


ESE

8. Experiment Planning

8.3 Validity Evaluation



Validity Evaluation

- Threats to the validity of an experiment
- Types of validity and their map to the experiment principia
- Basic questions concerning experiment validity
- Types of validity and their map to threats possible for them
 - Internal validity
 - External validity
 - Construct validity
 - Conclusion validity (`Statistical validity`)
- List of Validity Threats
- Conflicts and priority between types of validity threats

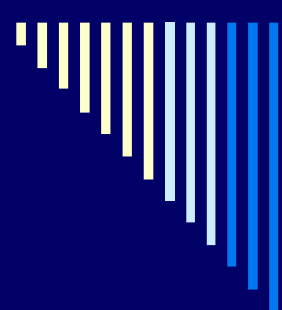


Threats to the validity of an experiment

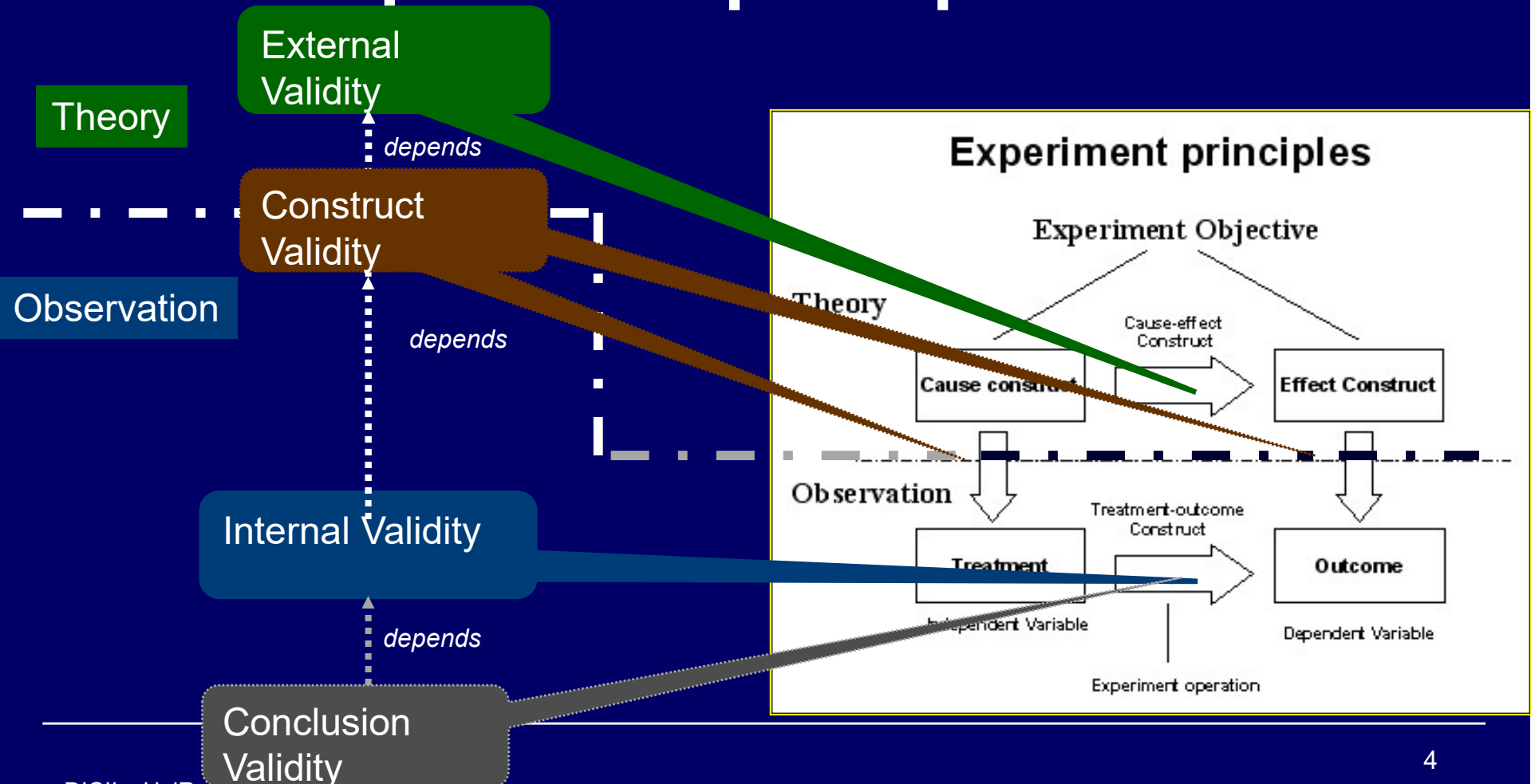
- Results should be valid for the **reference population** (from which the sample is drawn)
- We might be interested in generalizing results.

