

Empirical Quantitative Methods in Computer Science

Lecture 1: Measurements

Chat Wacharamanotham
Spring Semester 2026

<https://chatw.ch/quant>



Chatchavan Wacharamanotham

Lecturer at University of Zurich

PhD in Human-Computer interaction from RWTH Aachen University, Germany

Previously:

Assistant professor at University of Zurich

Lecturer at Swansea University, UK

Research: Improving how computer can help people do better and transparent science <https://chatw.ch>



<https://chatw.ch/quant>



Topics

- Comparisons
- Null-hypothesis significance tests (t-test, ANOVA)
- Linear models and friends
- Statistical assumptions
- Data transformation
- Data imputation
- Simulating data
- Experimental design
- Statistical graphics

Goal

- Being able to choose statistical procedures and models for empirical research in a responsible manner
- Not this course's goal: Inventing new statistical procedures

<https://chatw.ch/quant>



Course logistics

See the syllabus

<https://chatw.ch/quant>



Empirical Quantitative Methods in Computer Science — 2026

Instructor: Chat Wacharamanatham (to email, see OLAT) Office hours: <https://chatw.ch/h>

Course format: Lecture with in-class exercises

COURSE DESCRIPTION

This course introduces students to the following topics in empirical quantitative research:

- Controlled experiments
- Survey research
- Data visualization for statistical analysis
- Exploratory and inferential statistical analysis
- Research transparency

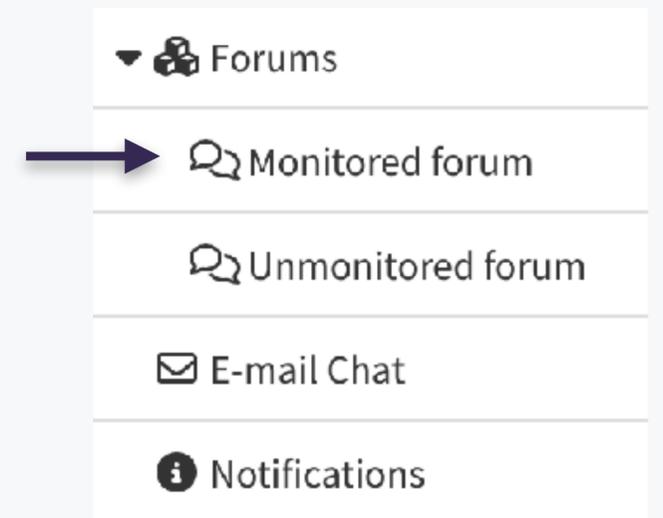
Students will learn these methods hands-on through assignments and project work.

INTENDED LEARNING OUTCOMES

1. Students can explain the properties of measurements and can choose appropriate measurements for their research.
2. Students understand types of validity in quantitative research.
3. Students can design and conduct controlled experiments.
4. Students can design and conduct survey studies.
5. Students can visualize data from their experiments for exploratory analysis.
6. Students can explain types of missing data and can choose appropriate techniques to address them.
7. Students know how to perform inferential statistics by choosing correct procedures and checking statistical assumptions.
8. Students can organize their research materials and prepare them for sharing in a transparent manner.

Get help

- For [questions that are not sensitive](#) → post on OLAT forum “Monitored forum”. Chat respond on Wednesdays and Fridays
- “Unmonitored forum” are for students only to form teams
- [Chat’s office hours](#) are on Wednesdays 14:00–15:30. With exceptions: chatw.ch/h No advance appointments needed. Just come by at BIN 2.B.07.
- Last resort: Email via OLAT’s “Email Chat” section
- If all other channels fail, email chat@ifi.uzh.ch



Research contribution types

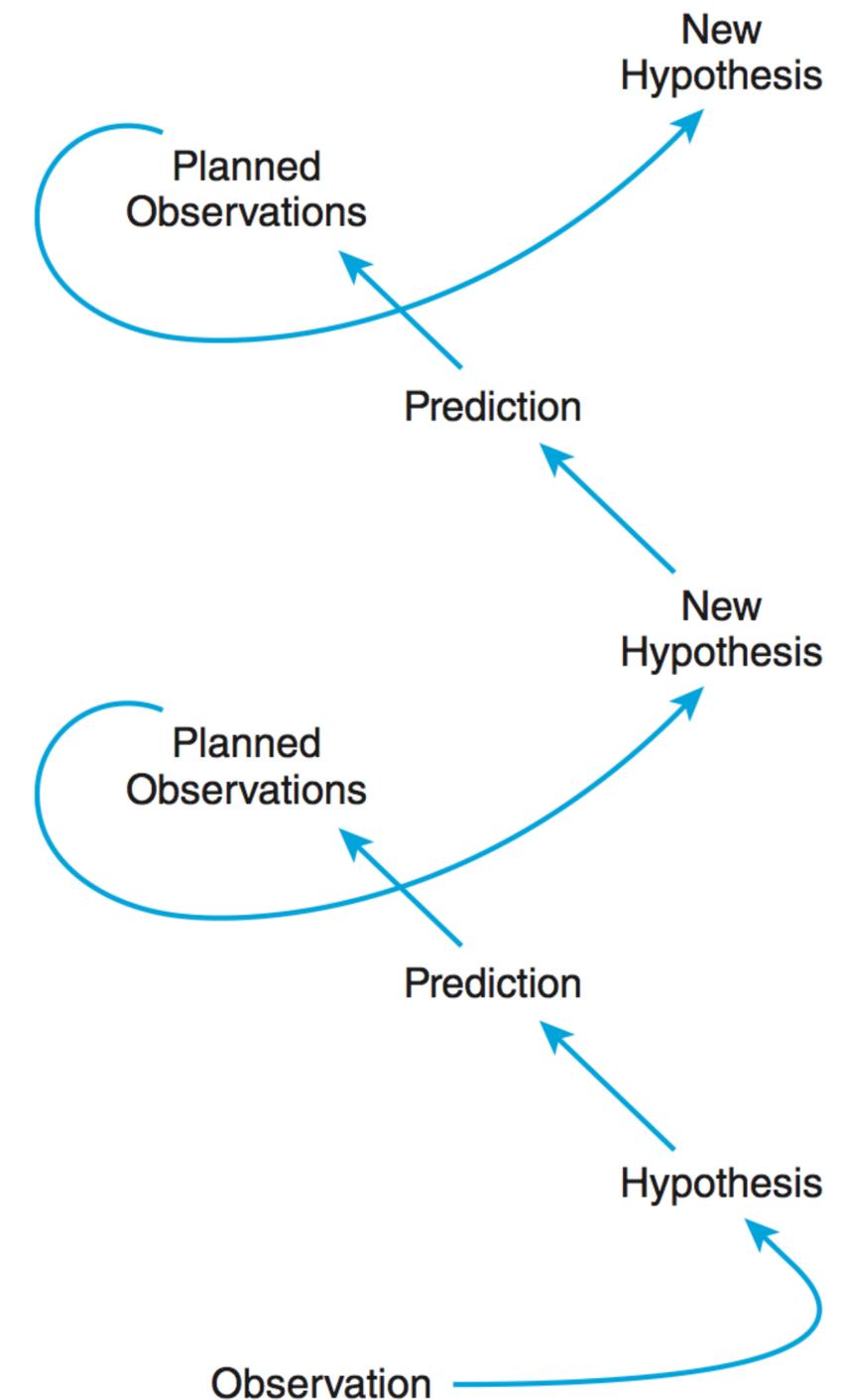
- **Empirical:** Knowledge about users or system usage
- **Artifact:** Design and engineering of interactive systems
- **Methodological:** Knowledge about new way of research
- **Dataset** that can enable future research
- **Literature-survey** that structure the current state of knowledge
- **Theoretical** principles or models
- **Opinion** pieces that aim to change readers' mind

