M506

Research Method and Scientific Work:

Quality Checks for Data Collection: Reliability, Validity & Sampling

Week 6, Feb 2022

Prof. Sara Ramzani; Prof. Phungmayo Horam; Prof. Dr. Tilmann Lindberg

Online Teaching and Learning

- Camera On
- Microphone Off please
- If you need to save on bandwidth, switch off camera
- Sessions will be recorded
- If you have something to say – feel welcome – unmute and chip in

(1)
Collecting Secondary Data – How to do that?

Secondary data collection method in research

- Secondary data is a type of data that has already been generated and published in books, newspapers, magazines, journals, online portals etc.

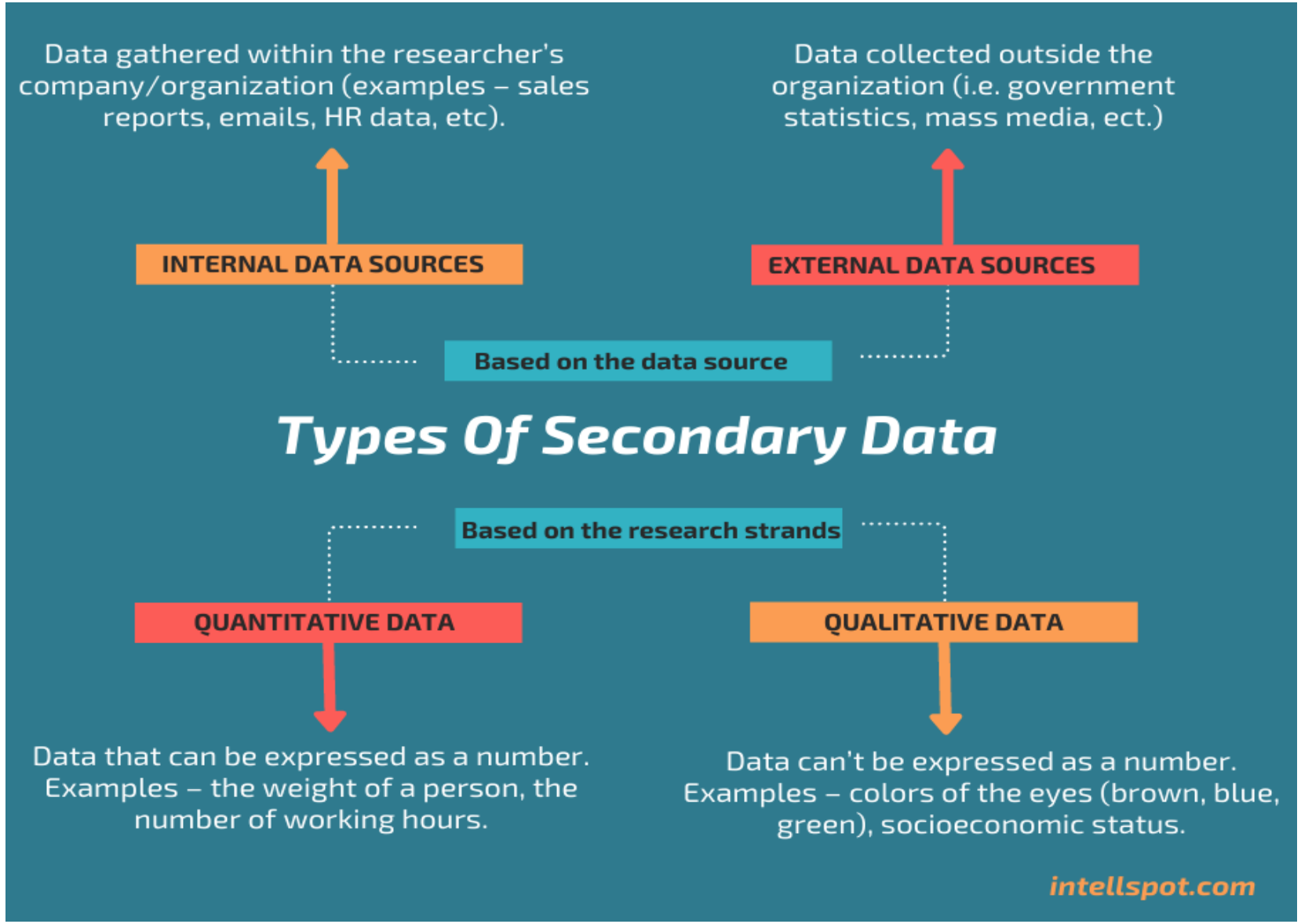
  “In recent years, however, secondary data has become important also in research studies ranging from education and health care to corporate governance, ethics, and social responsibility. From a demand side, this growth has also been driven by an expanded interest in large comparative studies undertaken nationally and internationally. From a supply side, technology has greatly enhanced the capacity of data providers to leverage their databases for commercial purposes, and this has increased the nature and extent of the data they provide.” (Hair, Page and Brunsveld, 2019)