Validity has to be **planned** if we want to have an **adequate validity**, i.e., results valid for the target population at least.

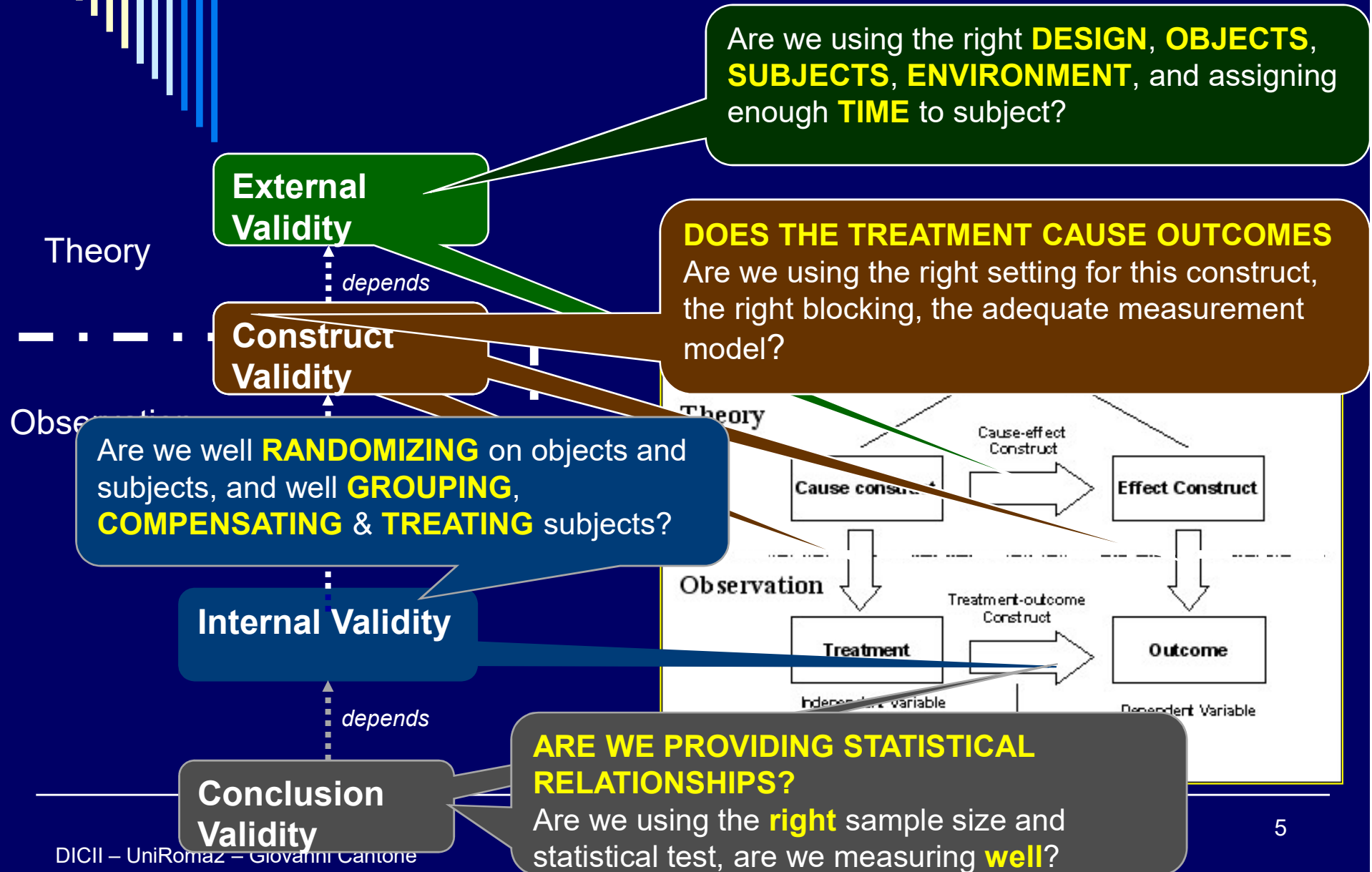
In addition to *Internal validity* and *External validity*, a modern vision of threats to validity of an experiment is also concerned *Construct validity* and *Conclusion validity*.



