Utrecht University

Faculty of Science
Department of Information and Computing Sciences

Sample-based experiments

Advanced Research Methods 2018-2019

Marcela Ruiz
Department of Information and Computing Sciences
Utrecht University
m.ruiz@uu.nl

December 4, 2018

Research project assignment

Main topic of the research assignment is: *you have selected one from the list of topics.*

For this main topic, you have selected two artefacts. You will compare these two artefacts by conducting a sample-based experiment

Example: From goal models to business process models

The *iStar2BPc-guidelines* for a top-down scenario has as a main purpose to guide the mapping from *iStar* (a goal-oriented modelling method) elements into Choreography (a BPM notation for business process modelling) elements.

The *iStar2BPc-discovery* presents a process for discovering traceability links between iStar models and Business process models.

Sample-based or case-based?

Sample-based experiments:

- Typical in domains with high (non-deterministic) variability
- Also used in engineering when validating artefacts with a concrete sample

Case-based experiments

- Typical in engineering domains
- High reproducibility
- Also used in social sciences to study many variables in one case

Example: From goal models to business process models

We are curious!

Which one is better?

Conduct a comparative experiment to evaluate the benefits and drawbacks of using the *iStar2Bc-guidelines* vs the *iStar2BPc-discovery* process. The experimental tasks are intended to compare the subjects' performance and perceptions when they use the discovery process and the guidelines.

The researchers

We are the researchers of this project!

Get familiarized with the treatments: guidelines and process
Build the experimental objects: the cases to evaluate
Design the experimental task: next week

The researchers

We are the researchers of this project!

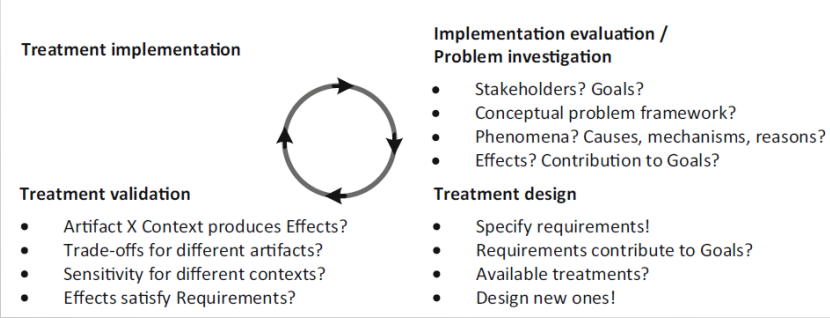
Get familiarized with the treatments: Example guidelines and process

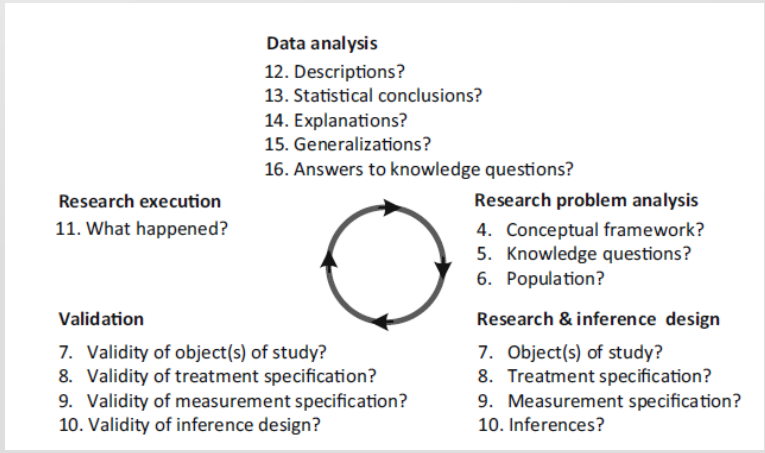
Build the experimental objects: the cases to evaluate

Design the experimental task: next week

Example: Experimental design

We have two artefacts that we want to test: guidelines and process
We want to evaluate the artefacts in context: Design science!





Design cycle: for your master thesis (maybe), we don't have the artefacts

Empirical cycle: for the ARM course, we have the artefacts!

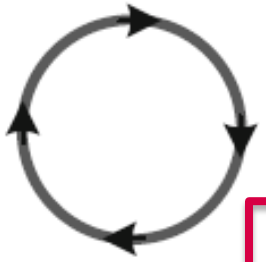
Experimental design

Data analysis

12. Descriptions?
13. Statistical conclusions?
14. Explanations?
15. Generalizations?
16. Answers to knowledge questions?

Research execution

11. What happened?



Research problem analysis

4. Conceptual framework?
5. Knowledge questions?
6. Population?

Validation

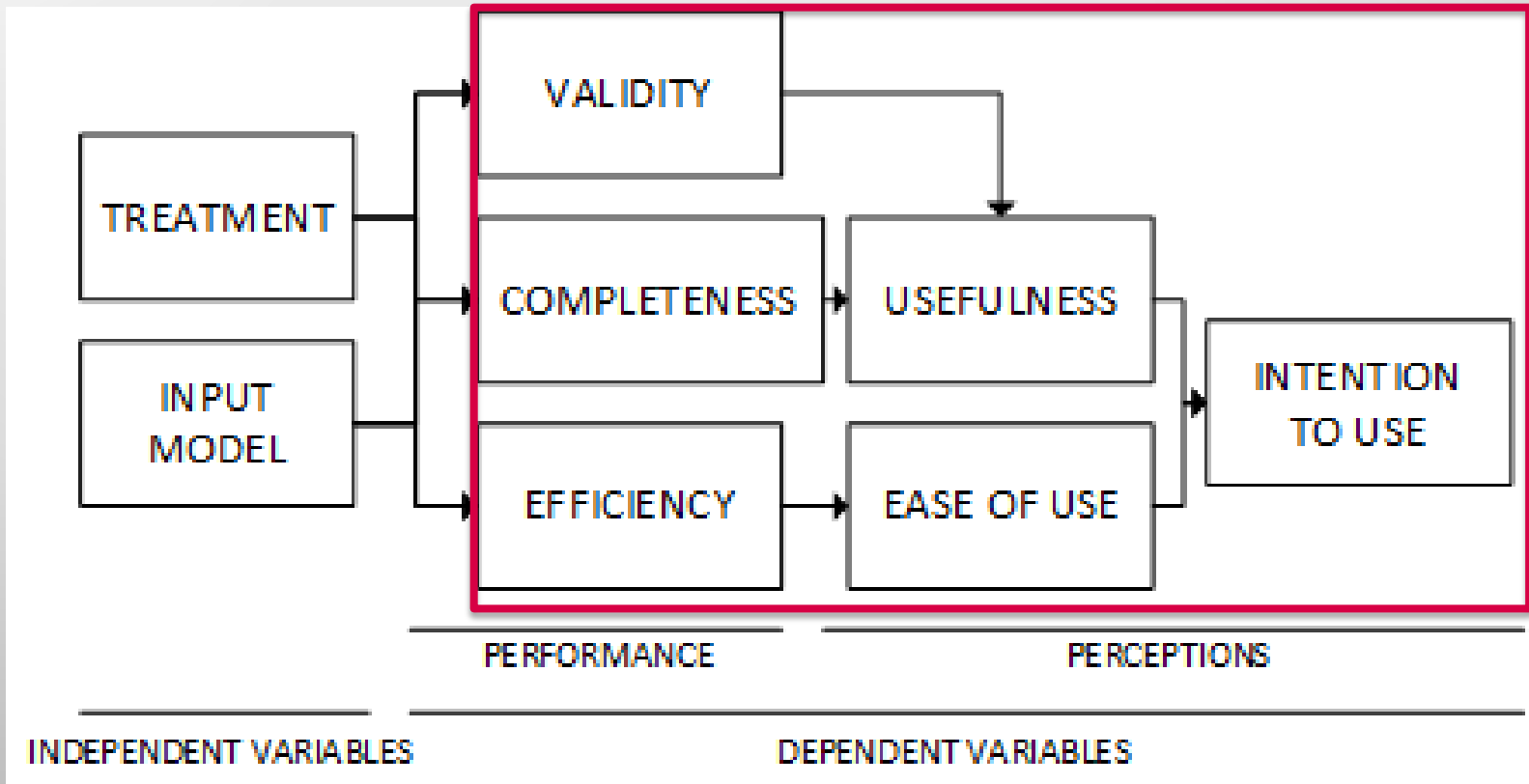
7. Validity of object(s) of study?
8. Validity of treatment specification?
9. Validity of measurement specification?
10. Validity of inference design?