Secondary data collection method in research

Advantages	Disadvantages
• Resource efficiency	• Misalignment of purpose
• Capacity for evaluation	• Access complications
• Potential for comparative analysis	– Cost
– Longitudinal	– Familiarity
– Cross-sectional	– Impact of reporting methods
– Contextual	• Quality concerns
• Avoids respondent fatigue	– Source
• Potential for triangulation	– Data collection methods
• Potential for new insights	– Definitions
	• Age of data

(Hair, Page and Brunsveld, 2019)

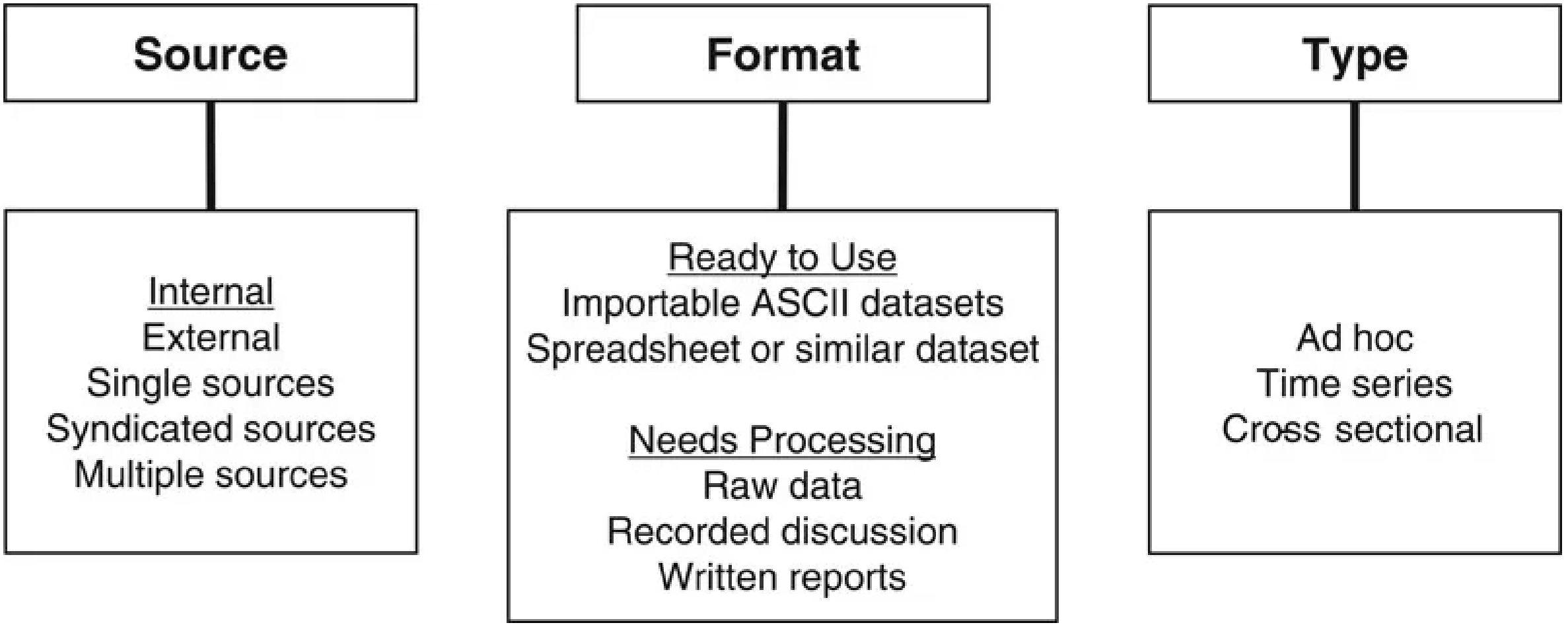
Types of secondary data



Secondary data may be qualitative or quantitative data and may be obtained inside an organization through its standard information-gathering processes or as the result of a prior specific piece of research. Alternatively, it can be sourced from external providers, including the originators of the research as well as organizations that consolidate and distribute data collected by others.