Types of validity and their map to relationships of the experiment principia



Basic Questions



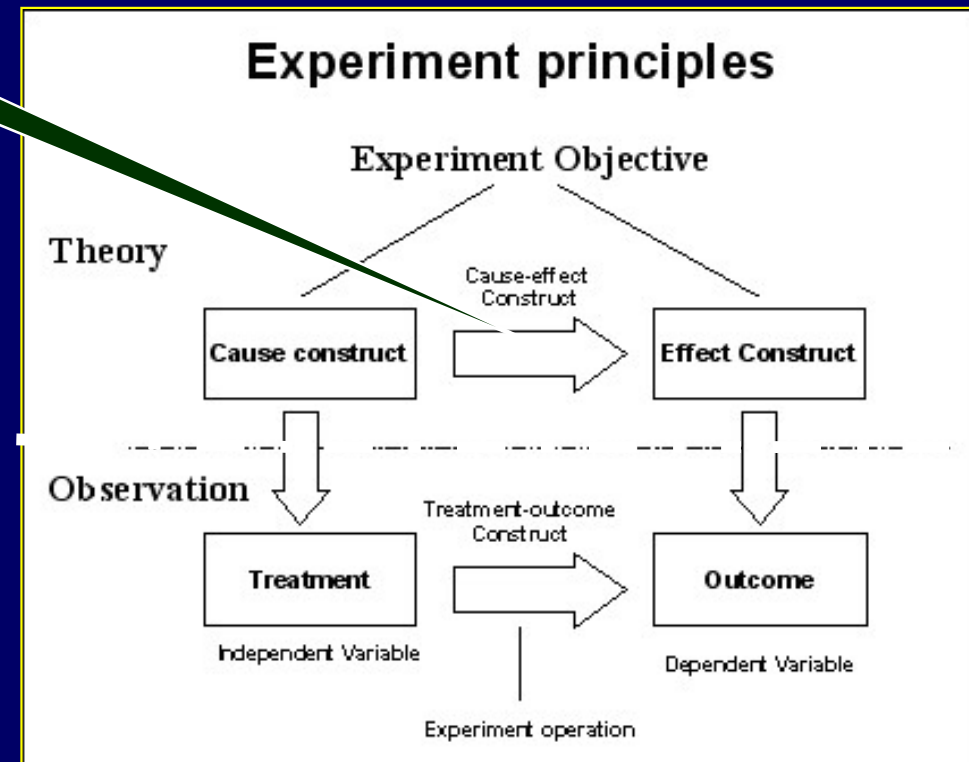
External Validity

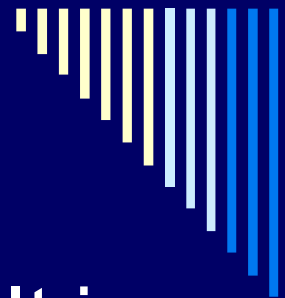
(1/3)

Theory

External
Validity

Q: Is there an adequate
→ general relationship
between treatment and
response?



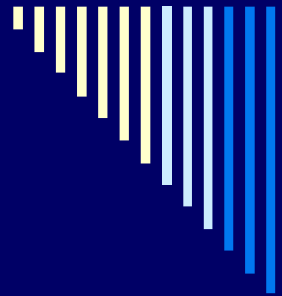


External Validity (2/3)

It is concerned with *generalization* of the results *outside* the scope of the study, hence the experiment setting.