Research & inference design

7. Object(s) of study?
8. Treatment specification?
9. Measurement specification?
10. Inferences?

Experimental design

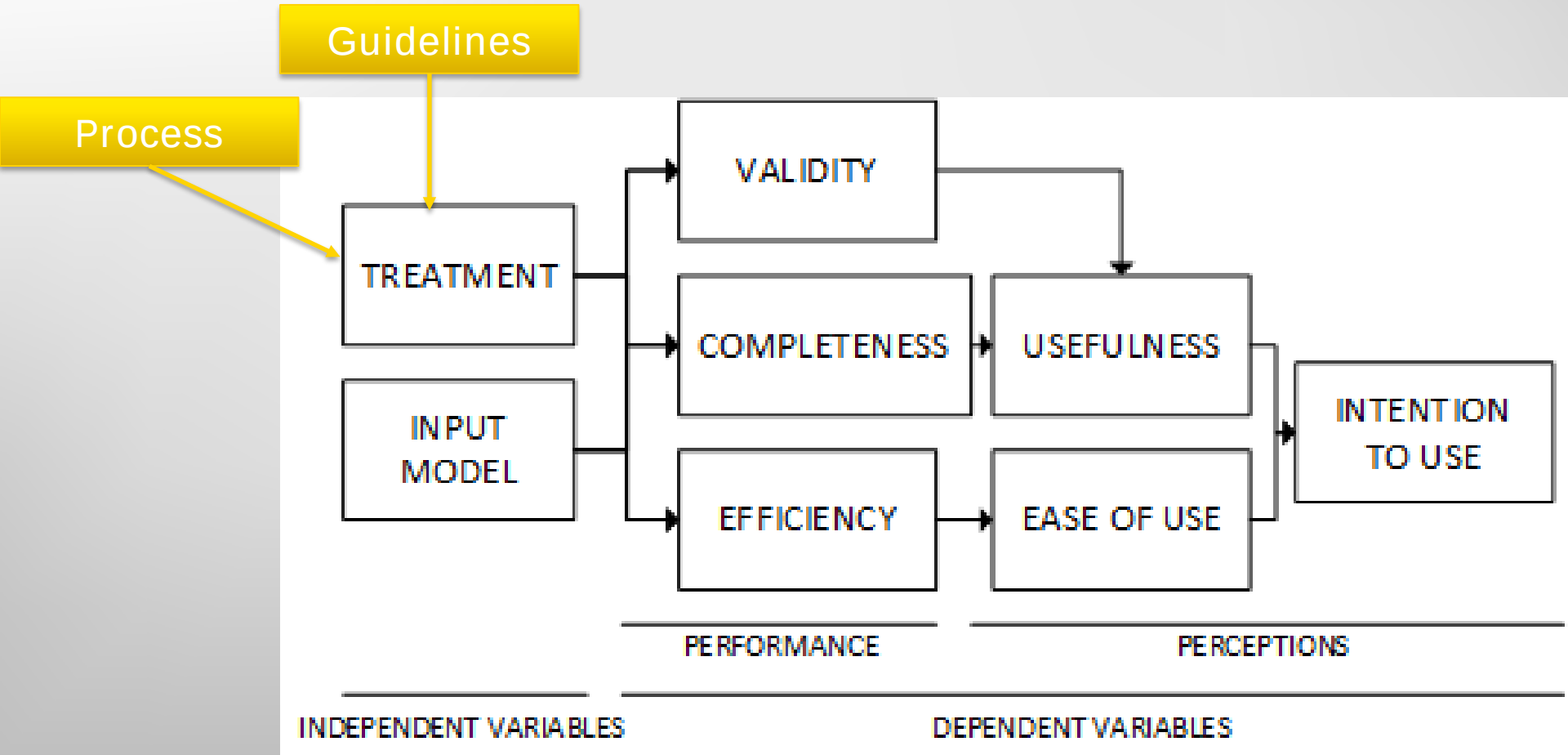
We are the researchers of this project. What to test?



Example: Experimental design

We are the researchers of this project. What to test?

Selected artefacts to test



Example; Experimental design

We are the researchers of this project. What to test?

Selected artefacts to test

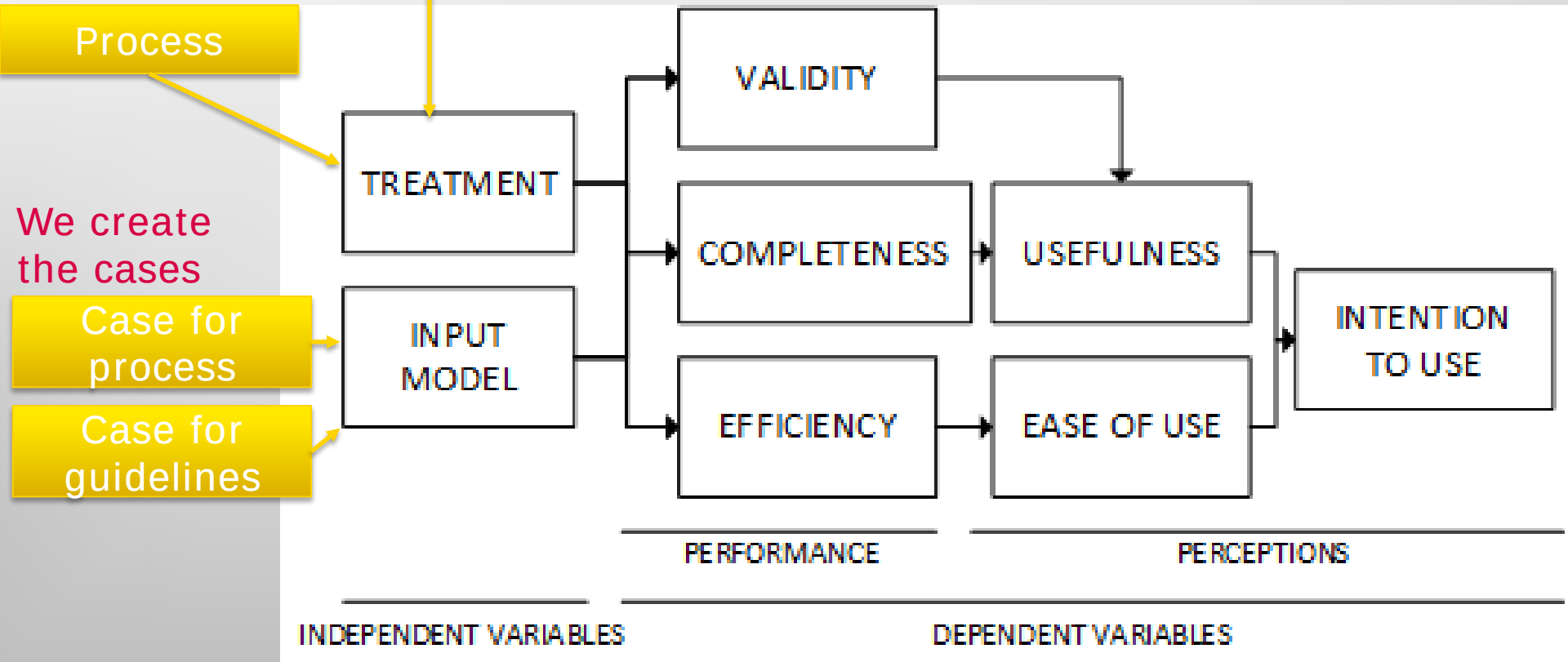
Guidelines

Process

We create the cases

Case for process

Case for guidelines



Utrecht University

We are the researchers of this project. What to test?