A research paper may make multiple contributions

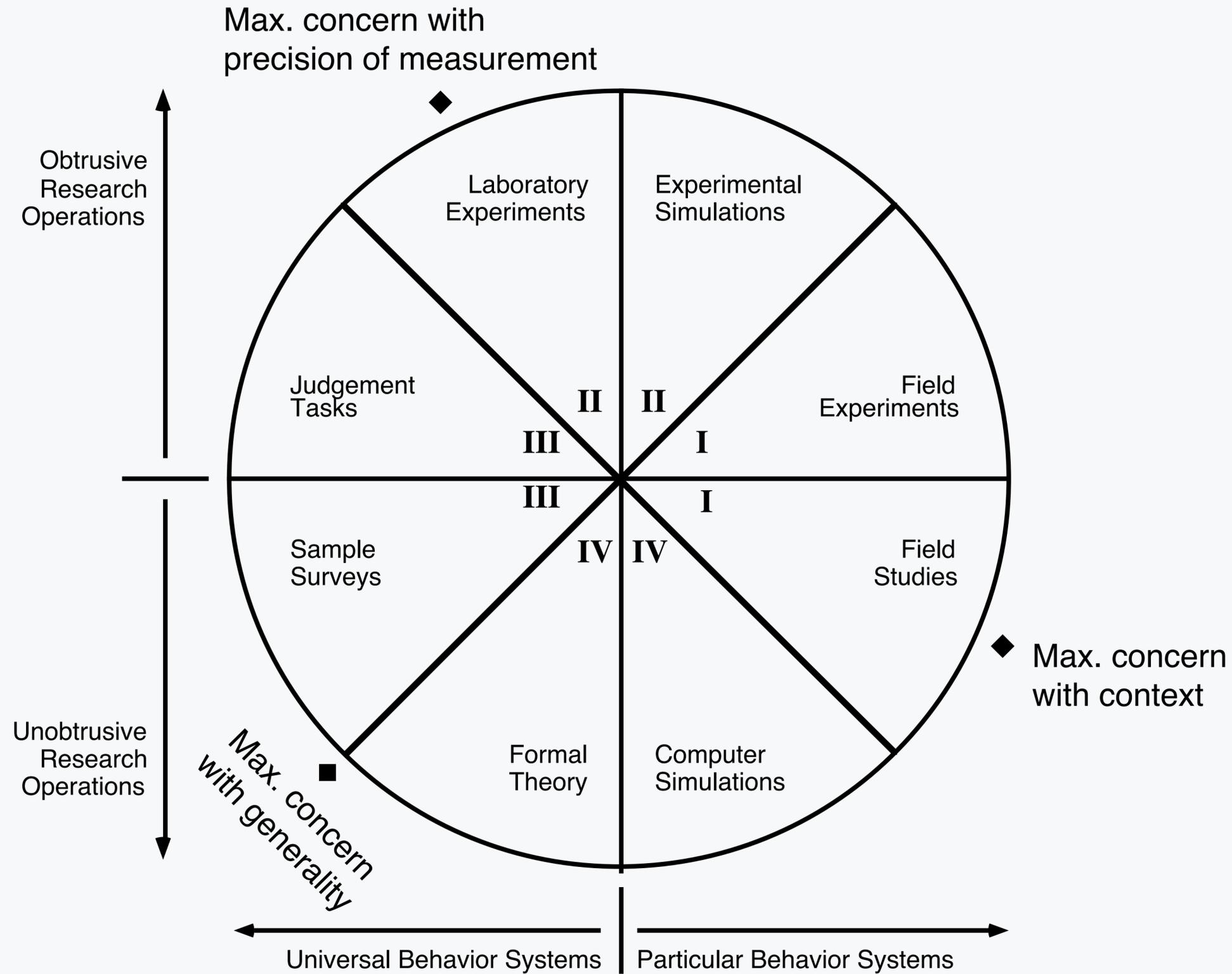
Empirical research

Empirical method uses observation or direct sensory experience to obtain knowledge

Scientific method is a method of acquiring knowledge that uses observations to develop a hypothesis, and then uses the hypothesis to make logical predictions that can be empirically tested by making additional, systematic observations. Typically, the new observations lead to a new hypothesis, and the cycle continues.



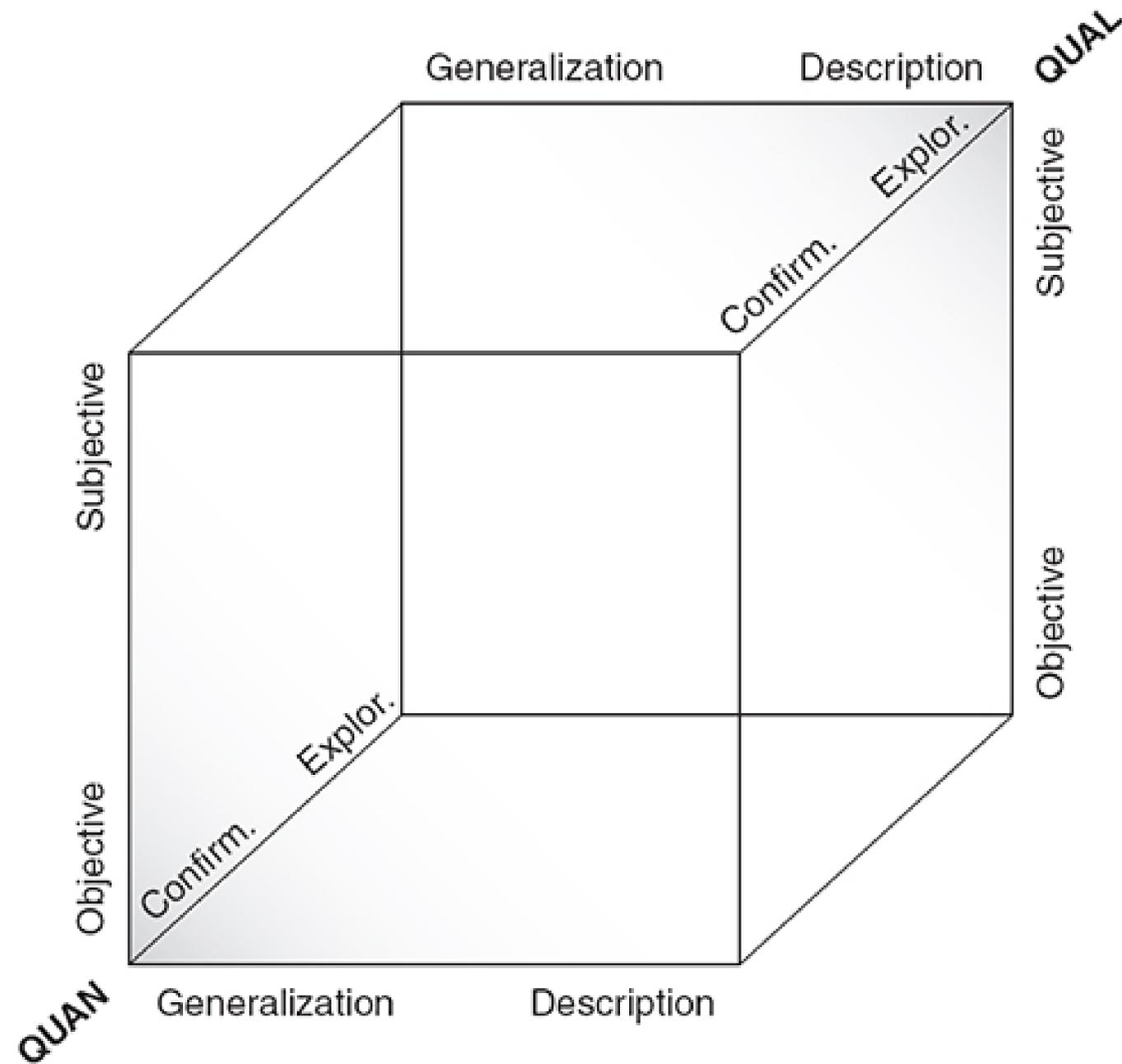
Strategies for empirical research



McGrath, J. E. (1981). Dilemmas: The study of research choices and dilemmas. *American Behavioral Scientist*, 25(2), 179-210.
 McGrath, J. E. (1995). Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human-Computer Interaction* (pp. 152-169). Morgan Kaufmann.

not in exam

CONTINUUM OF RESEARCH PARADIGMS (SELECTED 3 DIMENSIONS)



Objectivity: The extent to which the researchers are detached from the thing or people being studied

Research studies that are concerned with human experiences may inevitably be **subjective**. The subjectivity could be in the study participants, the researchers, or both

Generalizability: The extent to which we can use the results of a research study (which is based on a relatively small set of specific observations) to form a general statement about a larger set of possible observations (different people, settings, times, measures, and characteristics other than those used in that study).

Some research projects might not aim for generalizability. Instead, they aim to provide detailed and rich **descriptions** of specific phenomena

Confirmatory: The extent to which researchers are certain about specific outcomes prior to conducting the study

In many situations, humanity may not have enough knowledge about the topic of research. Thus, researchers may approach the project without predefined answers or specific questions. Such projects are called **exploratory**.