Threats to External Validity

External validity is *affected* by the *experiment* design chosen, but also the *objects* in the experiment and the *subjects* chosen.



External Validity

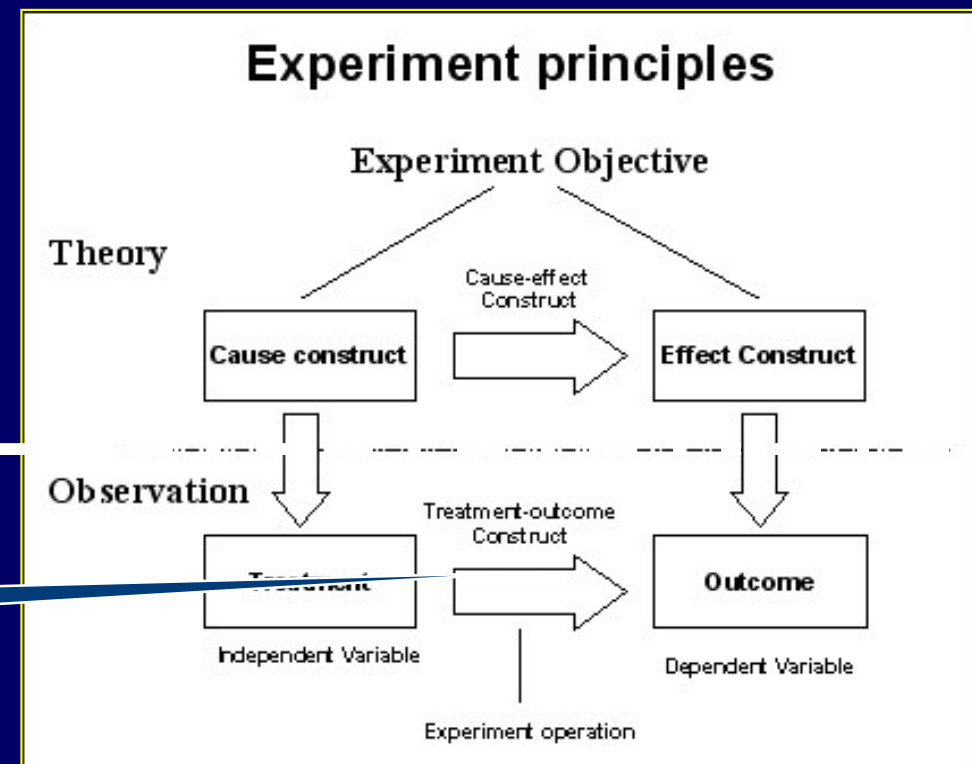
(3/3)

Threats to External Validity

There are three main **risks**:

- Having **wrong participants** as subjects
- Performing the experiment in the **wrong environment**
- Performing it with a **timing** that affects the results.

Internal Validity

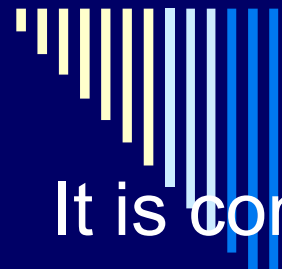


Observation

Internal Validity

Q: Is it a causal relationship?

Intone



Internal Validity

It is concerned with a stated **statistical relationship between treatment and outcome** and wants to make sure that such a relationship is a **causal relationship**, i.e., **the treatment** rather than **other uncontrolled or unmeasured factors** is the cause of the outcome.

Threats to Internal Validity

They are concerned with issues that may indicate a causal relationship, although there is none.

These issues include how **subject** are **selected**, **divided** in different strata and grouped, **treated** and **compensated** during the experiment.