Experimental design

Given artefacts to test

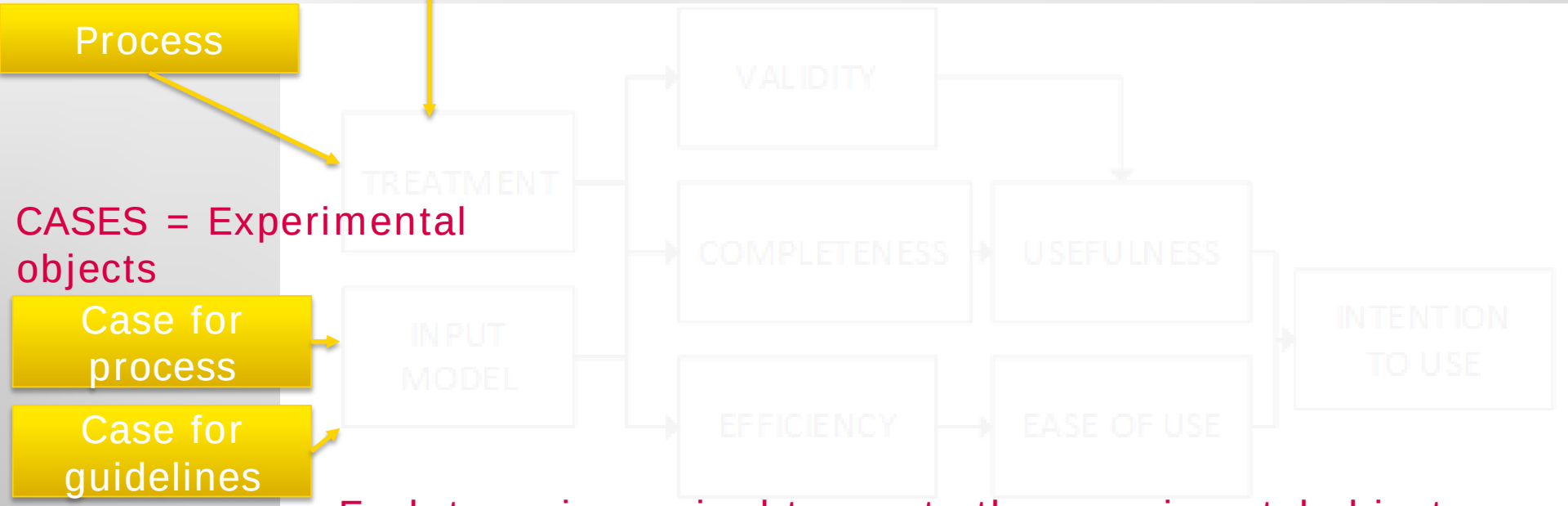
Guidelines

Process

CASES = Experimental objects

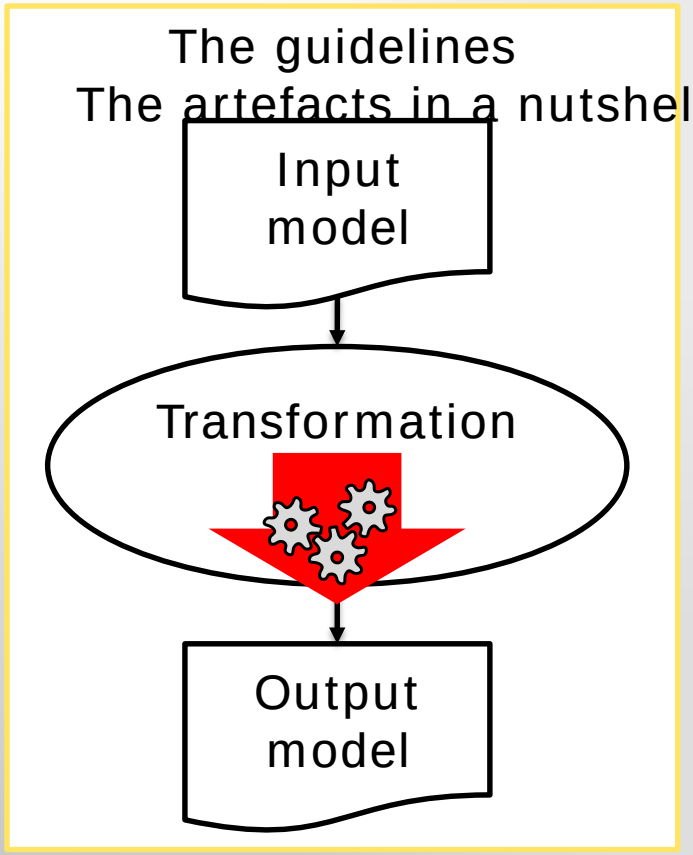
Case for process

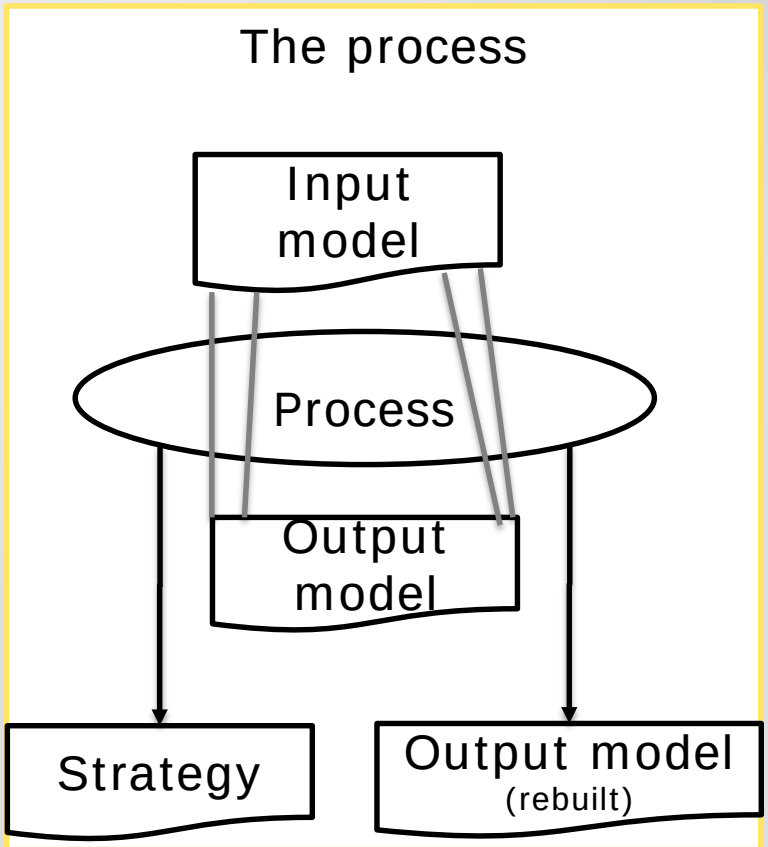
Case for guidelines



Each team is required to create the experimental objects

Example: you need to master your artefacts





Agenda

1. Object of study
2. Treatment specification
3. Measurement specification
4. Planning
   a) Context selection
   b) Hypothesis formulation
   c) Variables selection
   d) Selection of subjects
   e) Experiment design
   f) Instrumentation

Agenda

1. Object of study
2. Treatment specification
3. Measurement specification
4. Planning
   a) Context selection
   b) Hypothesis formulation
   c) Variables selection
   d) Selection of subjects
   e) Experiment design
   f) Instrumentation

Object of study

Part of the world that the researcher interacts with in order to learn something about population elements

Entity where the phenomena occur from which measurements are taken

Objects of study can be selected: How do you know that a selected entity is a population element?

Objects of study can be constructed: How do you construct a population element?

Objects of study can be selected and constructed

Example: (MDD vs traditional SW development)

"The study participants are master students with some professional experience. In the experimental investigation, subjects worked in pairs (mainly for logistic reasons). Both groups used a traditional software development method in the first session and MDD in the second session to develop a software system. The objects used two requirements specifications created for this experiment"

What is the object of study? Selected? Created? Both?

Agenda

1. Object of study
2. Treatment specification
3. Measurement specification
4. Planning
   a) Context selection
   b) Hypothesis formulation
   c) Variables selection
   d) Selection of subjects
   e) Experiment design
   f) Instrumentation

Treatment specification

An experimental treatment is a treatment of an object of study by a researcher, performed with the goal of finding out what the effects of the treatment are

It is the exposure of an object of study to a treatment in order to answer a knowledge question of the researcher

Example: (MDD vs traditional SW development)

"We now define factors and their levels to operationalize the cause of our experiment construct. Factors are variables (independent) whose effect on the dependent variables we want to understand. The experiment studies one factor: development method. The control in this experiment is a traditional method while the treatment is MDD (OO-Method and its tool)".

Which treatment(s) will be applied?
Which treatment instruments will be used? Instruction sheets, videos, lessons, software, computers, actuators, rooms, etc.
What is the treatment schedule?

What is your treatment?

Agenda

1. Object of study
2. Treatment specification
3. Measurement specification
4. Planning
   a) Context selection
   b) Hypothesis formulation
   c) Variables selection
   d) Selection of subjects
   e) Experiment design
   f) Instrumentation

Measurement specification

Measurement of a variable is the assignment, according to a rule, of a value to the phenomenon denoted by the variable

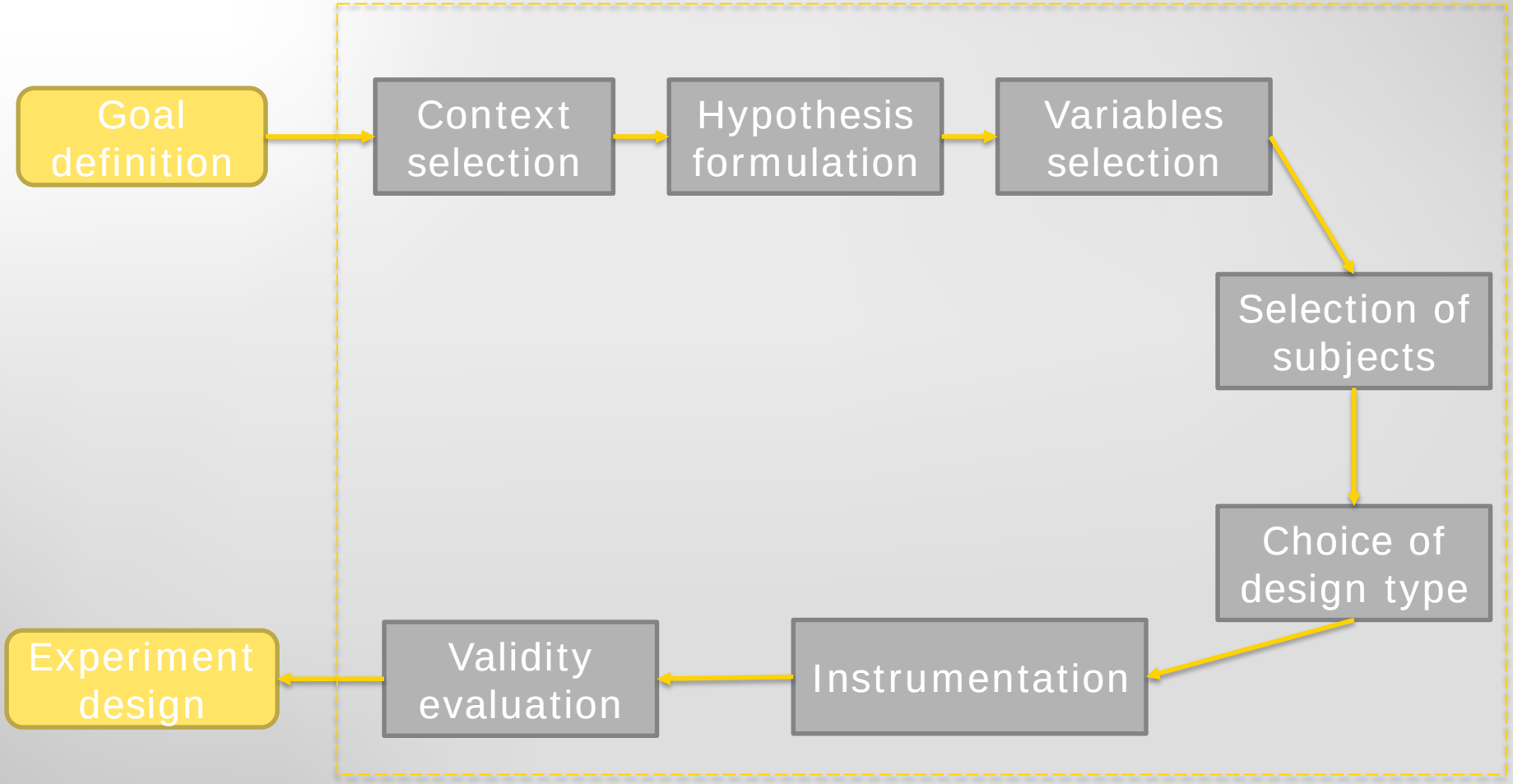
Definition of a measurement rule involves defining a measurement scale. A scale is a set of values with manipulation rules, where values and manipulation rules must have real-world meaning.

