Chapter 6
Experiment Process

Experimentation is not simple; we have to prepare, conduct and analyze experiments properly. One of the main advantages of an experiment is the control of, for example, subjects, objects and instrumentation. This ensures that we are able to draw more general conclusions. Other advantages include ability to perform statistical analysis using hypothesis testing methods and opportunities for replication. To ensure that we make use of the advantages, we need a process supporting us in our objectives in doing experiments correctly (the notion of experiments include quasi-experiments, unless clearly stated otherwise). The basic principles behind an experiment are illustrated in Fig. 6.1.

The starting point is that we have an idea of a cause and effect relationship, i.e. we believe that there is a relationship between a cause construct and an effect construct. We have a theory or are able to formulate a hypothesis. A hypothesis means that we have an idea of, for example, a relationship, which we are able to state formally in a hypothesis.

In order to evaluate our beliefs, we may use an experiment. The experiment is created, for example, to test a theory or hypothesis. In the design of the experiment, we have a number of treatments (values that the studied variable can take, see below) over which we have control. The experiment is performed and we are able to observe the outcome. This means that we test the relationship between the treatment and the outcome. If the experiment is properly set up, we should be able to draw conclusions about the relationship between the cause and the effect for which we stated a hypothesis.

The main objective of an experiment is mostly to evaluate a hypothesis or relationship, see also Sect. 2.4.1. Hypothesis testing normally refers to the former, and the latter is foremost a matter of building a relational model based on the data collected. The model may be derived using multivariate statistical methods, for example, regression techniques and then we evaluate it in an experiment. The focus in this book is primarily on hypothesis testing. Multivariate statistical methods are treated by, for example, Kachigan [90, 91] and Manly [118].

The experiment process presented in this chapter is formulated to make sure that the proper actions are taken to ensure a successful experiment. It is unfortunately not

C. Wohlin et al., *Experimentation in Software Engineering*,

DOI 10.1007/978-3-642-29044-2_6,

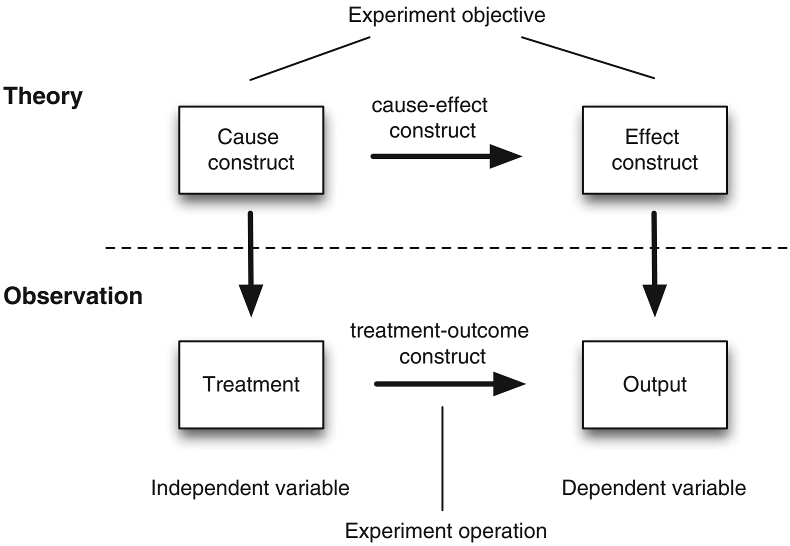

Fig. 6.1 Experiment principles (Adapted from Trochim [171])

uncommon that some factor is overlooked before the experiment, and the oversight prevents us from doing the planned analysis and hence we are unable to draw valid conclusions. The objective, of having a process, is to provide support in setting up and conducting an experiment. The activities in an experiment are briefly outlined in this chapter and treated in more detail in the following chapters, see Chaps. 7–11.

6.1 Variables, Treatments, Objects and Subjects

Before discussing the experiment process, it is necessary to introduce a few definitions in order to have a vocabulary for experimentation. When conducting a formal experiment, we want to study the outcome when we vary some of the input variables to a process. There are two kinds of variables in an experiment, independent and dependent variables, see Fig. 6.2.