Empirical Quantitative Methods in Computer Science

Lecture 1: Measurements

Chat Wacharamanotham
Spring Semester 2026

<https://chatw.ch/quant>



Intended learning outcomes

By the end of this course, students will be able to

- ❑ Explain the relationship between constructs and operationalization
- ❑ Identify scales of measurement and their limitations

<https://chatw.ch/quant>



Constructs and operationalization

Defining “success”

STEP 1

🕒 3 minutes 👤 work individually, brainstorm ways to complete the following sentences

A computing system is successful when...

Write your ideas in the following shared doc

🕒 3 minutes 👤 work individually



chatw.ch/quant

and select lecture 1 shared doc

Defining “success”

STEP 1

In pairs, brainstorm ways to complete the following sentences

A computing system is successful when...

Write your ideas in the following shared doc

STEP 2

For each cluster of “success”, brainstorm ways to measure them.

If the ideas in the cluster is already a measurement, move them down accordingly.

🕒 5 minutes 👤 work in pairs



chatw.ch/quant

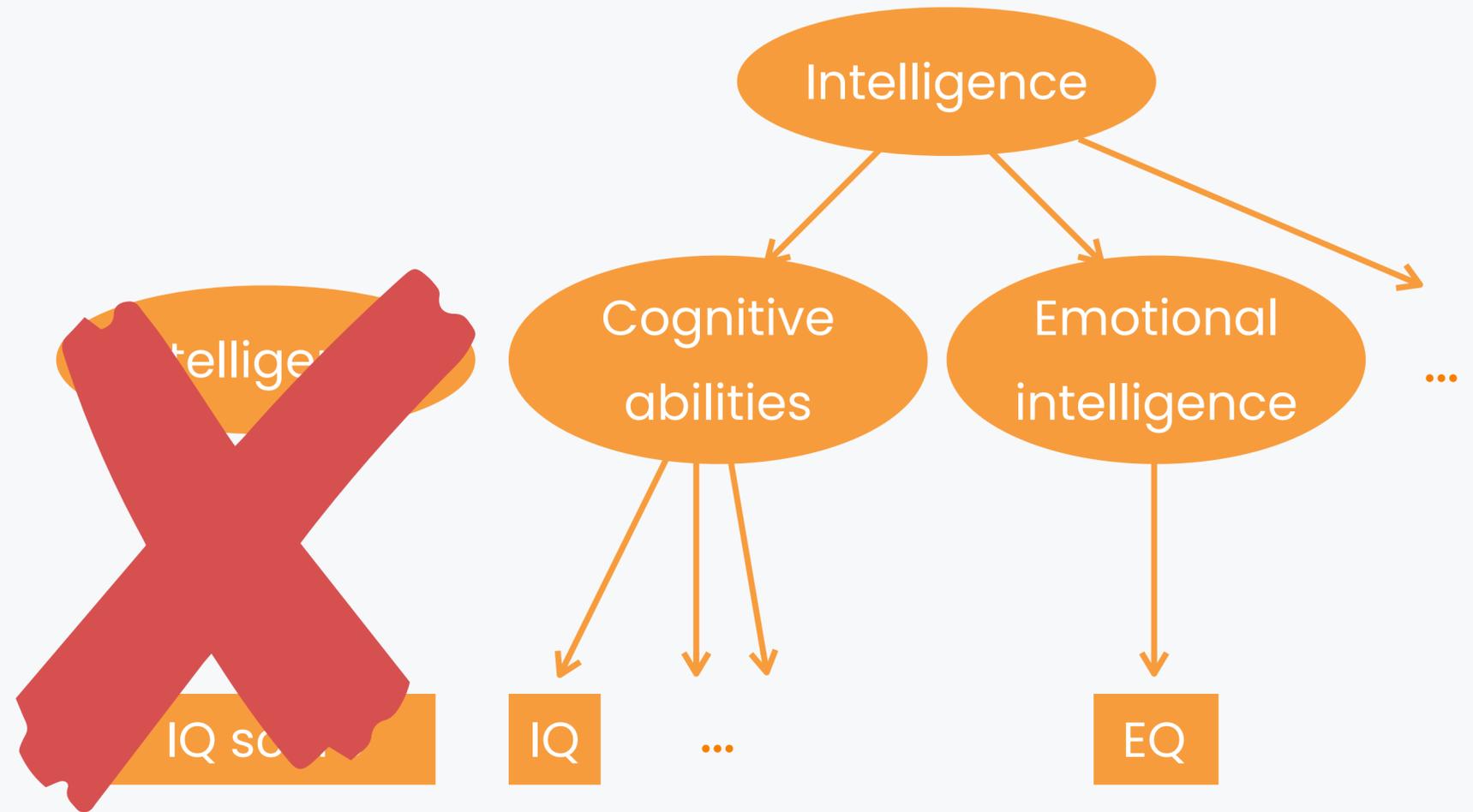
and select lecture 1 shared doc

Constructs vs. Operational definition

Construct: Theoretical and latent concepts that (we expected to) help explain and predict a phenomenon

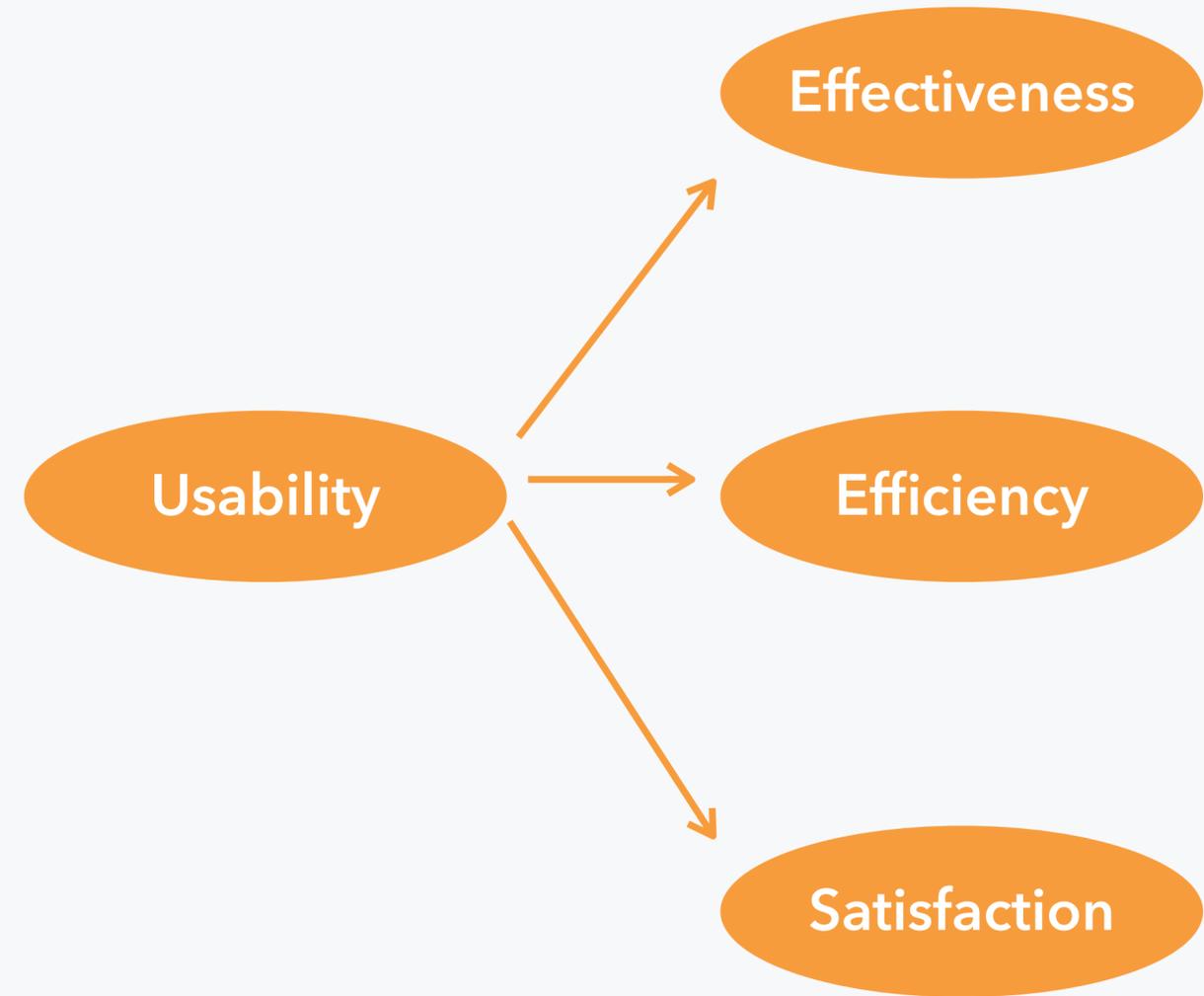
Operational definition: A specification of a procedure to manipulate or measure an external, observable the phenomenon