The measurement scales are:
Nominal: Different phenomena are given different laves, and we can test whether two phenomena have received the same or a different label
Ordinal: Phenomena are given ordered labels, and in addition to equality, we can test which order two labels have
Interval: Distances between numbers are meaningful, but the position of the zero is arbitrary. It tests phenomena with natural degrees of difference
Ratio: Distances between numbers are meaningful, but there is a zero

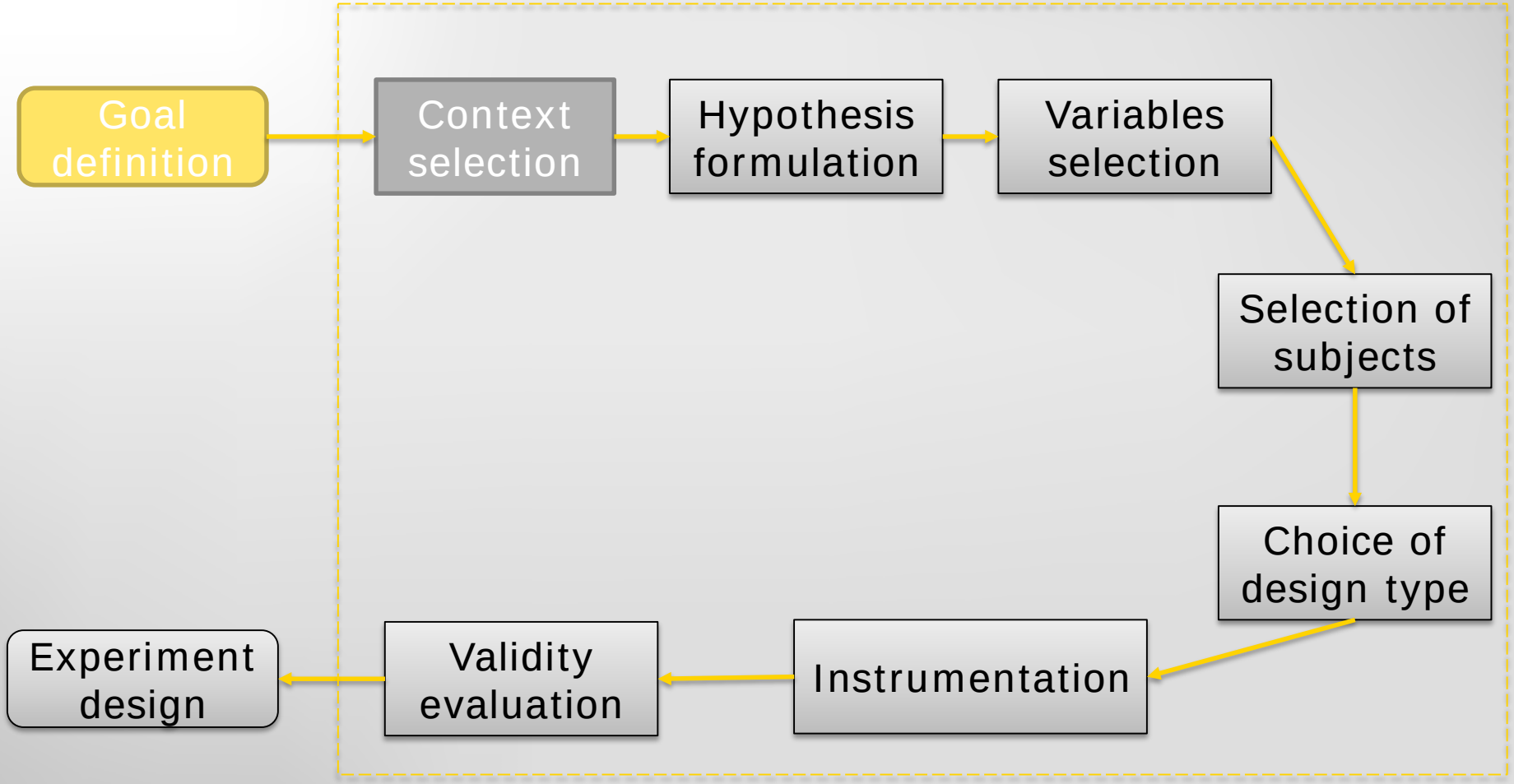
Agenda

1. Object of study
2. Treatment specification
3. Measurement specification
4. Planning
   a) Context selection
   b) Hypothesis formulation
   c) Variables selection
   d) Selection of subjects
   e) Experiment design
   f) Instrumentation

Planning phase overview



Planning phase overview



Context selection

Do you want to achieve general results in an experiment?
It should be executed in large, real software projects with professional staff. This is the ideal situation ☺

Hold on! Conducting an experiment involves risks:
The new method to be examined is not as good as expected and causes delays
Mortality of experimental subjects
High costs (time, resources, instruments)
What else could happen?

What are the options?

Context selection

Off-line vs. on-line
Student vs. professional
Toy vs. real problems
Specific vs. general

Context selection

Off-line vs. on-line
Student vs. professional
Toy vs. real problems
Specific vs. general

Context selection
Student vs. professional

Benefits

Cheap

Easy to control

Context fixed (experience and background)

Context selection
Student vs. professional

Benefits

Cheap

Easy to control

Context fixed (experience and background)

Drawbacks

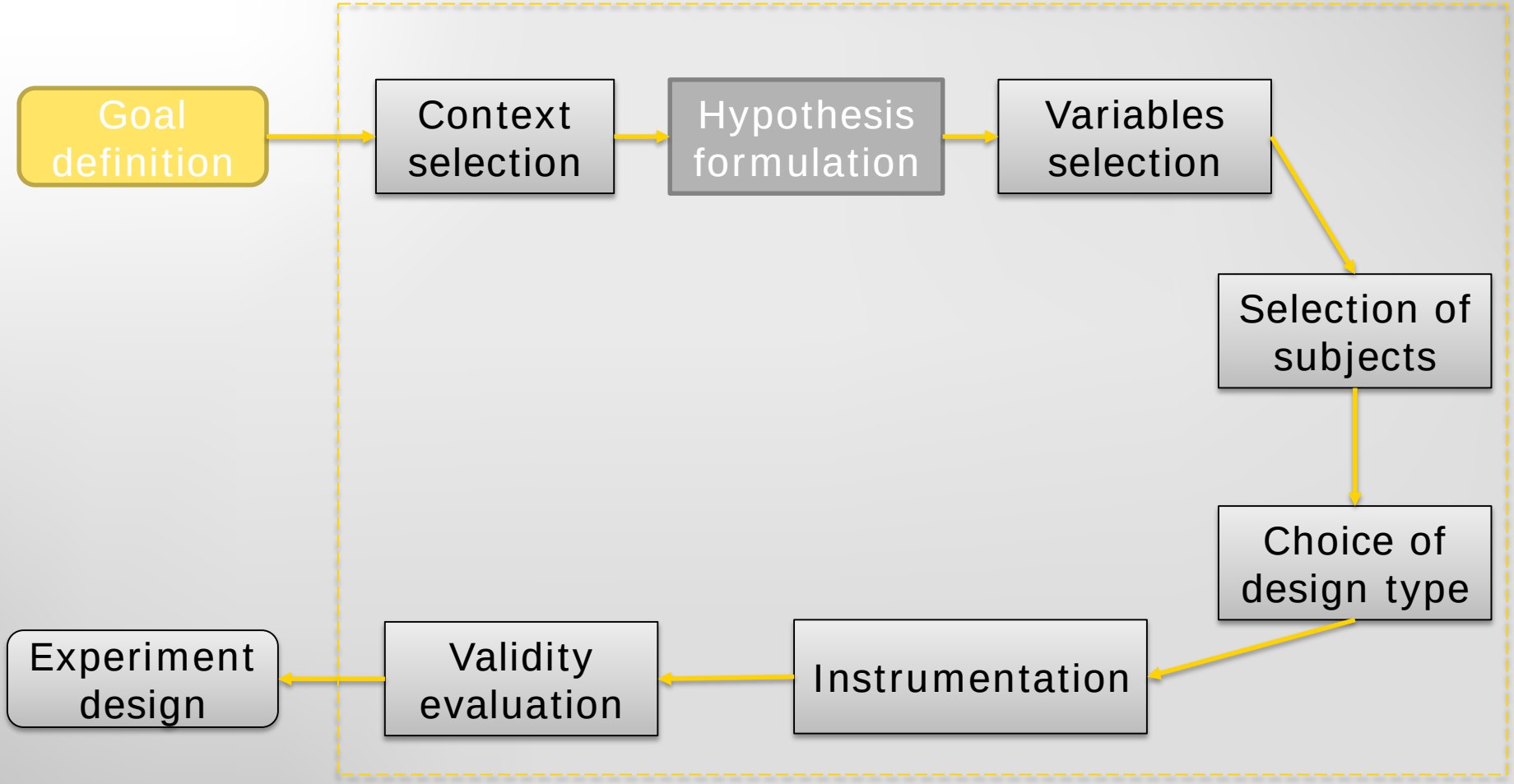
Context fixed -> difficult to generalize the results