Those variables that we want to study to see the effect of the changes in the independent variables are called *dependent variables* (or response variables). Often there is only one dependent variable in an experiment. All variables in a process that are manipulated and controlled are called *independent variables*.

Example. We want to study the effect of a new development method on the productivity of the personnel. We may have chosen to introduce an object-oriented design method instead of a function-oriented approach. The *dependent variable* in

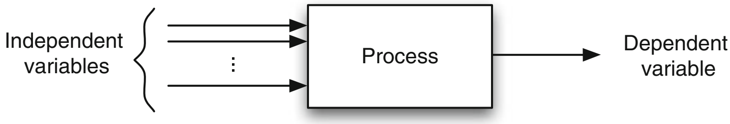

Fig. 6.2 Illustration of independent and dependent variables

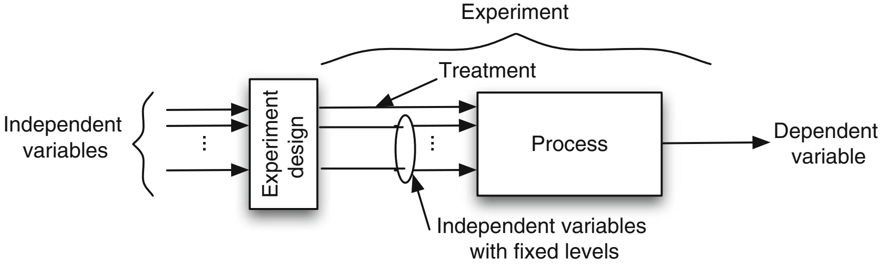

Fig. 6.3 Illustration of an experiment

the experiment is the productivity. *Independent variables* may be the development method, the experience of the personnel, tool support, and the environment.

An experiment studies the effect of changing one or more independent variables. Those variables are called *factors*. The other independent variables are controlled at a fixed level during the experiment, or else we cannot say if the factor or another variable causes the effect. A *treatment* is one particular value of a factor.

Example. The factor for the example experiment above, is the development method since we want to study the effect of changing the method. We use two treatments of the factor: the old and the new development method.

The choice of treatment, and at which levels the other independent variable shall have, is part of the experiment design, see Fig. 6.3. Experiment design is described in more detail in Chap. 8.

The treatments are being applied to the combination of *objects* and *subjects*. An object can, for example, be a document that shall be reviewed with different inspection techniques. The people that apply the treatment are called *subjects*.[1] The characteristics of both the objects and the subjects can be independent variables in the experiment.

[1]Sometimes the term *participant* is used instead of the term subject. The term subject is mainly used when people are considered with respect to different treatments and with respect to the analysis and the term participant mainly when it deals with how to engage and motivate people in a study.

Example. The *objects* in the example experiment are the programs to be developed and the *subjects* are the personnel.

An experiment consists of a set of *tests* (sometimes called trials) where each test is a combination of treatment, subject and object. It should be observed that this type of test should not be confused with the use of statistical tests, which is further discussed in Chap. 10. The number of tests affects the experimental error, and provides an opportunity to estimate the mean effect of any experimental factor. The experimental error helps us to know how much confidence we can place in the results of the experiment.

Example. A *test* can be that person *N* (*subject*) uses the new development method (*treatment*) for developing program *A* (*object*).

In human-oriented experiments, humans are the subjects, applying different treatments to objects. This implies several limitations to the control of the experiment. Firstly, humans have different skills and abilities, which in itself may be an independent variable. Secondly, humans learn over time, which means that if one subject applies two methods, the order of application of the methods may matter, and also the same object cannot be used for both occasions. Thirdly, human-oriented experiments are impacted by all sorts of influences and threats, due to the subject's ability to guess what the experimenter expects, their motivation for doing the tasks etc. Hence it is critical for the outcome of the experiment how subjects are selected and treated.

Technology-oriented experiments are easier to control, since the technology may be made deterministic. The independent variable out of control in this type of experiments may instead be the objects selected for the experiment. One tool or technique may be well suited for one type of programs, and not for another. Hence it is critical for the outcome how objects are selected.

6.2 Process

A process provides steps that support an activity, for example, software development. Processes are important as they can be used as checklists and guidelines of what to do and how to do it. To perform an experiment, several steps have to be taken and they have to be in a certain order. Thus, a process for how to perform experiments is needed.

