to start, e.g., purchasing tools, configuring hardware, training personnel, etc.
   - Instructions must be written out or recorded properly.
4. **Execution**
   - Conduct the experiment
   - Following the steps laid out in the plan
   - Measuring attributes as prescribed by the plan, you apply the treatment to the experimental objects.

Planning Experiments(cont.)

5. **Analysis**
   - The analysis phase has two parts
     i. Review all the measurements taken to make sure that they are valid and useful.
     ii. Analyze the sets of data according to the statistical principles
6. **Dissemination & decision making**
   i. Documenting conclusions to duplicate experiments & confirm conclusions in a similar setting.
   ii. To support decisions how to develop or maintain software in the future

Key Experimental Design Concepts

- Simple designs - make the experiment practical, minimize time, money, personnel, easier to analyze & experimental resources.
- The three key experimental design concepts: Replication, Randomization, & Local control.

Replication:

- Involves repeating an experiment under identical conditions, rather than repeating measurements on the same experimental unit.
  - To know how much confidence we can place in the results of the experiment.
  - Replication enables us to estimate the mean effect of any experimental factor.

Key Experimental Design Concepts(cont.)

II. **Randomization:**

- The random assignment of subjects to groups or of treatments to experimental units, so that we can assume independence (conclusion & internal validity) of results.

III. **Local control:**

II. Local control refers to the control that you have over the placement of subjects in experimental units and the organization of those units.

- ***Blocking*** means allocating experimental units to groups so that the units within a block are relatively homogeneous.
- ***Balancing*** is assignment of treatments – an equal number of subjects is assigned to each treatment.

Types of Experimental Designs

- An independent variable is called a *factor* in the experimental design
- Various values or classifications for each factor are called the *levels* of the factor. *Levels* can be continuous or discrete, quantitative or qualitative.
- Eg. A study to determine the effect of experience and language on the productivity of programmers has two factors: experience and language. The dependent variable is productivity.
- Most designs in software engineering research are based on two simple relations between factors: <u>**Crossing**</u> and <u>**Nesting**</u>

Types of Experimental Designs(cont.)

1. **Crossing:**
2. Expressing the design in terms of factors, called the *factorial design*
3. How many different treatment combinations are required.
4. Two factors, A and B, in a design are *crossed* if each level of each factor appears with each level of the other factor. This relationship is denoted as $A \times B$

		Factor B		
		Level 1	Level 2	Level 3
Factor A	Level 1	a_1b_1	a_1b_2	a_1b_3
	Level 2	a_2b_1	a_2b_2	a_2b_3

Types of Experimental Designs(cont.)

2. Nesting:

- Factor B is *nested* within factor A if each *meaningful* level of B occurs in conjunction with only one level of factor A.
- The relationship is depicted as $B(A)$, where B is the nested factor and A is the nest factor.

Factor A		
Level 1		Level 2
Factor B		Factor B
Level 1	Level 2	Level 3
a_1b_1	a_1b_2	a_2b_3

- Levels 1 and 2 of B occur only with level 1 of A,
- level 3 of B only occurs with level 2 of A.
- Thus, B is nested within A. By nesting we have reduced the number of treatment combinations from 6 to 3

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.

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.