No real-world problems (toy cases due to constrains in cost and time)

Conduct a trade-off analysis and make studies valid to a specific context of valid to the general software engineering domain

Tip! Check H¨ost,M., Regnell, B.,Wohlin, C.: Using students as subjects – a comparative study of students and professionals in lead-time impact assessment. Empir. Softw. Eng. 5(3), 201–214 (2000)

Planning phase overview



Hypothesis formulation

A hypothesis is stated formally and the data collected during the course of the experiment is used to, if possible, reject the hypothesis

Two hypotheses have to be formulated:
H0: States that there are no real underlying trends or patterns in the experiment setting; the only reasons for differences in our observations are coincidental. This is the hypothesis that the researcher wants to reject with a high significance as possible

H1: States in favor of which the null hypothesis is rejected with a high significance

In order to test the hypotheses, the statistical tools will be provided the weeks 49-50

Tip:
For each H0 and H1, one variable should be established

Example: (MDD vs traditional SW development)

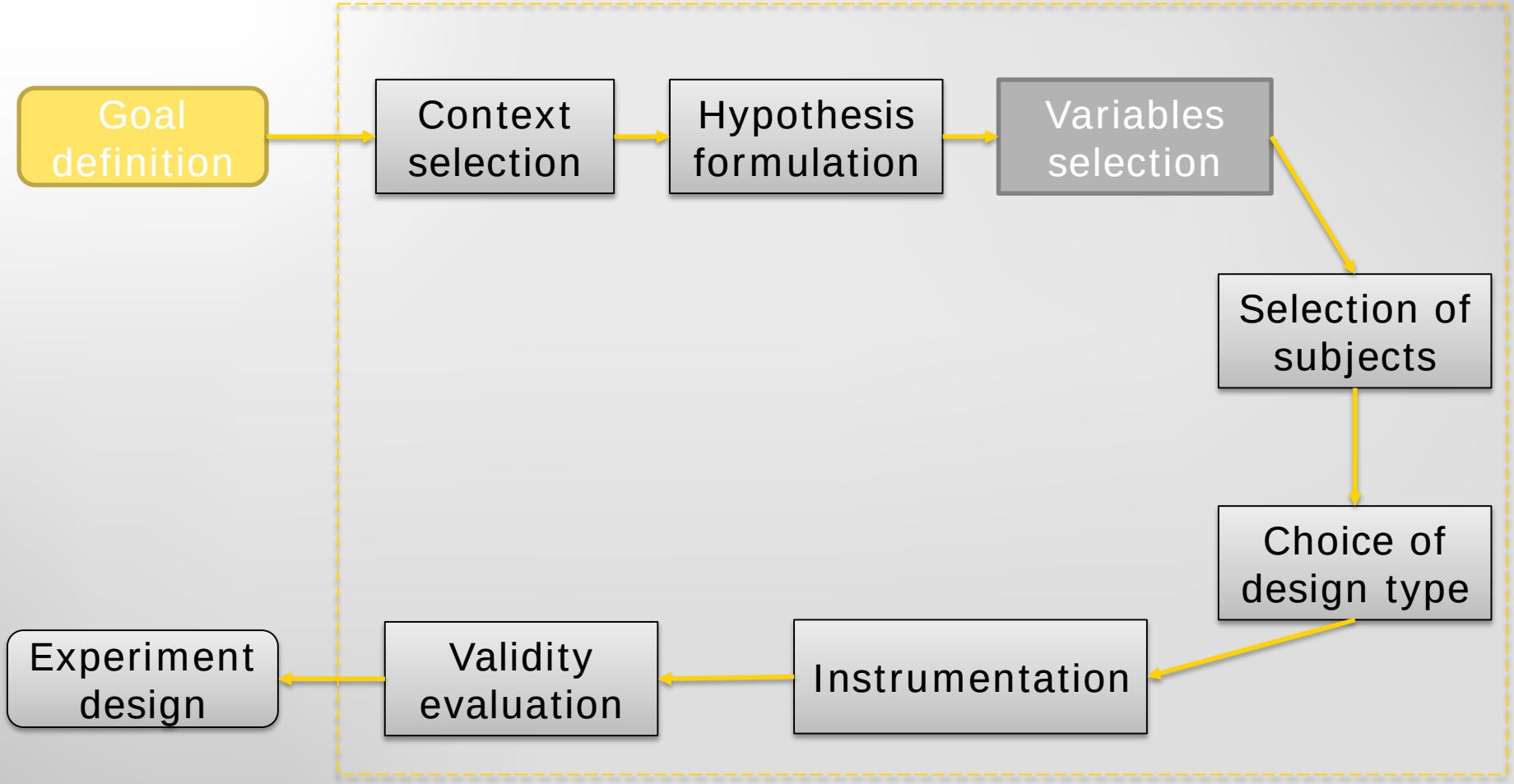
H01: The software quality of a system built using MDD is similar to software quality using a traditional method

H02: The developer effort to build a system using MDD is similar to effort using a traditional method

H03: The developer productivity using MDD to build a system is similar to productivity using a traditional method

H04: The developer satisfaction using MDD to build a system is similar to satisfaction using a traditional method

Planning phase overview



Variables selection

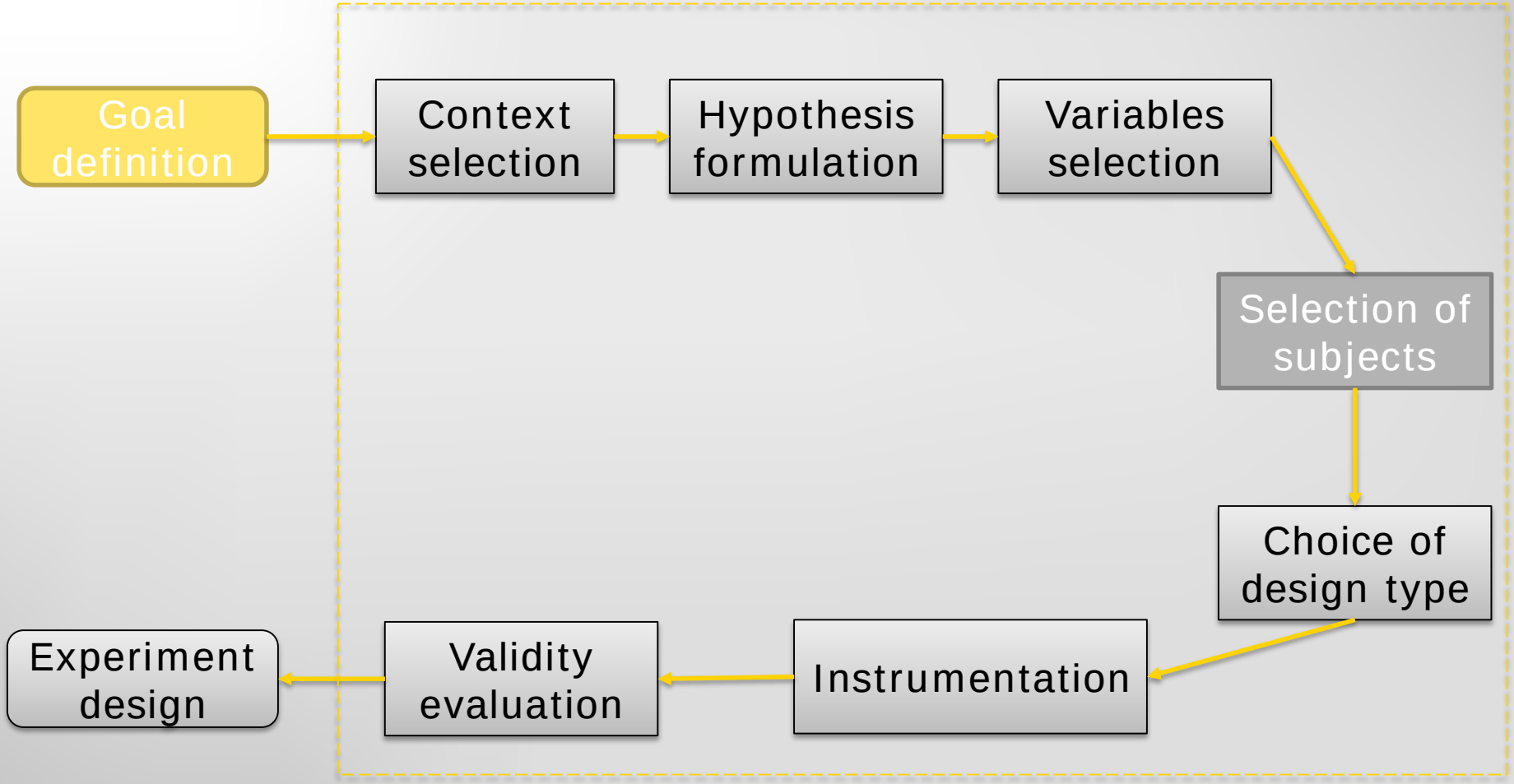
Before any design can start, it is vital to choose the dependent and independent variables

Example

RQs	Hypotheses	Response Variables	Metric
RQ1	$H_0 1$	Accuracy	Test cases passed
RQ2	$H_0 2$	Effort	Time
RQ3	$H_0 3$	Productivity	Accuracy/effort
RQ4	$H_0 4$	Satisfaction	PU, PEOU, ITU

Keep the traceability among knowledge questions, hypothesis, and independent variables

Planning phase overview



Selection of subjects

Selection of subjects impact the potential generalization of the results. The selection of subjects is also called a sample from a population.