The process presented is focused on experimentation, but the same basic steps must be performed in any empirical study, as illustrated for the case study process in Sect. 5.1.2. The main difference is the work within a specific activity, for example, the design of a survey, experiment and case study differ, but they all need to be designed. Further, as case studies are flexible design studies, there are several iterations over the process steps, while experiments and surveys, as fixed design studies, primarily execute the steps once. Thus, the basic process may be used for

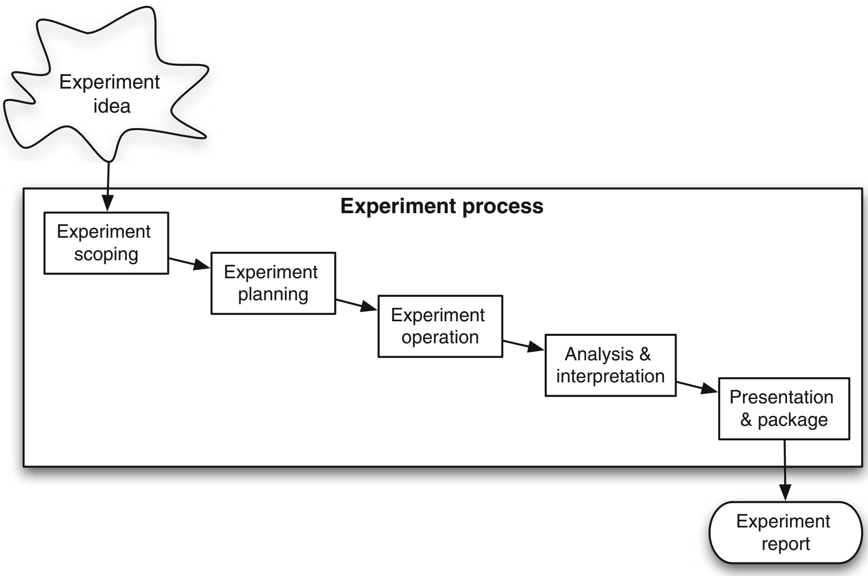

Fig. 6.4 Overview of the experiment process

other types of studies than experiments, but it has to be tailored to the specific type of study being conducted, for example, a survey using e-mail or a case study of a large software project. The process is as it is presented, however, suited for both randomized experiments and quasi-experiments. The latter are often used in software engineering when random samples of, for example, subjects (participants) are infeasible.

The starting point for an experiment is insight, and the idea that an experiment would be a possible way of evaluating whatever we are interested in. In other words, we have to realize that an experiment is appropriate for the question we are going to investigate. This is by no means always obvious, in particular since empirical studies are not frequently used within computer science and software engineering [170, 181]. Some argumentation regarding why computer scientist should experiment more is provided by Tichy [169]. If we assume that we have realized, that an experiment is appropriate then it is important to plan the experiment carefully to avoid unnecessary mistakes, see Sect. 2.9.

The experiment process can be divided into the following main activities. *Scoping* is the first step, where we scope the experiment in terms of problem, objective and goals. *Planning* comes next, where the design of the experiment is determined, the instrumentation is considered and the threats to the experiment are evaluated. *Operation* of the experiment follows from the design. In the operational activity, measurements are collected which then are analyzed and evaluated in *analysis and interpretation*. Finally, the results are presented and packaged in *presentation and package*. The activities are illustrated in Fig. 6.4 and further elaborated below, and then each of the activities is treated in-depth in Chaps. 7–11. An overview of the experiment process including the activities, is presented in Fig. 6.5.

The process is not supposed to be a 'true' waterfall model; it is not assumed that an activity is necessarily finished prior to that the next activity is started. The order of activities in the process primarily indicates the starting order of the activities. In other words, the process is partly iterative and it may be necessary to go back and refine a previous activity before continuing with the experiment. The main exception is when the operation of the experiment has started, then it is not possible to go back to the scoping and planning of the experiment. This is not possible since starting the operation means that the subjects are influenced by the experiment, and if we go back there is risk that it is impossible to use the same subjects when returning to the operation phase of the experiment process.