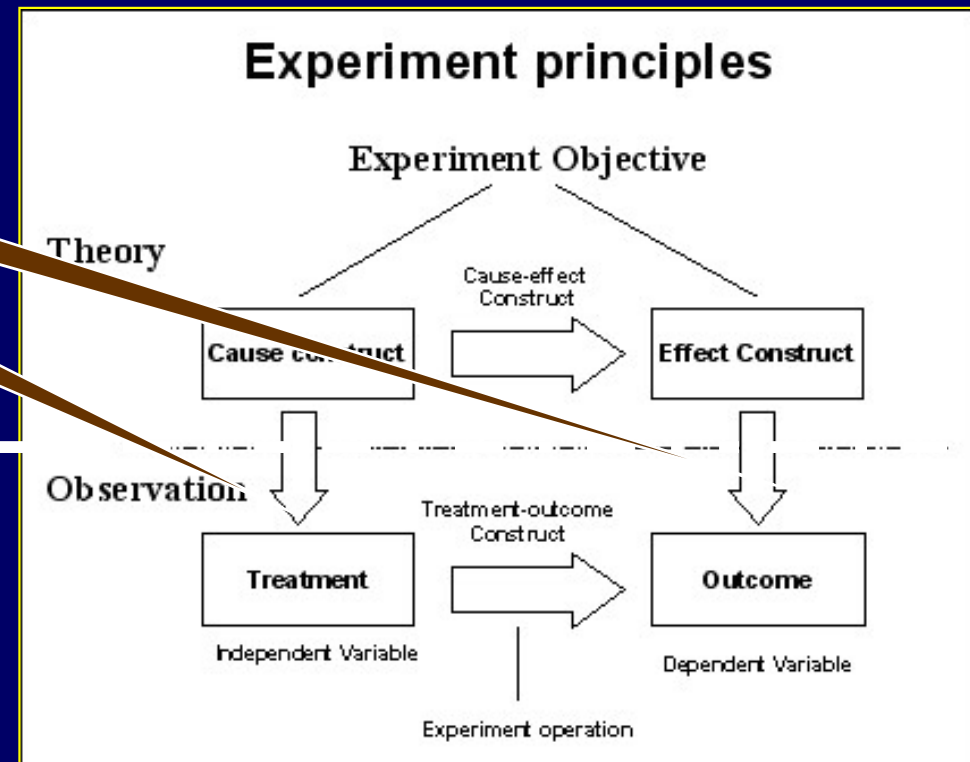
Construct Validity

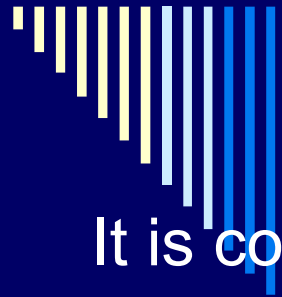
Theory

Observation

Construct
Validity

Q: Are treatments and
outcomes correct
instances of the theory
level's entities





Construct Validity

It is concerned with **relationship between the level of the *theory* and the level of the *observation*.**

If the relationship between cause and effect is a causal relationship, we must ensure two things: 1) that **the treatment reflects the construct of cause well, and** 2) that **the outcome reflects the construct of effect well.**