Construct validity: How much the operationalized measurement correspond to the construct of interest

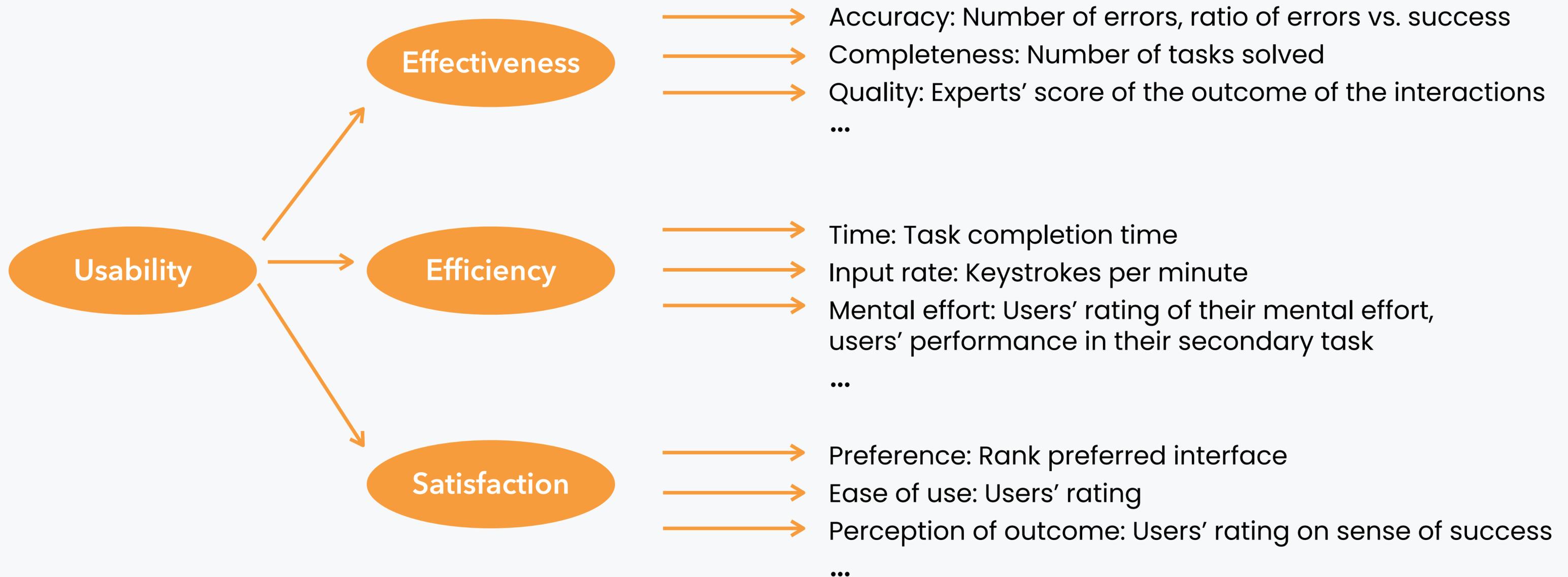


A construct may have multiple facets

- **Effectiveness:** accuracy and completeness with which users achieve specified goals.
- **Efficiency:** resources expended in relation to the accuracy and completeness with which users achieve goals
- **Satisfaction:** freedom from discomfort, and positive attitudes towards the use of the product



A construct may be operationalized in multiple ways



Multiple measurements for one construct

Effectiveness

- Accuracy: Number of errors, ratio of errors vs. success
- Completeness: Number of tasks solved
- Quality: Experts' score of the outcome of the interactions
- ...

When multiple measurements of the same construct produce strongly correlated results → high **convergent validity** → Increase confidence in the results

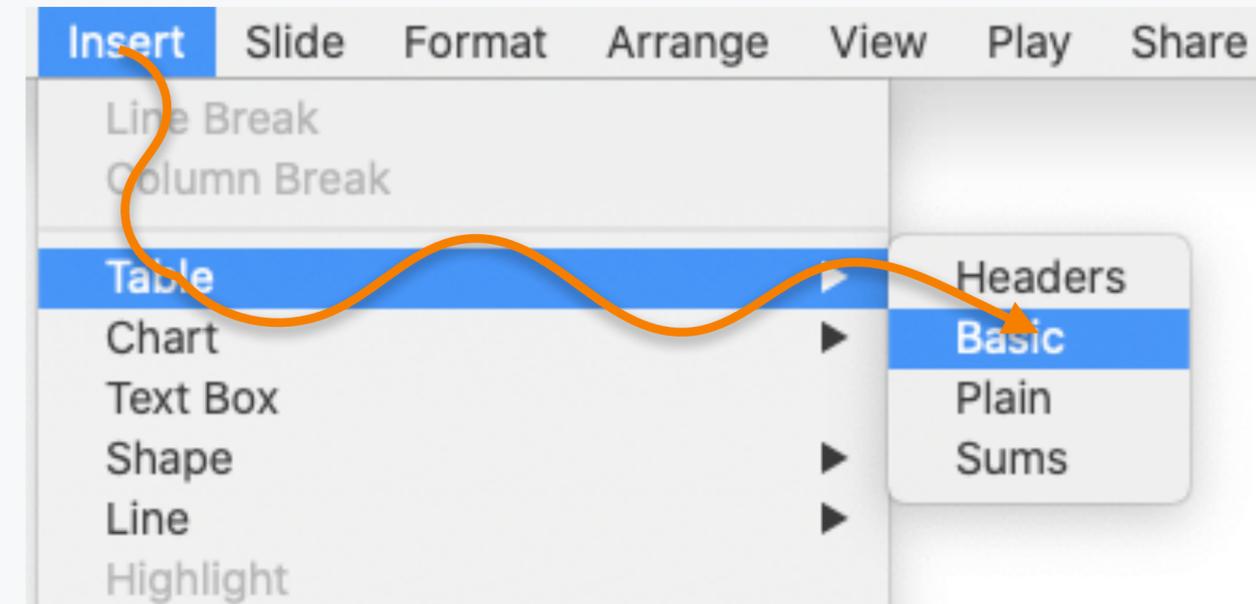
Operationalization in the task context

When operationalizing a construct, the context of the task matters

Example: mouse movement accuracy

- Hit vs. missed the target
- Smoothness of the movement path

Using the latter operationalization is more likely to lead to a better user experience