Scoping. The first activity is scoping. The hypothesis has to be stated clearly. It does not have to be stated formally at this stage, but it has to be clear. Furthermore, the objective and goals of the experiment must be defined. The goal is formulated from the problem to be solved. In order to capture the scope, a framework has been suggested [13]. The framework consists of the following constituents:

- Object of study (what is studied?),
- Purpose (what is the intention?),
- Quality focus (which effect is studied?),
- Perspective (whose view?), and
- Context (where is the study conducted?).

These are further discussed in Chap. 7.

Planning. The planning activity is where the foundation for the experiment is laid. The context of the experiment is determined in detail. This includes personnel and the environment, for example, whether the experiment is run in a university environment with students or in an industrial setting. Moreover, the hypothesis of the experiment is stated formally, including a null hypothesis and an alternative hypothesis.

The next step in the planning activity is to determine variables (both independent variables (inputs) and the dependent variables (outputs). An important issue regarding the variables is to determine the values the variables actually can take. This also includes determining the measurement scale, which puts constraints on the method that we later can apply for statistical analysis. The subjects of the study are identified.

Furthermore, the experiment is designed, which includes choosing a suitable experiment design including, for example, randomization of subjects. An issue closely related to the design is to prepare for the instrumentation of the experiment. We must identify and prepare suitable objects, develop guidelines if necessary and define measurement procedures. These issues are further discussed in Chap. 8.

As a part of the planning, it is important to consider the question of validity of the results we can expect. Validity can be divided into four major classes: internal,

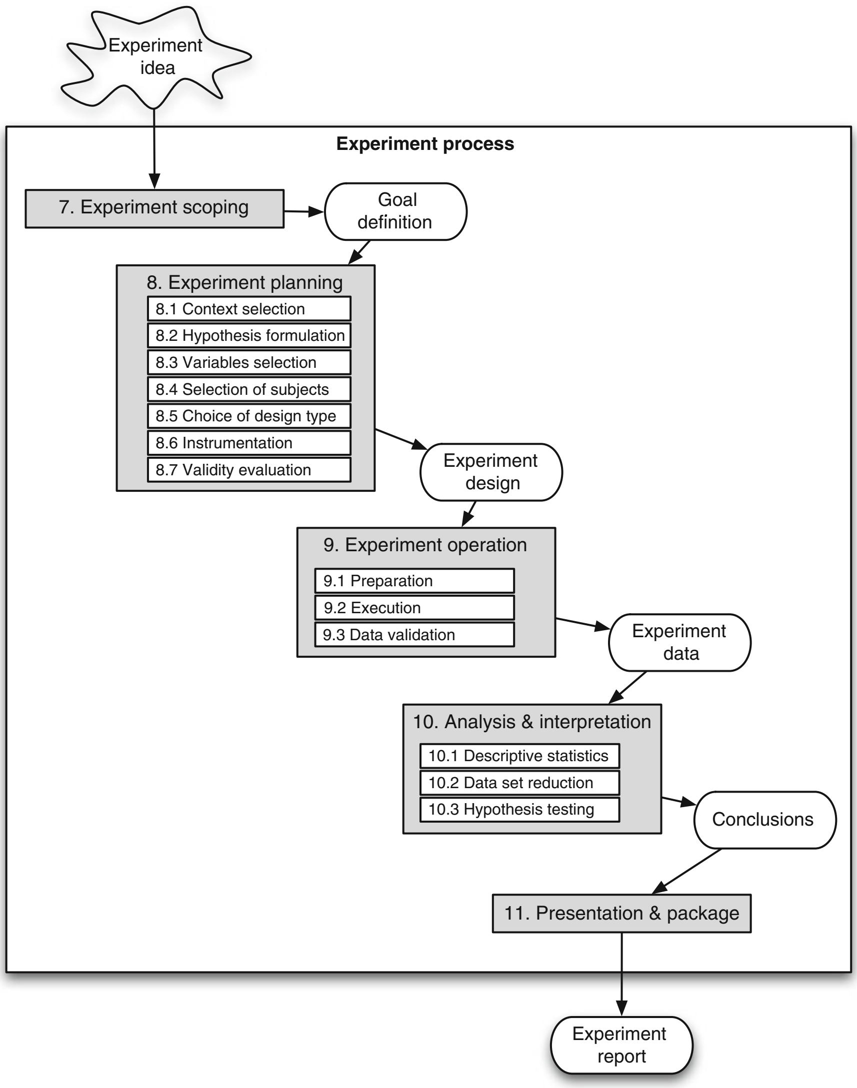

Fig. 6.5 Overview of the experiment process and artefacts with references to chapters and sections of this book