Two types of sampling design

Probability

The probability of selecting each subject is known

Simple random sampling

Systematic sampling

Stratified random sampling

Non-probability

The probability of selecting each subject is unknown

Convenience sampling

Quota sampling

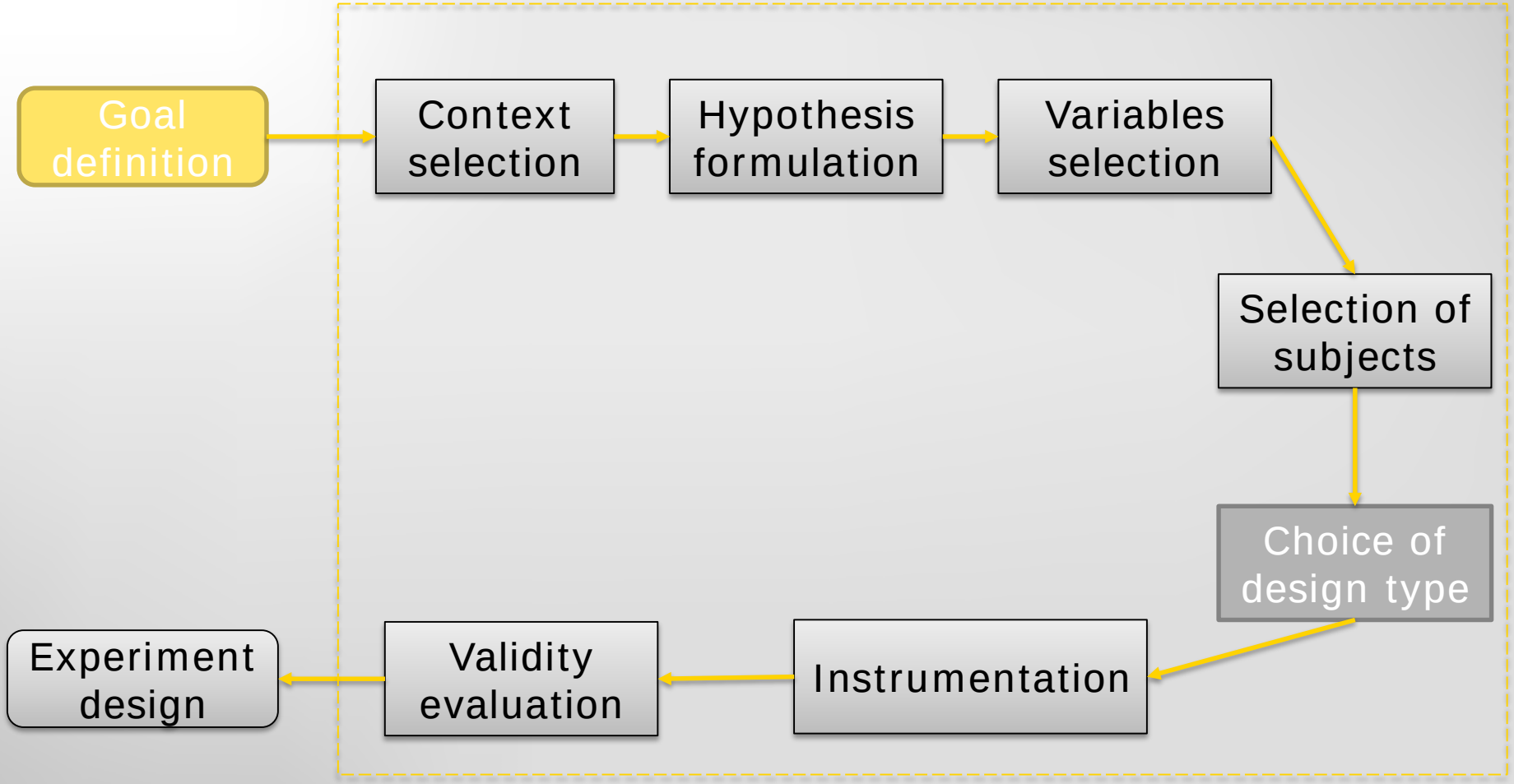
Benefits and drawbacks?

Selection of subjects

The size of the sample impacts the results when generalizing:
The larger the sample is, the lower the error becomes when generalizing the results
Big sample, more power of the statistical test to use

Tip:
This is useful to justify the selection of statistical tests

Planning phase overview



Choice of design type

Focus in the independent variables

One factor with two treatments: Compare two treatment against each other

One factor with more than two treatments: Compare the treatments against each other

Two factors: compare the treatments against each other and the factors against each other

2*2 factorial design: Two factors each with two treatments

More than two factors: compare the treatments against each other and the factors against each other

2k factorial design: k factors each with two treatments

What is your design type?