(Intellspot, 2023; Hair, Page and Brunsveld, 2019)

Categorizing and understanding secondary data



(Hair, Page and Brunsveld, 2019)

Evaluating the Quality of Secondary Data

Source of Data	Research Design	Data-Collection Methods
Reputation	Objective	Definition of constructs Examined relationships Report structure
Expertise	Sampling	Respondent intention and response rate Measurement techniques

(Hair, Page and Brunsveld, 2019)

Ensuring Validity of Secondary Data and avoiding Potential Bias

Measurement validity is difficult to establish with any degree of certainty, and judgment is often necessary when considering secondary data.12 For this reason, you should investigate the extent to which the approach you intend to take has been used by prior researchers who have experienced similar difficulties.

Measurement bias can also arise when a technique does not truly measure the topic of interest. For example, inflation data that is computed using a stable basket of goods suffers from this deficiency. Cost-of-living increases for the typical household in a given country will be correctly measured only if the basket of goods used to determine the consumer price index represents the actual purchasing preferences of an average household.

Numerous ethical dilemmas need to be considered when deciding whether to use secondary data. These include attempting to use the data when the specificity of the research question requires that primary data be obtained, insisting on collecting primary data when appropriate secondary data is inexpensive or even free, using secondary data gathered under guarantees of anonymity in a manner that may undermine that initial promise, and using secondary data that has been collected using questionable methods.

(Hair, Page and Brunsveld, 2019)

References on Secondary Data Collection

Hakim, C., 1982. Secondary Analysis in Social Research: A Guide to Data Sources and Methods with Examples. London: Allen and Unwin.

Intellspot, 2023. https://www.intellspot.com/secondary-data/

Hair, J., Page, M. and Brunsveld, N. (2019) Essentials of Business Research Methods. 4th edn. Taylor and Francis. Available at: https://www.perlego.com/book/1601160/essentials-of-business-research-methods-pdf (Accessed: 8 February 2023).

Quality Checks for Questionnaires: Reliability vs. Validity: Definition and Threats

Reliability

→ Does the measurement produce "stable and consistent results"? (Wilson 2014)

Threats to reliability:

- **Cross-application error:**
  It could be that a certain question format develops reliable answers in one setting but not in another (e.g. due to cross-cultural differences)
- **Subject error:**
  Are the target subjects (e.g. audiences) of your research the same or do they unknowingly change at different point of time (e.g. different groups of visitors at a certain sight)
- **Observer Influence:**
  Does the person of the observer / interviews changes create bias to the subjects questioned?

Validity

→ Is the measurement indeed measuring what it is supposed to measure?

Threats to validity:

- **Current events creating short-term changes of responses** (e.g. certain political events, a ransomware attack, etc.)
- **Effects of pre-tests with the same group on answers** (e.g. people whose performance is measured be dislike retesting; routines behaviours may determine the second application of the tool)
- **Mortality of participants in longitudinal research projects**
- **Changes in conditions and outer influences on test groups** (e.g. changing leadership or leadership style in a department)
- **Wrong attribution of cause and effect** (correlations do not determine causal relationship)

Wilson 2014, 129f

Quality Checks for Questionnaires: Approaches to Measuring Reliability

GISMA BUSINESS SCHOOL

Reliability

→ Does the measurement produce “stable and consistent results“? (Wilson 2014)

Inter-judgmental Reliability

What: Level of agreement between different assessors

When: In cases where subjective influences of researchers can create biases of results

How: Repeated application of the same data collection tool by different researchers to the same group of people, cross-comparison of results

Testing-and-Retesting Reliability

What: Consistency of results of the repeated application of a data collection tool

When: In cases where the consistency of data collection tools has to be measured

How: Repeated application of a data collection tool to the same group of people; cross- comparison of results

Parallel forms of Reliability

What: Consistency of results across different data collection tools

When: In cases where various methods with the same research focus used and tested for cross-method consistency

How: Application of different data collection tools for the same research focus; cross- comparison of results

Wilson 2014, 130

Quality Checks for Questionnaires: Approaches to Measuring Validity

Validity

→ Is the measurement indeed measuring what it is supposed to measure?