Threats to Construct Validity

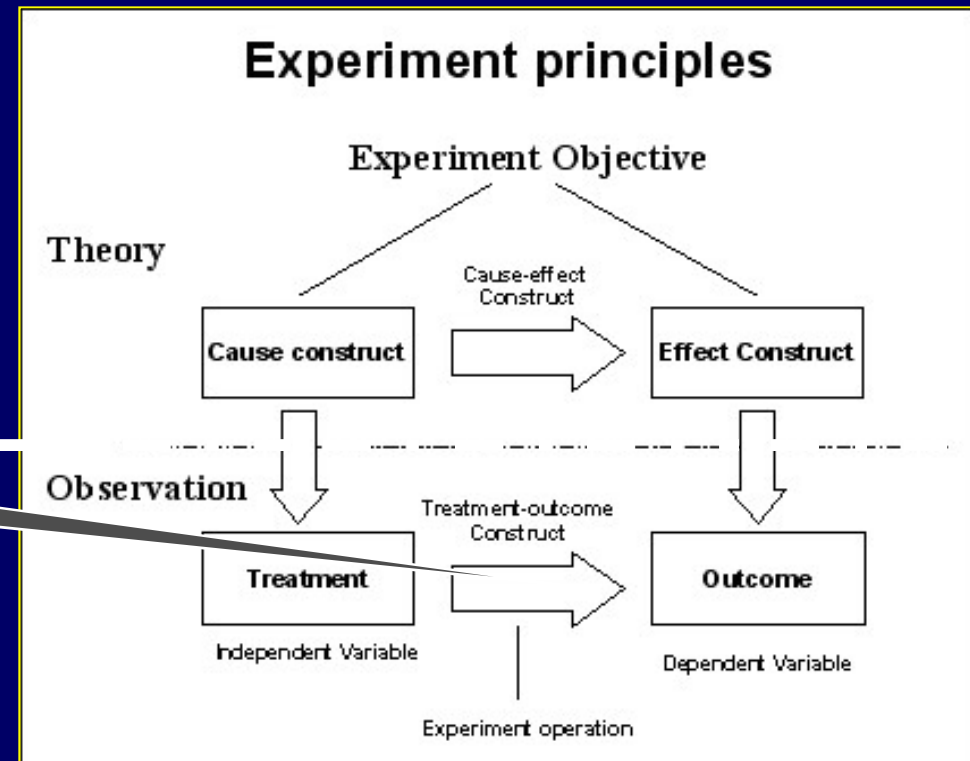
They refer to the extent to which the experiment ***setting*** actually reflects the ***construct*** under study. This involves the adequacy for the experiment of the utilized ***measurement models***.

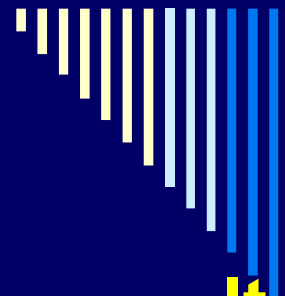
Conclusion Validity

Observation

**Conclusion
Validity**

**Q: Is there a statistical
significant relationship
between treatment and
outcome?**





(“Statistical”) Conclusion Validity

It is concerned with relationship between treatment and outcome.

We want to be sure that there is a statistical relationship with a given significance.

Threats to Conclusion Validity

They are concerned with issues that affect the ability to draw the correct conclusion about that relationship.

These issues include the choice of the **statistical test**, the choice of the **sample size**, care taken in **implementation** and **measurement** of an experiment.



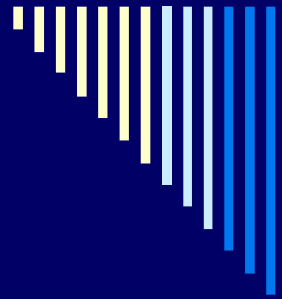
SUBJECTS, OBJECTS & DESIGN

Are we using the right design, objects, subjects, environment, and assigning enough time to subject?

List of Validity Threats (1/4)

External Validity

1. Interaction among setting and treatment
2. Interaction of history and treatment
3. Interaction of selection and treatment



External Validity Threats

1. **Interaction among setting and treatment:** When using old technologies, toy objects, etc.
2. **Interaction of history and treatment:** e.g., conduction questionnaire on software-based safety the day after a software crash.
3. **Interaction of selection and treatment:** Subjects are not representative of the population.

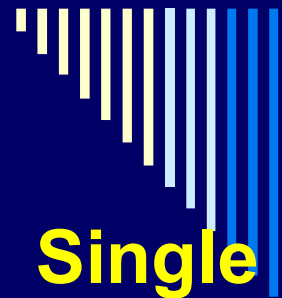
GROUPING & RANDOMIZATION

Are we **well** randomizing on objects and subjects, and **well** grouping, compensating & treating subjects?

List of Validity Threats (2/4)

Internal Validity

| | |
|---------------------------|--|
| 1. History | 8. Ambiguity about direction of causal influence |
| 2. Maturation | 9. Interaction with selection |
| 3. Testing | 10. Diffusion of imitation of treatments |
| 4. Instrumentation | 11. Compensatory equalization of treatments |
| 5. Statistical regression | 12. Compensatory rivalry |
| 6. Selection | 13. Resentful demoralization |
| 7. Mortality | |

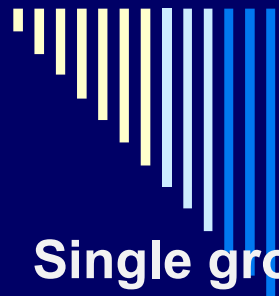


Internal Validity Threats (1/4)

Single group threats (1/2): There is no control group (to which we do not apply treatment).