external, construct and conclusion validity. Internal validity is concerned with the validity within the given environment and the reliability of the results. The external validity is a question of how general the findings are. Many times, we would like to state that the results from an experiment are valid outside the actual context in which the experiment was run. The construct validity is a matter of judging if the treatment reflects the cause construct and the outcome provides a true picture of the effect construct, see Fig. 6.1. The conclusion validity is concerned with the relationship between the treatment and the outcome of the experiment. We have to judge if there is a relationship between the treatment and the outcome.

The planning is a crucial step in an experiment to ensure that the results from the experiment become useful. Poor planning may ruin any well-intended study.

Operation. The operation consists in principle of three steps: preparation, execution and data validation. In the preparation step, we are concerned with preparing the subjects as well as the material needed, for example, data collection forms. The participants must be informed about the intention; we must have their consent and they must be committed. The actual execution is normally not a major problem. The main concern is to ensure that the experiment is conducted according to the plan and design of the experiment, which includes data collection. Finally, we must try to make sure that the actually collected data is correct and provide a valid picture of the experiment. The operation activity is discussed in Chap. 9.

Analysis and interpretation. The data collected during operation provide the input to this activity. The data can now be analyzed and interpreted. The first step in the analysis is to try to understand the data by using descriptive statistics. These provide a visualization of the data. The descriptive statistics help us to understand and interpret the data informally.

The next step is to consider whether the data set should be reduced, either by removing data points or by reducing the number of variables by studying if some of the variables provide the same information. Specific methods are available for data reduction.

After having removed data points or reduced the data set, we are able to perform a hypothesis test, where the actual test is chosen based on measurement scales, values on the input data and the type of results we are looking for. The statistical tests together with a more detailed discussion of descriptive statistics and data reduction techniques can be found in Chap. 10.

One important aspect of this activity is the interpretation. That is, we have to determine from the analysis whether the hypothesis was possible to reject. This forms the basis for decision-making and conclusions concerning how to use the results from the experiment, which includes motivation for further studies, for example, to conduct an enlarged experiment or a case study.

Presentation and package. The last activity is concerned with presenting and packaging of the findings. This includes primarily documentation of the results, which can be made either through a research paper for publication, a lab package for replication purposes or as part of a company's experience base. This last activity is important to make sure that the lessons learned are taken care of in an appropriate way. Moreover, an experiment will never provide the final answer to a question, and hence it is important to facilitate replication of the experiment. A comprehensive and thorough documentation is a prerequisite to achieve this objective. Having said that, the use of lab packages should done with care since using the same experimental design and documents may carry over some systematic problems and biases from the original experiment, as discussed in Sect. 2.6. Independently, we must take some time after the experiment to document and present it in a proper way. The presentation of an experiment is further elaborated in Chap. 11.

6.3 Overview

The steps in this experiment process are described in more detail subsequently, and to support the understanding of the process, an example is presented in Chap. 12. The objective of the example is to closely follow the defined process in order to illustrate the use of it. A summarizing overview of the experiment process can be found in Fig. 6.5.

6.4 Exercises

6.1. What is a cause and effect relationship?

6.2. What is a treatment, and why is it sometime necessary to apply treatments in a random order?

6.3. What are dependent and independent variables respectively?

6.4. What are quasi-experiments? Explain why these are common in software engineering.

6.5. Which are the main steps in the experiment process, and why is it important to have distinct steps?

Springer
http://www.springer.com/978-3-642-29043-5

Experimentation in Software Engineering
Wohlin, C.; Runeson, P.; Höst, M.; Ohlsson, M.C.;
Regnell, B.; Wesslén, A.
2012, XXIV, 236 p., Hardcover
ISBN: 978-3-642-29043-5