Selected artefacts to test

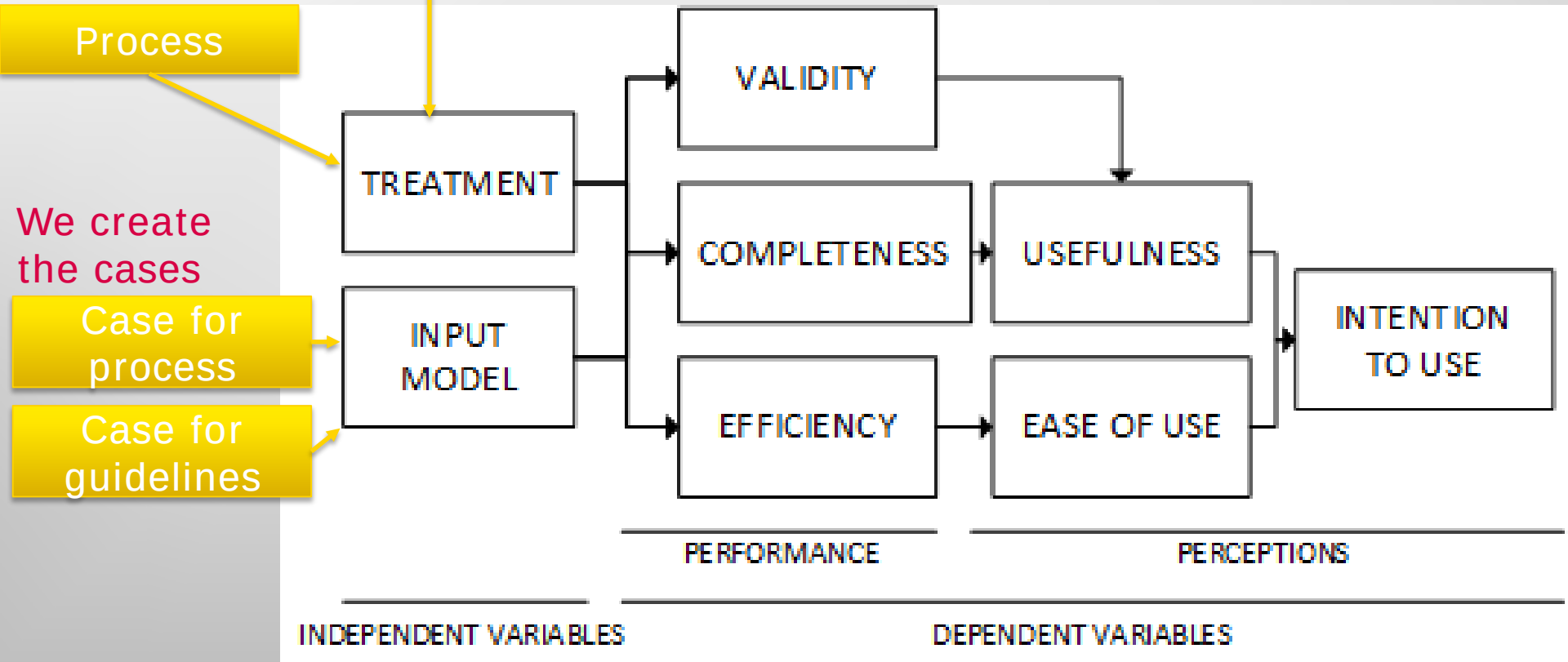
Guidelines

Process

We create the cases

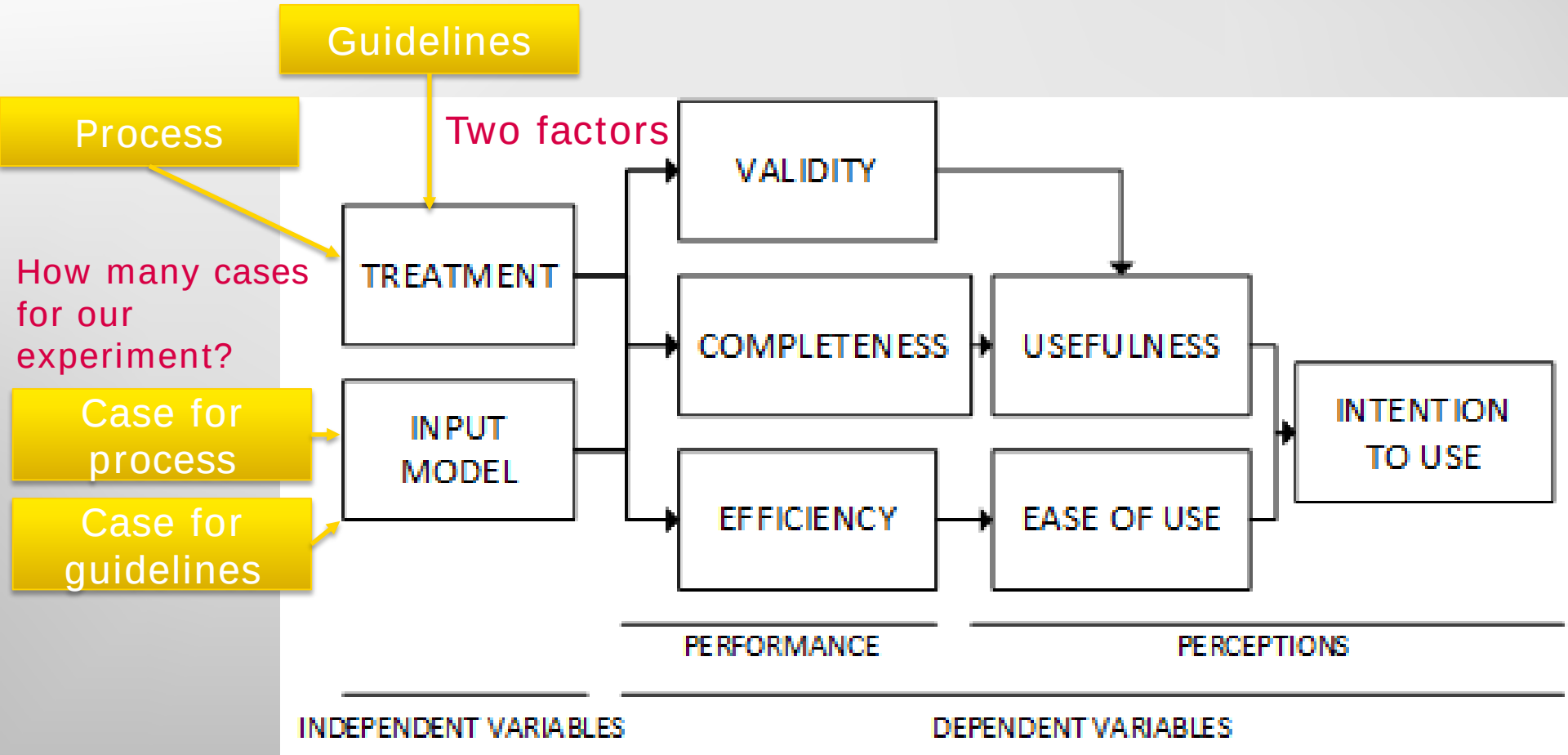
Case for process

Case for guidelines

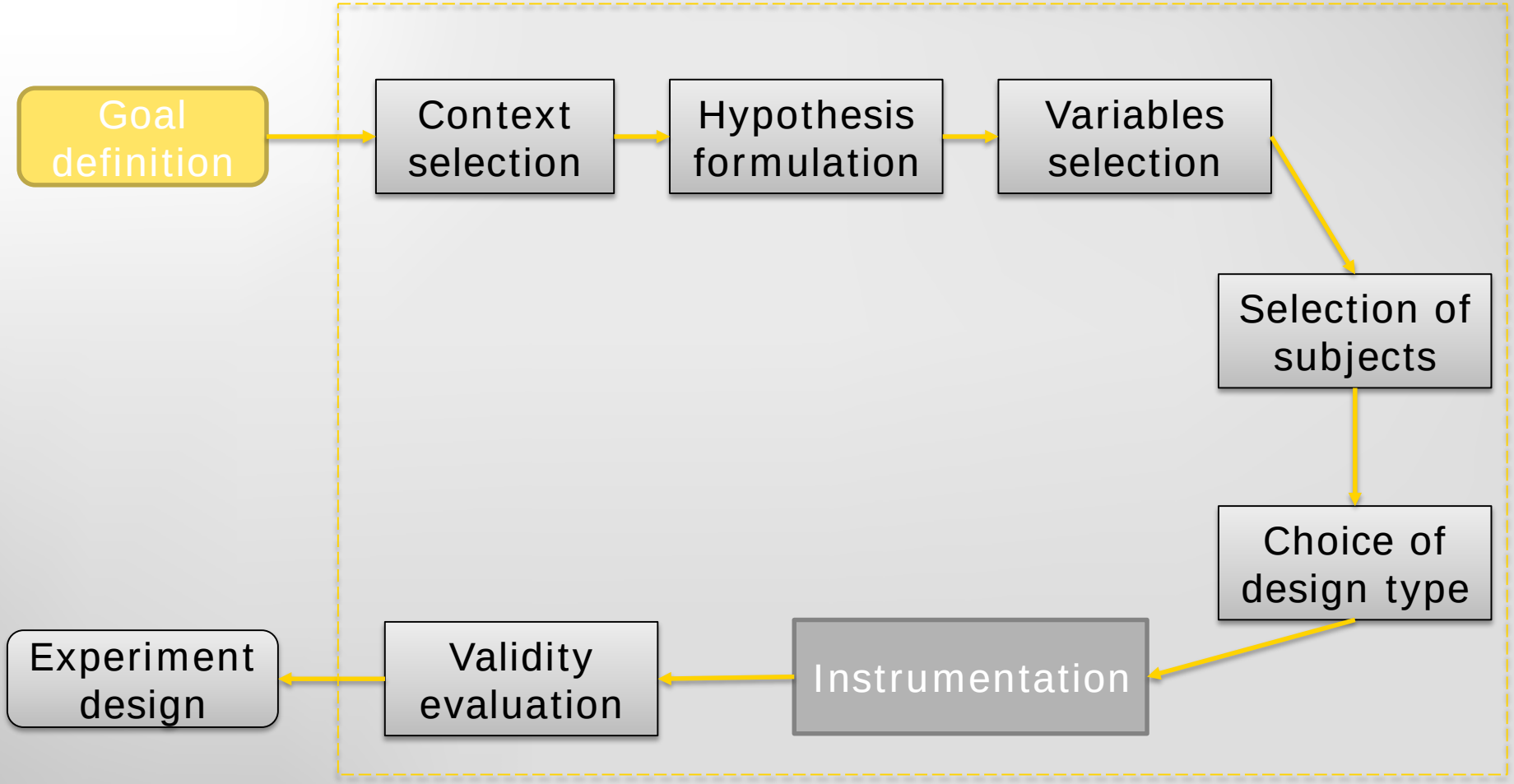


What is your design type?

Two treatments



Planing phase overview



Instrumentation

There are three types of instruments: objects, guidelines and measurement instruments

Experimental object: The objects used in the experimental investigation. E.g., models, code documents, textual descriptions, etc.

Guidelines: The guidelines to guide the subjects in the experiment. It includes process descriptions and checklists, training in the method to be used, etc.

Measurement instruments: Objects to collect data during the experimental tasks. In human experiments, data is generally collected via questionnaires, interviews, results of performed tasks, time used when performing a certain task, etc.

Experimental task

Include the experimental objects that will be evaluated
What are your experimental objects?
Present the instructions for the subjects
Guide them to solve certain tasks
Help you to collect data for experimental analysis!

Let’s check a possible template:

[Your project]:
An Empirical Validation

Dear colleague,

We would like to thank you for participating in this validation activity related to **[Your project]**. Please, read carefully the instructions below and be sure of following them as you proceed. Otherwise, the validation activity will not be valid.

Please, take into account that you are kindly requested to measure accurately the time **[a variable to measure efficiency]** it takes you to complete the activity using these 2 metrics:

- T1: The time it takes to read and understand section *1. **[Treatment]***.
- T2: The time it takes to elaborate and write down the solution. Section *2. **[Task to complete]***. If, during the elaboration of the section 2, you need to revise again the supporting material or the statement of the section 1, include this revision time into T2, not T1.

Note: in the next page starts the exercise; be aware of the previous instructions; and prepare a chronometer to measure first T1 and then T2. When you have finished the solution, stop measuring time and fill the form of final survey in this document (section 3). Afterwards, please give this document back to the researcher in charge.

If you have any doubt, do not hesitate to contact us before starting.