There is problem in determining if the treatment or some other factor caused the observed results.

- 1. History:** E.g., Run the experiment in a day after holiday.
- 2. Maturation:** Subjects perform differently as time passes or because of dependencies from the experimenter (e.g., of students from their teacher).
- 3. Testing:** Concern repetition of a test. If a test is repeated, subjects may respond differently at different times, since they know how the test is conducted.

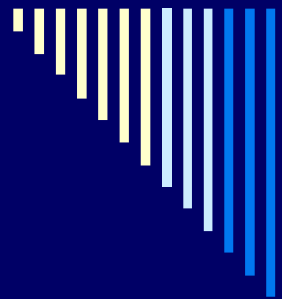


Internal Validity Threats (2/4)

Single group threats (2/2): There is no control group (to which we do not apply treatment).

There is problem in determining if the treatment or some other factor caused the observed results.

- 4. Instrumentation:** Is the effect caused by data collection forms, documents to be inspected, etc. If they are badly designed the experiment is affected negatively.
- 5. Selection:** Volunteers are generally more motivated than the whole population.
- 6. Statistical regression:** Subjects are stratified, based on their performances in previous experiments.

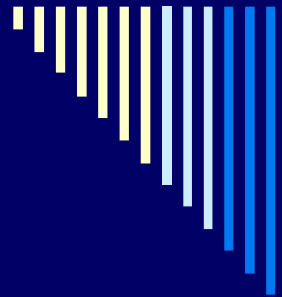


Internal Validity Threats (3/4)

Social threats: Applies, for instance, when new method, tool, etc. is compared with one already in use.

10. Compensatory equalization of the treatments: When compensating subjects in the control group for not getting treatment and compensation affect their behavior.

11. Compensatory rivalry: (1) Who gets the less desirable treatment may tend to reduce or reverse the expected outcomes. (2) People using traditional treatment may do their best.



Internal Validity Threats (4/4)

Social threats: Applies, for instance, when new method, tool, etc. is compared with one already in use.

12. Diffusion or imitation of the treatment: When the control group learn about the treatment and try to imitate the behavior of the group in the study

13. Resentful demoralization: When (1) Subjects who get the less desirable treatment tend to reduce their performances, or (2) Subjects who get the new treatment tend to perform at the best.

DOES THE TREATMENT CAUSE THE OUTCOMES

Are we using the right design for this experiment, the right setting for this construct, the right blocking, the adequate measurement model? (E.g. What is the level of the subjects?)

List of Validity Threats (3/4)

Construct Validity

| | |
|--|--|
| 1. Violated assumptions of statistical test | 6. Interaction of different treatments (through the same subjects) |
| 2. Inadequate preoperational explication of constructs | 7. Restricted generalization among constructs |
| 3. Mono-operational bias | 8. Hypothesis guessing |
| 4. Mono-method bias | 9. Experiment expectancies |
| 5. Confounding constructs and levels of constructs | 10. Evaluation apprehension |



Construct Validity Threats (1/4)

Design threats: concerns the design and its ability to reflect the construct to be studied.

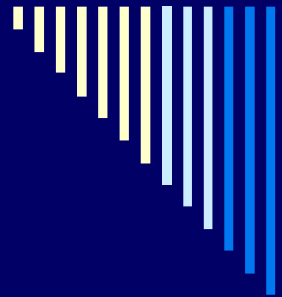
1. **Violated assumptions of statistical tests:** When trying to jump over some critical statistical rules (i.e. normality, independent sample). In general, when utilizing a test with samples or outcomes, which do not meet the test assumptions.
2. **Inadequate preoperational explication of constructs:** Construct are not sufficiently defined before they are translated into measures or treatments.
E.g., if you measure productivity of a method then you should also measure maintainability of products.