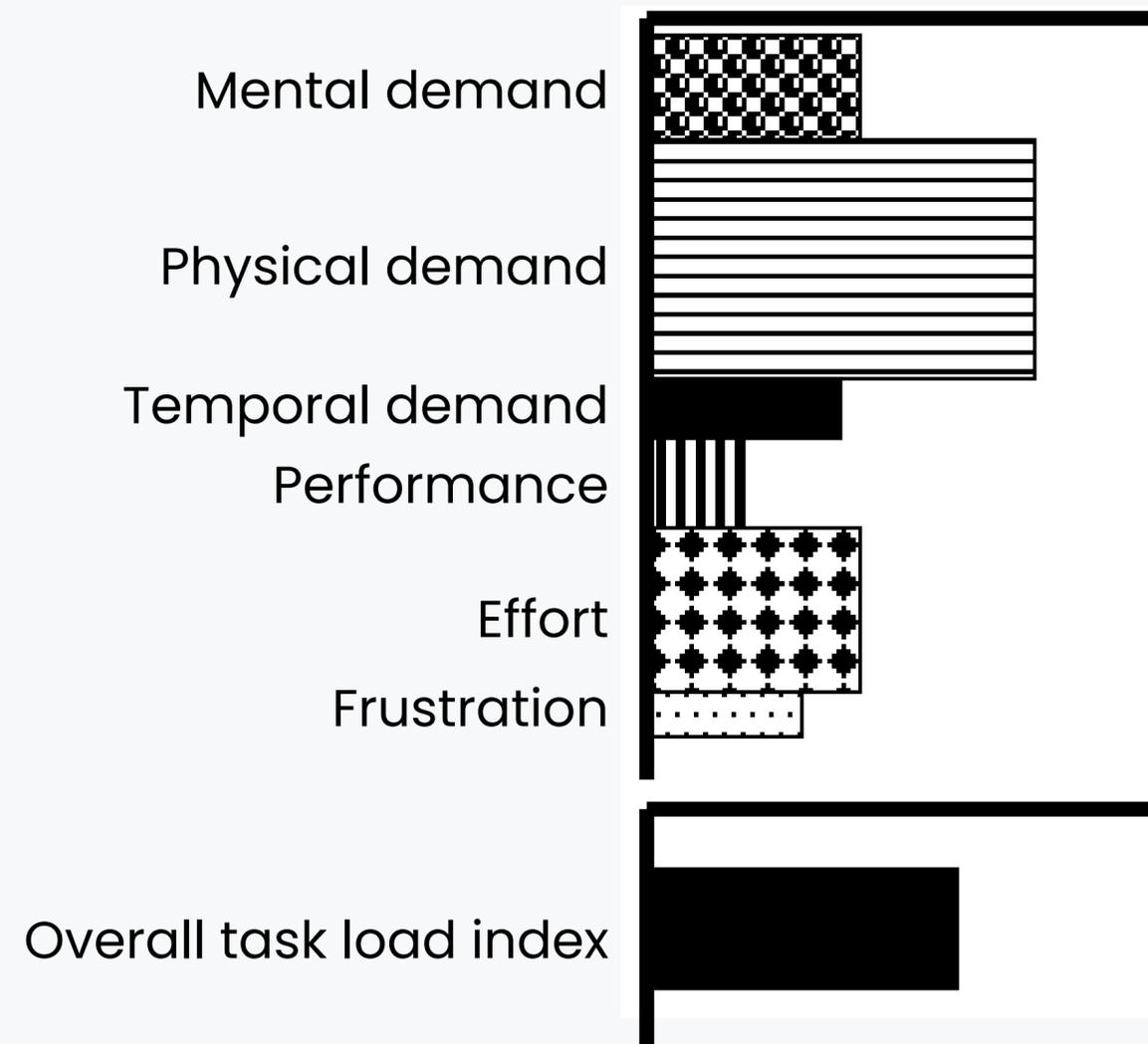
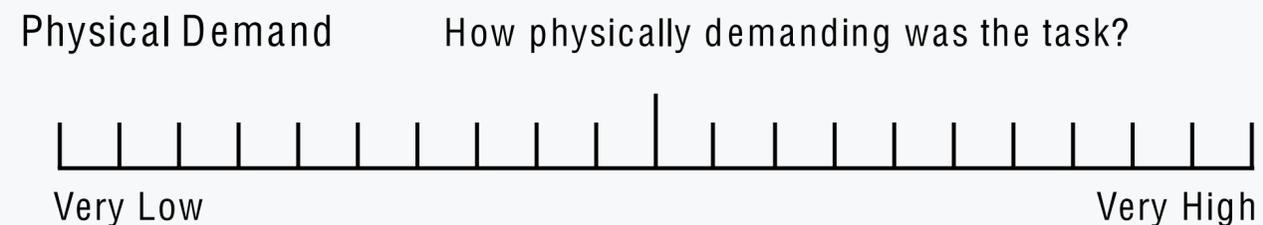
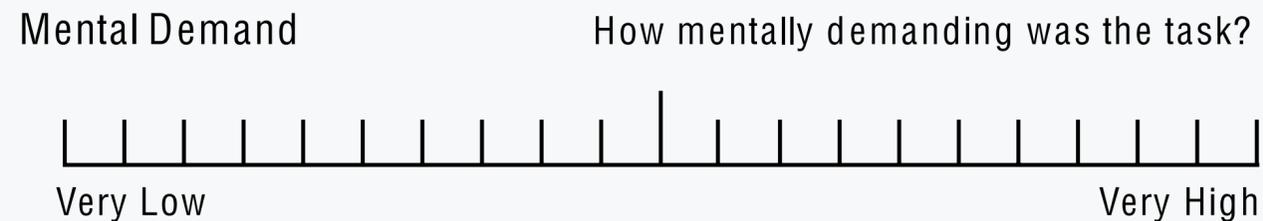
Constructs can be measured or manipulated

Example: measuring mental workload

NASA Task Load Index (NASA-TLX)

- Rating on a 100-point range
- Pairwise comparison of subscales' importance

Example questions:



Constructs can be measured or manipulated

Example: manipulating mental workload

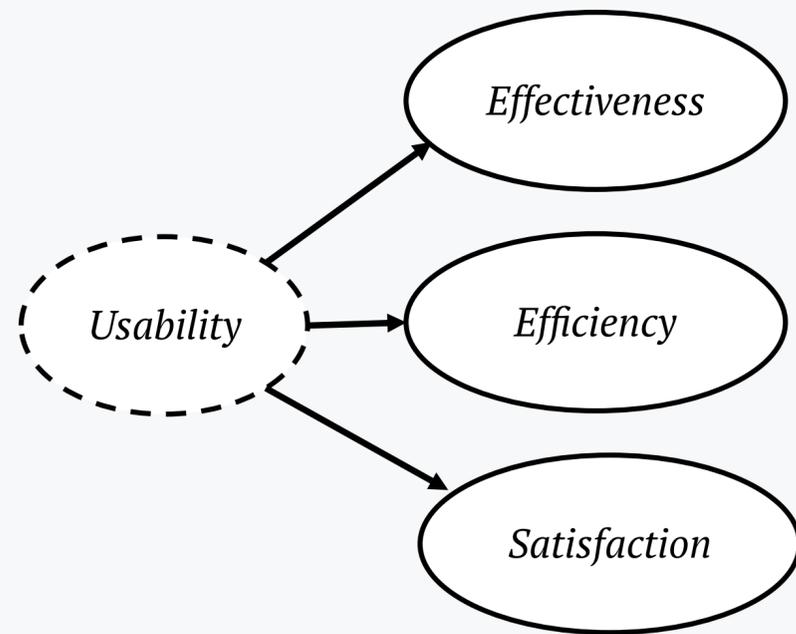
- Purpose: Determine how much eye tracking could be used to estimate mental workload
- Manipulation: Asking the driver to remember single digit number and repeat back verbally
 - Immediately
 - After another number was presented
 - After another two numbers were presented



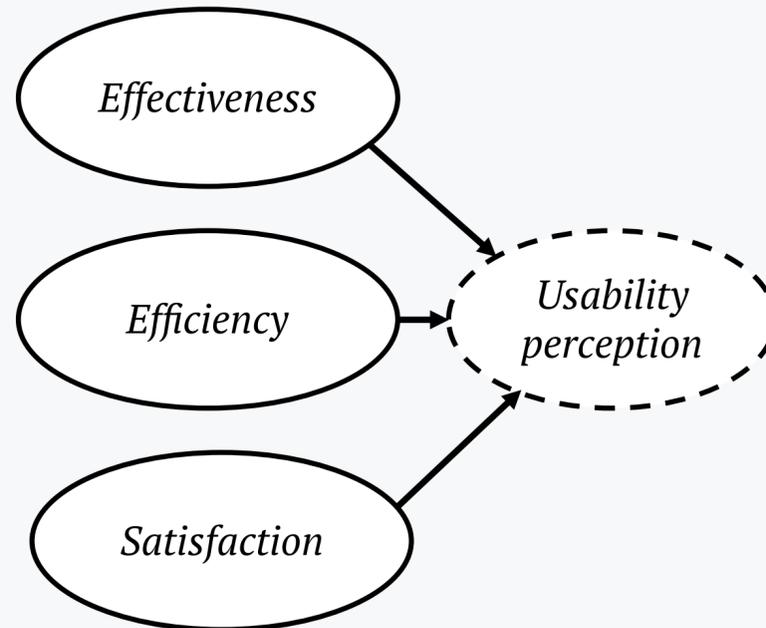
Causal model for explaining relationships among variables