Internal Validity (Content)

What: Do the indicators represent the construct according to reason and judgment?

When: When sound judgement is required to determine the validity of a data collection tool

How:
a) Does each indicator reasonably reflect the part of the construct that it is supposed to reflect? ***(Face Validity)***
b) Is the choice of indicators reasonably reflecting the width of the construct? ***(Sampling Validity)***

Internal Validity (Construct)

What: Do the indicators represent the construct according to statistical calculations?

When: When a statistical argument for the validity of a data collection tool is expected

How:
a) Do indicators with similar characteristics correlate with each other while indicators with dissimilar don't?
b) Do the indicators help to distinguish groups that are also by theory distinguished?

External Validity

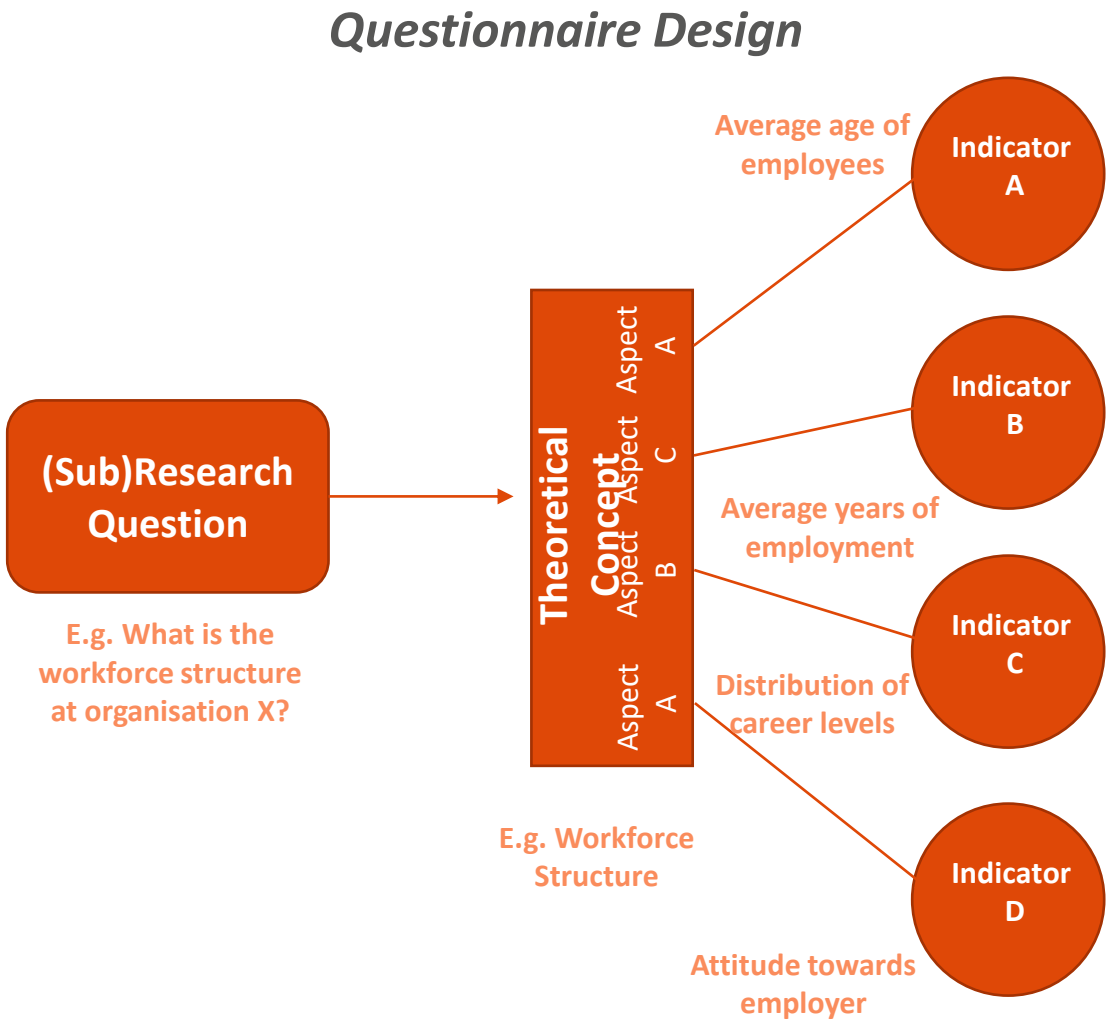
What: To what extent can findings from a study be generalised / transferred to other situations?

