HUMAN MACHINE INTERACTION

ING. OTTO PARRA GONZALEZ

ABRIL 019

EMPIRICAL EVALUATION

HUMAN MACHINE INTERACTION

Introduction

- Software engineering research has had a tremendous impact on software development in the past decades, contributing with advances in processes, for example with agile methods, in debugging methods, e.g., with integrated development environments, and in tools, in particular with refactorings.
- It has also contributed to improving the quality of software systems, bridging research and practice, by formalising, studying, and popularising good practices.

Introduction

- Software engineering research acknowledged early that software engineering is fundamentally an empirical discipline—thus further distinguishing computer science from software engineering—because (1) software is immaterial and does not obey physical laws and (2) software is written by people for people. Therefore, many aspects of software engineering are, by definition, impacted by human factors.

Introduction

- Many aspects of software engineering are, by definition, impacted by human factors. Empirical studies are needed to identify these human factors impacting software engineering and to study the impact of these factors on software development and software systems.

Definition of Empirical Evaluation

- An **evaluation** method in which results are derived by observation or experiment instead of theory.
- **Evaluation** methodology that employs users to interact with the system.
- A study where the research ends are based on evidence and not just theory. This is done to comply with the scientific method that asserts the objective discovery of knowledge based on verifiable facts of evidence. This includes observing a phenomenon under laboratory conditions (e.g., in a controlled experiment) or in the field (e.g., in a case study).

Empirical software engineering

- Empirical software engineering is a field that encompasses several research methods and endeavours, including—but not limited to—surveys to collect data on some phenomenon and controlled experiments to measure the correlation between variables.
- Therefore, empirical studies in software engineering are studies that use any of the "usual" empirical research methods:
  - survey (including systematic literature reviews),
  - case studies,
  - quasi experiments, and
  - controlled experiments.

Empirical Software Engineering

- It is nowadays rare for software engineering research works not to include some empirical studies.
- Indeed, empirical studies are among the top most popular topics in conferences like the ACM/IEEE International Conference on Software Engineering.

Empirical software engineering

- Software engineering research strives to follow the scientific method to offer sound and well-founded results and it thus requires observations and experimentations to create and to assess hypotheses and theories.
- Empirical research offers the methods needed to perform these observations and experimentations.
  - Observations can usually be readily made in real conditions, hence rooting empirical research into industry practices.
  - Experimentations may be more difficult to carry but allow to compare activities and tasks.
- On the other hand, empirical research has a long tradition in science, in particular in fields like medicine and–or physics. Therefore, empirical researchers can draw inspiration from other fields and apply known successful method

General concepts

- The **scientific method** typically consists in observations, measurements, and experimentations. Observations help researchers to formulate important questions about a phenomenon under study. From these questions, researchers derive hypotheses that can be tested through experimentations to answer the questions.
- Once the hypothesis are tested, researchers compile and communicate the results of the experiments in the form of laws or theories, which may be universal truths and which may still require testing. The scientific method is increasingly applied in software engineering to build laws and theories about software development activities.

General concepts

- At the heart of the scientific method are **empirical methods**, which leverage evidences obtained through observations, measurements, or experimentations, to address a scientific problem. Although empirical methods are not the only means to discover (universal) truths, they are often used by researchers
- A fundamental principle of the scientific method is that results must be backed by evidences and, in software engineering, such evidences should be based on concrete observations, measurements, or experimentations resulting from qualitative and quantitative research, often based on a multimethod or mixed-method methodology.

General concepts

- Observations, measurements, or experimentations are used to formulate or infirm/confirm hypotheses. These hypotheses must be refutable. The **refutability of an hypothesis** is the possibility to prove this hypothesis to be false. Without this possibility, an hypothesis is only a statement of faith without any scientific basis.

General concepts

- **Qualitative research** aims to understand the reasons (i.e., "why") and mechanisms (i.e., "how") explaining a phenomenon. A popular method of qualitative research is case-study research, which consists in examining a set of selected samples in details to understand the phenomenon illustrated by the samples.
- For example, a qualitative study can be conducted to understand why developers prefer one particular static analysis tool over some other existing tools. For such a study, researchers could perform structured, semi-structured, or unstructured interviews with developers about the tools. They could also run focus groups and–or group discussions. They could also analyse data generated during the usage of the tools to understand why developers prefer the tool.

General concepts

- **Quantitative research** is a data-driven approach used to gain insights about an observable phenomenon.
- In quantitative research, data collected from observations are analyzed using mathematical/statistical models to derive quantitative relationships between different variables capturing different aspects of the phenomenon under study.
- For example, if we want to investigate the relation between the structure of a program and its reliability measured in terms of post-release defects, we must define a set of variables capturing different characteristics of the program, such as the complexity of the code, or the cohesion of the code, then, using a linear regression model, we can quantify the relationship between code complexity values and the number of post-release defects.

General concepts

- Experiments are used by researchers to examined cause–effect relationships between different variables characterising a phenomenon. Experiments allow researchers to verify, refute, or validate hypotheses formulated about the phenomenon. Researchers can conduct two main types of experiments:
  - controlled experiments and
  - quasi-experiments.

General concepts

- In medicine, nursing, psychology and other similar fields of empirical research, acknowledging the tradition established by medicine, researchers often study the effect of a treatment on some subjects in comparison to another group of subjects not receiving the treatment. This latter group of subjects is called the controlled group, against which researchers measure the effect of the treatment.

General concepts

- Software engineering researchers also perform **controlled experiments** but, differently from other fields of empirical research, they are rarely interested by the effect of a treatment (a novel process, method, or tool) on some subjects but rather by the effect—and possibly the magnitude of the effect—of the treatment on the performance of the participants.
- Consequently, software engineering researchers often control external measures of the participants' performance, like the numbers of introduced bugs or the numbers of failed test cases.

General concepts

- In **quasi-experiments**, researchers do not or cannot assign randomly the participants in the experiments in the control and experimental groups, as they do in experiments.
- For example, researchers could perform a (quasi-)experiment to characterise the impact of anti-patterns on the comprehensibility of systems. They should select some systems containing instances of some anti-patterns. They would create "clean" versions of the systems without these instances of the anti-patterns. They would use these versions of the systems to compare participants' performance during comprehension activities.

General concepts

- While experiments are conceived by researchers and performed to produce new knowledge (observations, results), there is a special form of experiments that is essential to the scientific method: replication experiments. **Replication experiments** are experiments that reproduce (or quasi-reproduce) previous experiments with the objectives to confirm or infirm the results from previous experiments or to contrast previous results in different contexts.
- They are essential to the scientific method because, without reproducibility, experiments provide only a collection of unrelated, circumstantial results while their reproductions give them the status of universal truth.