Different theoretical assumptions

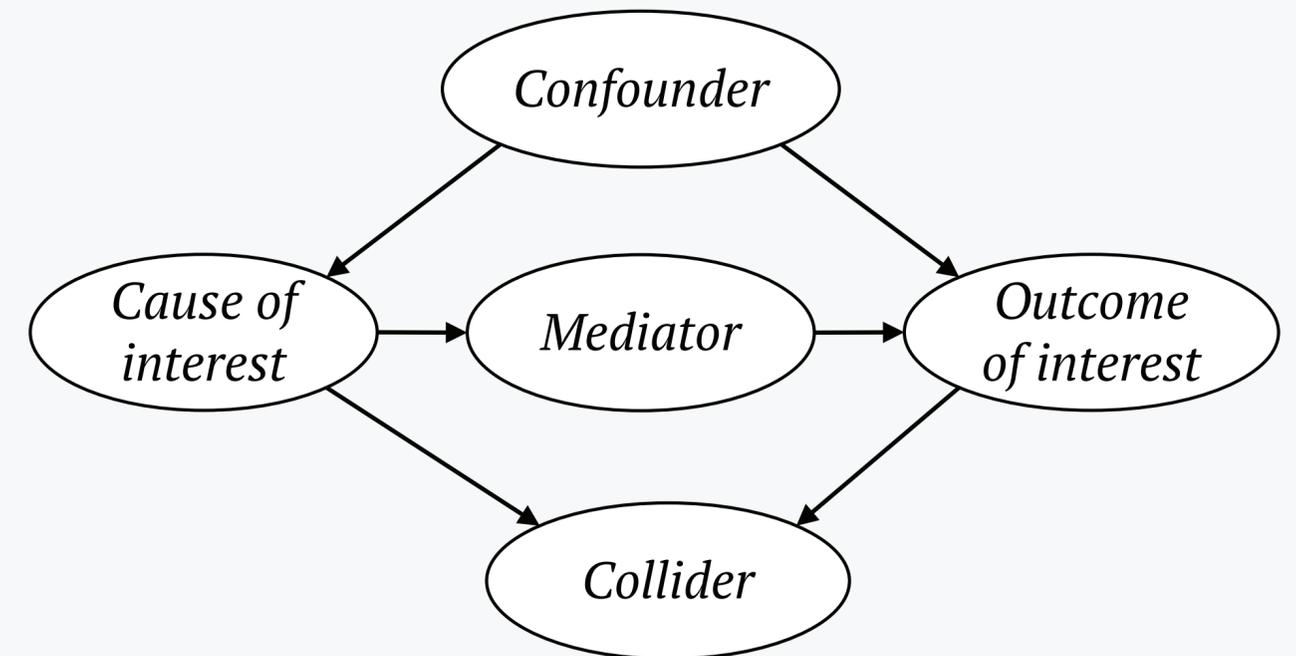
E,E,S estimates the latent usability construct



E,E,S predicts the latent usability construct

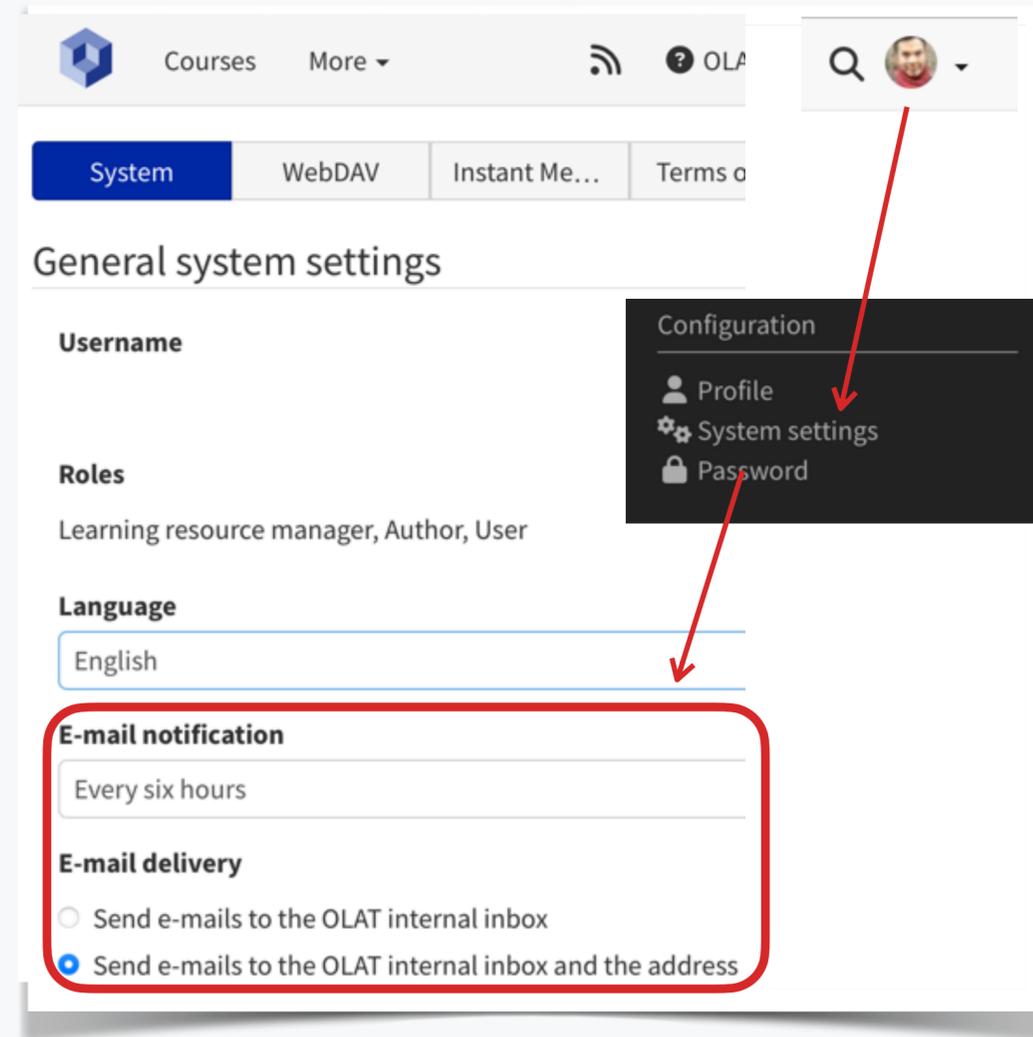
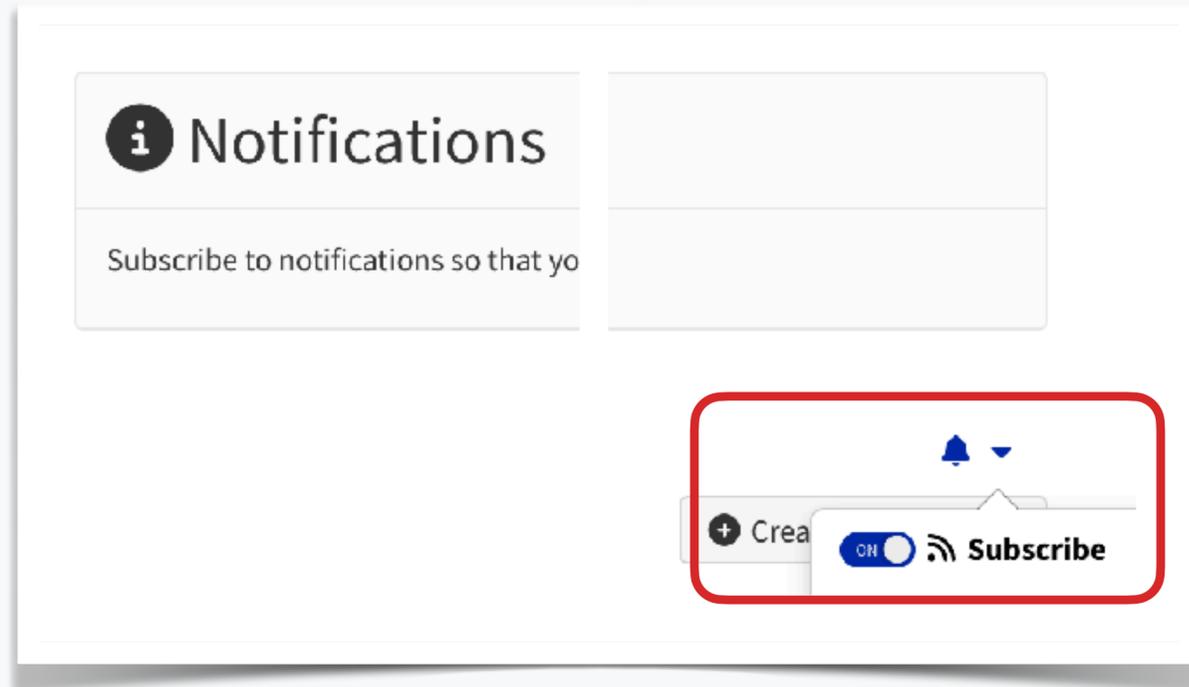