Yours faithfully,

The Researchers

Utrecht University

Name: _____

Student number: _____

1. [Treatment]

In this section, we introduce the **[treatment]**. Please find a short description bellow:

[Short description of the treatment]

Please, start the chronometer now for measuring T1.

[Introduce the treatment to evaluate]

[This time will help you to understand the amount of time the subject took to understand the treatment; it could be an interesting data to know]

Now you can stop your chronometer and take note of the final value for T1.

2. [Task to complete]:

Please read the following case: **[introduce the case]**. To apply the [treatment] please read the following case:

[Include here the input artefacts]

Start the chronometer now for measuring T2.

You are required to apply the [treatment] in order to obtain [output] from [input artefacts]

[The outcome will help you to measure variables related to performance, like completeness and validity]

Once completed, stop the chronometer and take note of T2.

3. Final survey

This survey is related to your **[perceptions, variable to measure perceptions]** of the application of the **[treatment]**. Please, rate the extent to which you agree with each statement.

[Survey]

[Finally, acknowledge the participation of your subjects!]

Thank you very much for your support! Please give back this document to the researcher in charge

Commit participants

Obtain consent: the participants have to be agree to the research objectives

Disclosure: Reveal the results of the experiment to the subjects

Announce all these details beforehand. It is important to state it together with the guidelines for the experiment.

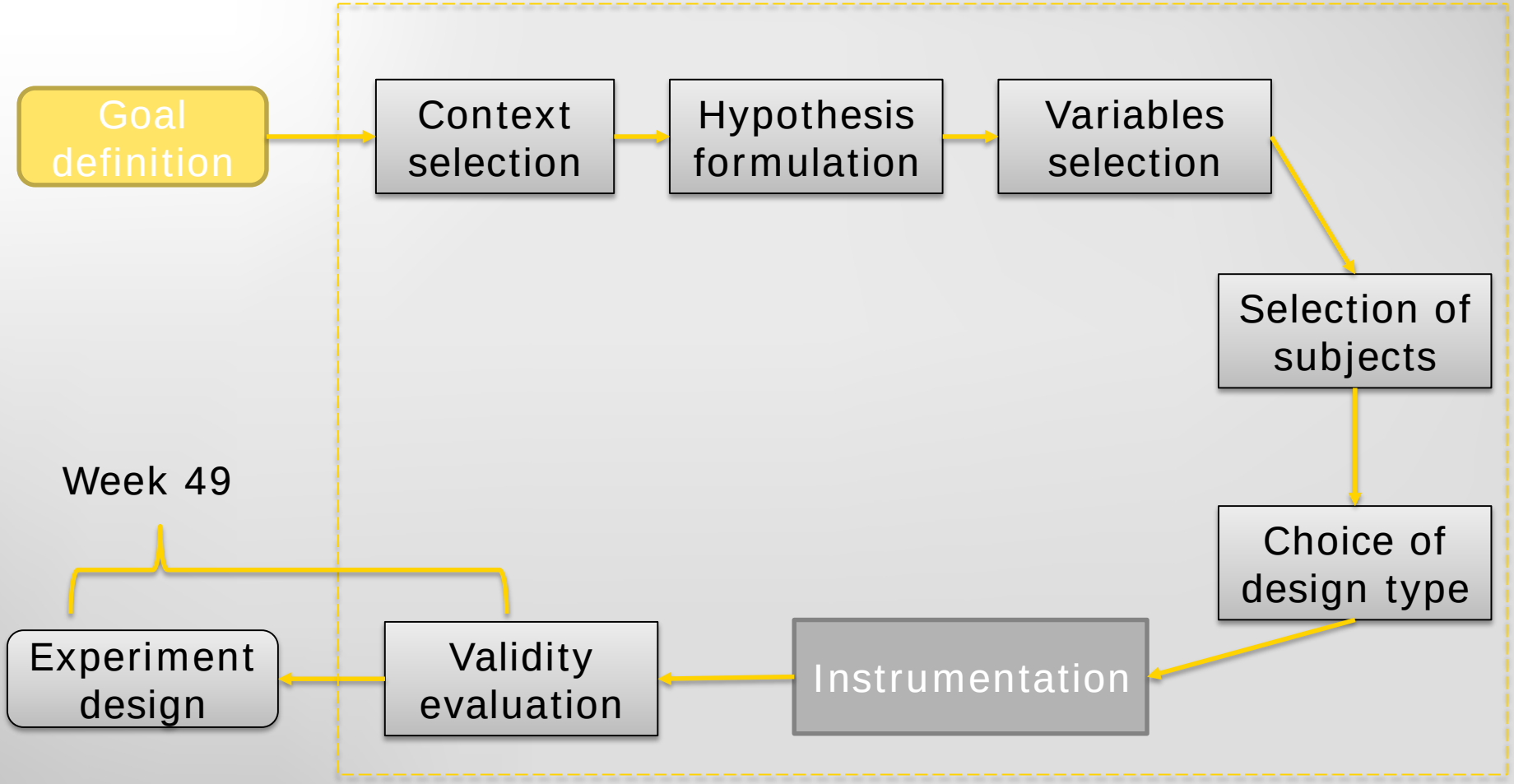
Commit participants Example

Dear Colleague,
I hope this message finds you well.
As part of our work in the SUPERSEDE H2020 project, www.supersede.eu, we are currently proposing an ontology-based approach for defining the architecture of a monitoring platform. At this point of time, we need some validation to assess whether the proposal is really bringing the value that we expect. To this aim, two of the members of the SUPERSEDE team at UPC, UPC_Researcher1 and UPC_Researcher2, have designed a validation exercise. We would very deeply grateful if you accept to participate in the study. According to our piloting of the exercise, the total time to finalize it should not exceed 30 minutes plus 10 minutes to fill and submit the form.

Please let us know if you are able to participate and send us the results by Nov. 18.

Thanks for your collaboration,

Planing phase overview



Experimental process overview

Example: (MDD vs traditional SW development)

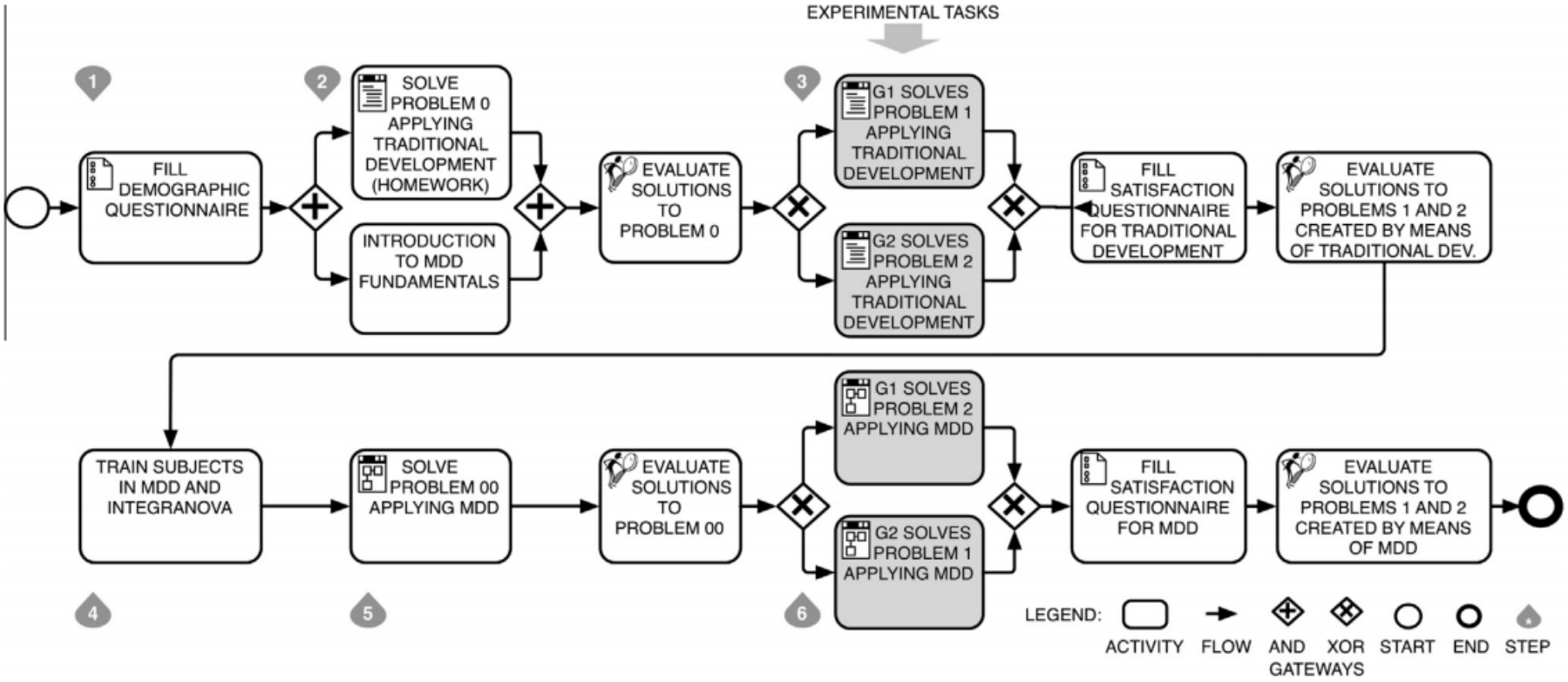


Fig. 2. Summary of how the experiment was performed.