Construct Validity Threats (2/4)

Design threats: concerns the design and its ability to reflect the construct to be studied.

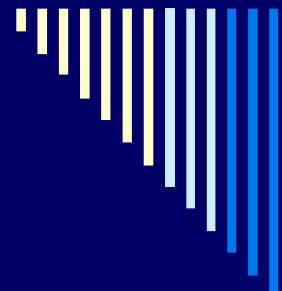
- 3. Mono-method bias:** When using a single type of measure, specially measure influenced by subjectivity, the experiment may bias the measure.
- 4. Mono-operational bias:** When the elementary experiments (subject, object, treatment) include a single independent variable, the experiment may not give a full picture of the theory.
- 5. Confounding constructs and levels of constructs:** Using presence/absence, rather than levels, of construct.



Construct Validity Threats (3/4)

Design threats: concerns the design and its ability to reflect the construct to be studied.

- 6. Interaction of different treatments (through the same subjects):** When subjects are involved in more than one experiment.
- 7. Restricted generalization among constructs:** The treatment affect the intended construct positively and unintended constructs negatively.



Construct Validity Threats (4/4)

Social threats: In the experiment, subjects might act differently than they do otherwise.

8. **Hypothesis guessing:** When subjects bias their behavior on their guesses about the hypotheses.
9. **Experiment expectancies:** When experimenters bias the results consciously or unconsciously based on what they expect from the experiment.
10. **Evaluation apprehension:** When people are afraid to be evaluated and try to perform better when being evaluated.

STATISTICAL RELATIONSHIP

Are we using the **right** sample size and statistical test, are we measuring well?

List of Validity Threats (3/4)

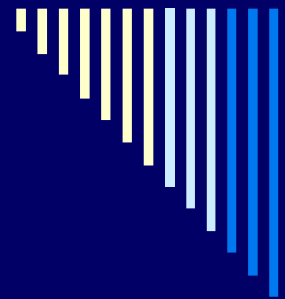
Conclusion Validity

1. Low statistical power
2. Violated assumptions of statistical tests
3. Reliability of measures
4. Fishing and the error rate
5. Reliability of treatment implementation
6. Random irrelevancies in experimental setting
7. Random heterogeneity of subjects



Conclusion Validity Threats (1/2)

1. **Statistical power**: Ability of the test to reveal a true pattern.
2. **Violated assumptions of statistical tests**: When trying to jump over some critical statistical rules (i.e. normality, independent sample). In general, when utilizing a test with samples or outcomes, which do not meet the test assumptions.
3. **Reliability of measures**: Depends on many factors, like using bad instrumentation or bad instrumentation layout, trying to use subjective rather than objective measures. We should use measures that are independent from human judgment, so that when trying to replicate the experiment once more, the same results should be obtained.



Conclusion Validity Threats (2/2)

4. **Fishing**: Searching for a specific result.
5. **Reliability of treatment implementation**: Takes in count (dis)-similarity in the application of treatment to different subject or in different situations.
6. **Random irrelevancies in experimental setting**: Concerns noises outside the experiment room that affect results.
7. **Random heterogeneity of subjects**: Differences among subjects is so large that variations due to subjects is higher than variation due to treatment.



Conflicts between Types of Validity Threats

When increasing one type of validity, another type may decrease.

Using students →

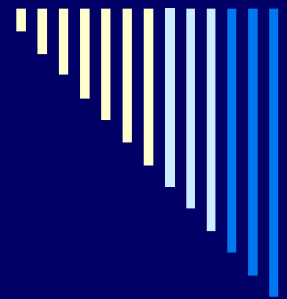
(Due to large study groups, reduced heterogeneity, reliable treatment implementation)

- High conclusion validity
- Reduced external validity

Subjects get measurements by themselves →

(Tedious measurements tend to reduce the reliability of measures)

- Reduced construct validity
- Reduced conclusion validity



Priority among Types of Validity Threats

Depending on the purpose of the experiment, different types of validity can be differently prioritized.