Section 3 in the paper includes good explanation of these term and concrete examples from HCI



Break, return by 15:07

- Try if you can access OLAT of this course. The link is in the URL below
- Enable “Notifications” tab. Ensure that you receive the notification (e.g., by an email or notification on the UZH app)
- A test notification will be sent on 15:40



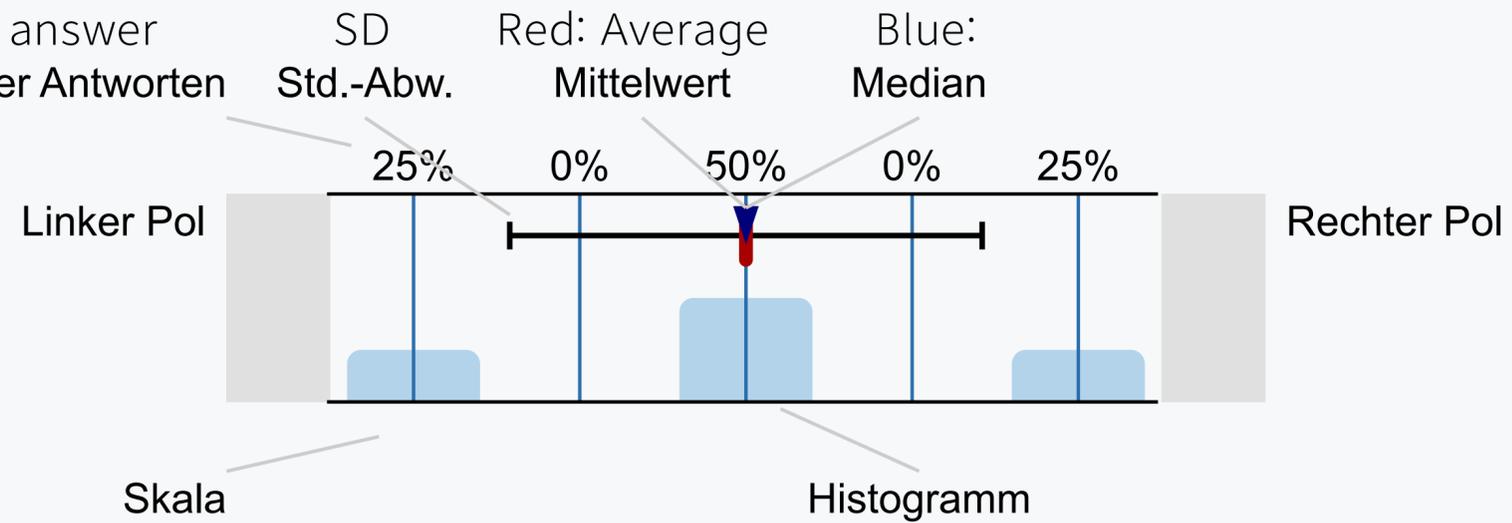
<https://chatw.ch/quant>



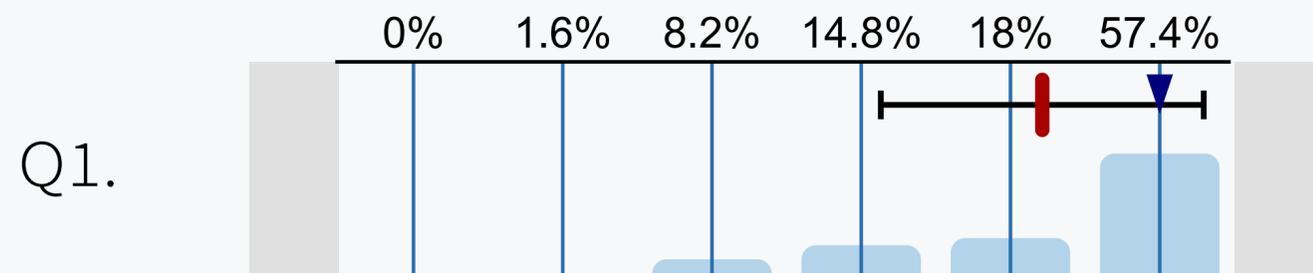
Scale of measurements

Example from the course evaluation questionnaire

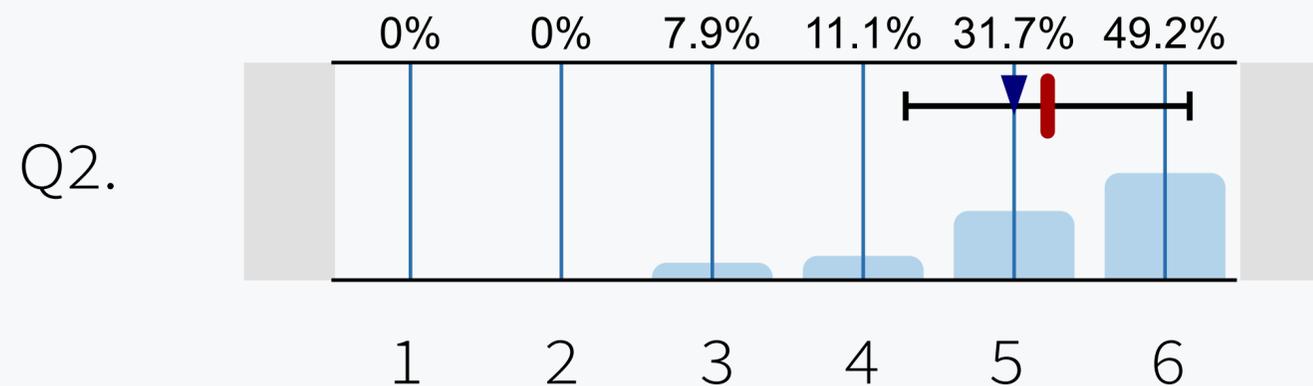
Relative frequency of each answer
Relative Häufigkeiten der Antworten



n=Anzahl
mw=Mittelwert
md=Median
s=Std.-Abw.
E.=Enthaltung

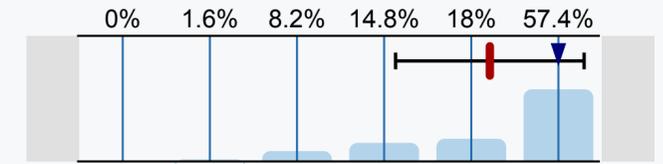


n=61
mw=5.2 (average)
md=6 (median)
s=1.1
E.=1 (abstention)



n=63
mw=5.2
md=5
s=0.9
E.=1

Exercise



Identify type and ordering direction of the following measurements.

	No meaningful order	Meaningful order	Meaningful order + Consistent interval	Ordering Direction
				→ Sequential → Diverging → Cyclic
Date				
Day of week				
Clock time				
Duration				
Weather condition (sunny, cloudy...)				
Temperature (°C)				
5-point response (Strongly disagree to strongly agree)		●		
City name				
Phone number				

Scales of measurement

	Properties	Examples of operations	Example
Categorical = Nominal	Can determine equality or difference	<ul style="list-style-type: none"> • Frequency • Mode 	Task success vs. error
Ordinal	+ Greater or less	<ul style="list-style-type: none"> + Median + Percentiles 	Rank preference of the UIs
Interval = Quantitative	+ One interval is the same size regardless of its location	<ul style="list-style-type: none"> + Calculate differences + Mean + Standard deviation 	Task completion time relative to average
Ratio	+ Has a meaningful zero point	+ Calculate ratio (e.g., "twice", "three times")	Task completion time

The distinction between ordinal vs. interval is more important than interval vs. ratio because many statistical procedures requires the interval property.

One construct, operationalized on multiple scales

	Properties	Example from the "Effectiveness" construct
Categorical = Nominal	Can determine equality or difference	Number of successful tasks in an open-ended usage
Ordinal	+ Greater or less	Rank of conditions based on the success rate
Interval = Quantitative	+ One interval is the same size regardless of its location	Success rate relative to the average
Ratio	+ Has a meaningful zero point	Success rate from a fixed number of tasks

Rating questions

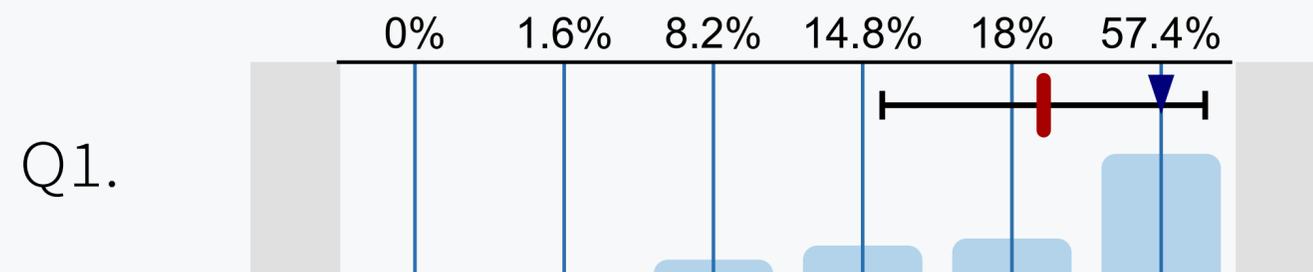
Likert-style question: Individual rating question

Example:

The university cafeterias are clean Strongly disagree 1 2 3 4 5 6 7 Strongly agree

Measurement on the ordinal scale

- ✓ Reporting frequency of each rating
- ✓ Reporting median
- ✗ Reporting average across participants



n=61
mw=5.2 (average)
md=6 (median)
s=1.1
E.=1 (abstention)

Direct interpretation is problematic
Useful for diagnosis

Likert scale

Scores aggregated from multiple rating questions (“item”) in a way that has been empirically show to have interval-scale properties

✓ Can analyze and report scores averaged across participants

	Do not agree at all							Do fully agree						
Simplicity facet of the Visual aesthetic scale														
1.	The layout appears too dense.							1	2	3	4	5	6	7
2.	The layout is easy to grasp.							1	2	3	4	5	6	7
3.	The layout appears well structured.							1	2	3	4	5	6	7
4.	The site appears patchy.							1	2	3	4	5	6	7
5.	Everything goes together on this site.							1	2	3	4	5	6	7

Reverse-score question 1 and 4 and calculate the mean → simplicity score

Summary

A theoretical construct may have several ways to measure

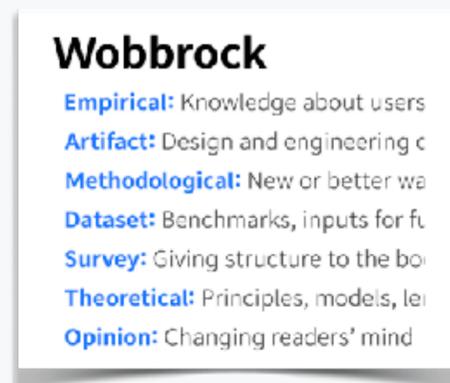
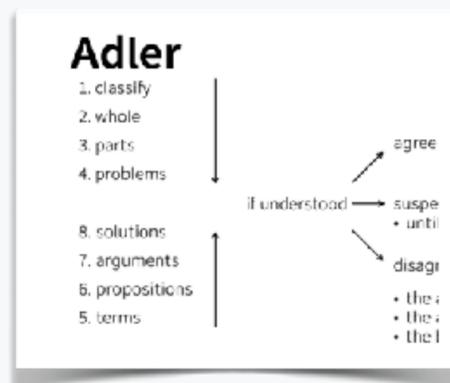
An operational definition specifies one way to measure the construct

Measurements are in different scales, and they allows answering different types of questions

Homework

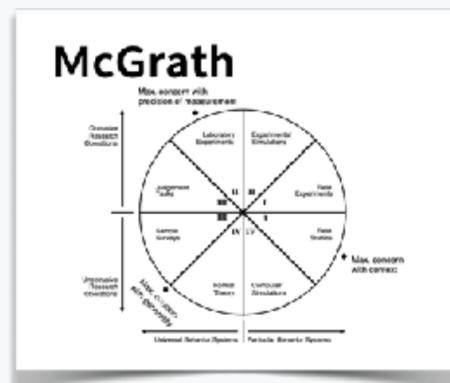
Everyone read: Gravetter & Forzano (2016): Chapter 3 Defining and measuring variables.

- 3.2 Constructs and Operational Definitions
- 3.4 Scale of measurement



According to your reading group: Review the materials and prepare to explain important concepts to your classmates next Tuesday

- **Adler:** Adler (1992) Chapters 11–13 + first two pages of Ch. 14
- **Wobbrock:** Wobbrock & Kientz (2016) first 4 pages
- **McGrath:** Read pages 8–15 in McGrath (1995)
- **Gravetter:** Gravetter & Forzano (2016) section 3.3 Validity and reliability of measurement
- **Hornbæk:** Hornbæk (2006) - Choose 3 operational definitions of a construct and prepare to explain and contrast different ways of measurements



Reading materials are provided on OLAT

Install R and RStudio: Instructions will be on OLAT ▶ Materials ▶ 00 Key Info ▶ R installation instructions

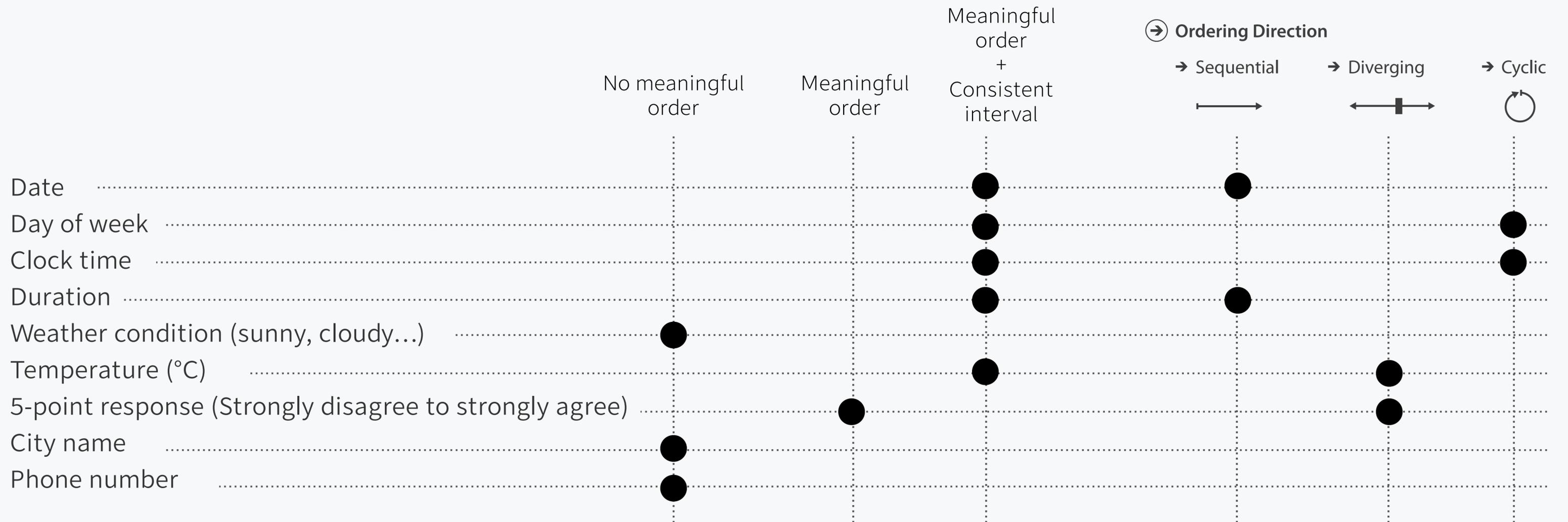
This Thursday: 19 Feb. No lecture
Self-study at home.

If you have not received the test
notification from OLAT, talk to Chat

Supplemental slides

Exercise solution

Identify type and ordering direction of the following measurements.



Student activation

Which semester of the master degree
are you currently in?

Which major are you in?

Which language do you speak most confidently?

Which of the following are you most familiar with?

If you are familiar with multiple concepts, choose one to balance make the groups roughly the same size

Adler

1. classify
2. whole
3. parts
4. problems
5. terms
6. propositions
7. arguments
8. solutions

if understood

- agree
- suspe
- until
- disagr
- the a
- the a
- the b

Wobbrock

- Empirical:** Knowledge about users
- Artifact:** Design and engineering c
- Methodological:** New or better wa
- Dataset:** Benchmarks, inputs for fu
- Survey:** Giving structure to the bo
- Theoretical:** Principles, models, ler
- Opinion:** Changing readers' mind

McGrath

Max. concern with precision of measurement

Obtrusive Research Operations

Unobtrusive Research Operations

Max. concern with generality

Universal Behavior Systems

Particular Behavior Systems

None

Note down your choice