When: When research data is supposed to be used to formulate general principles / theories.

How:
Developing predictions / explanations or testing causal relationships in dissimilar settings based on the research results.

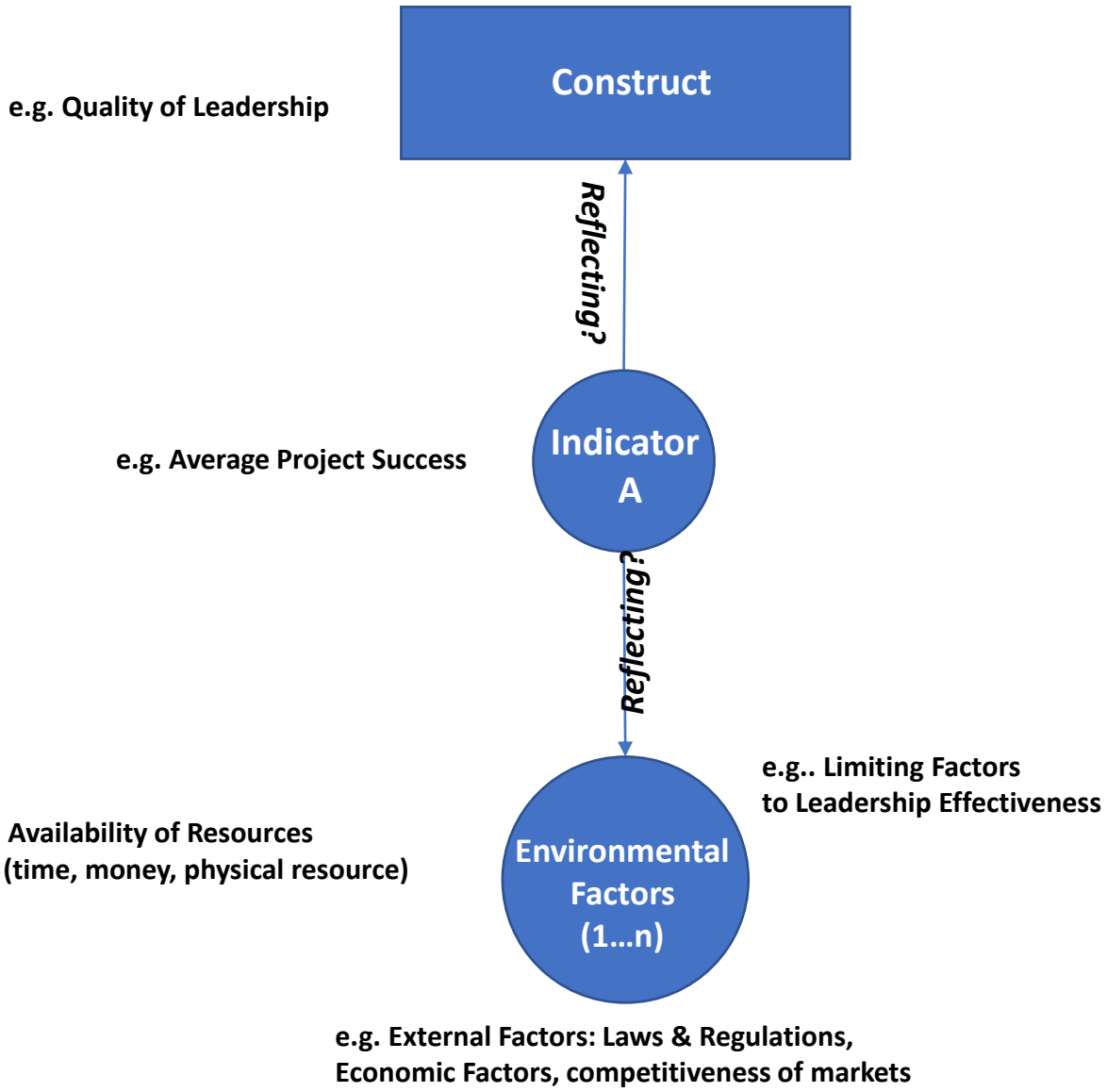
Wilson 2014, 133f; https://www.sciencedirect.com/topics/social-sciences/construct-validity (29.10.21)

Questionnaire Design 1: Principles of measurement

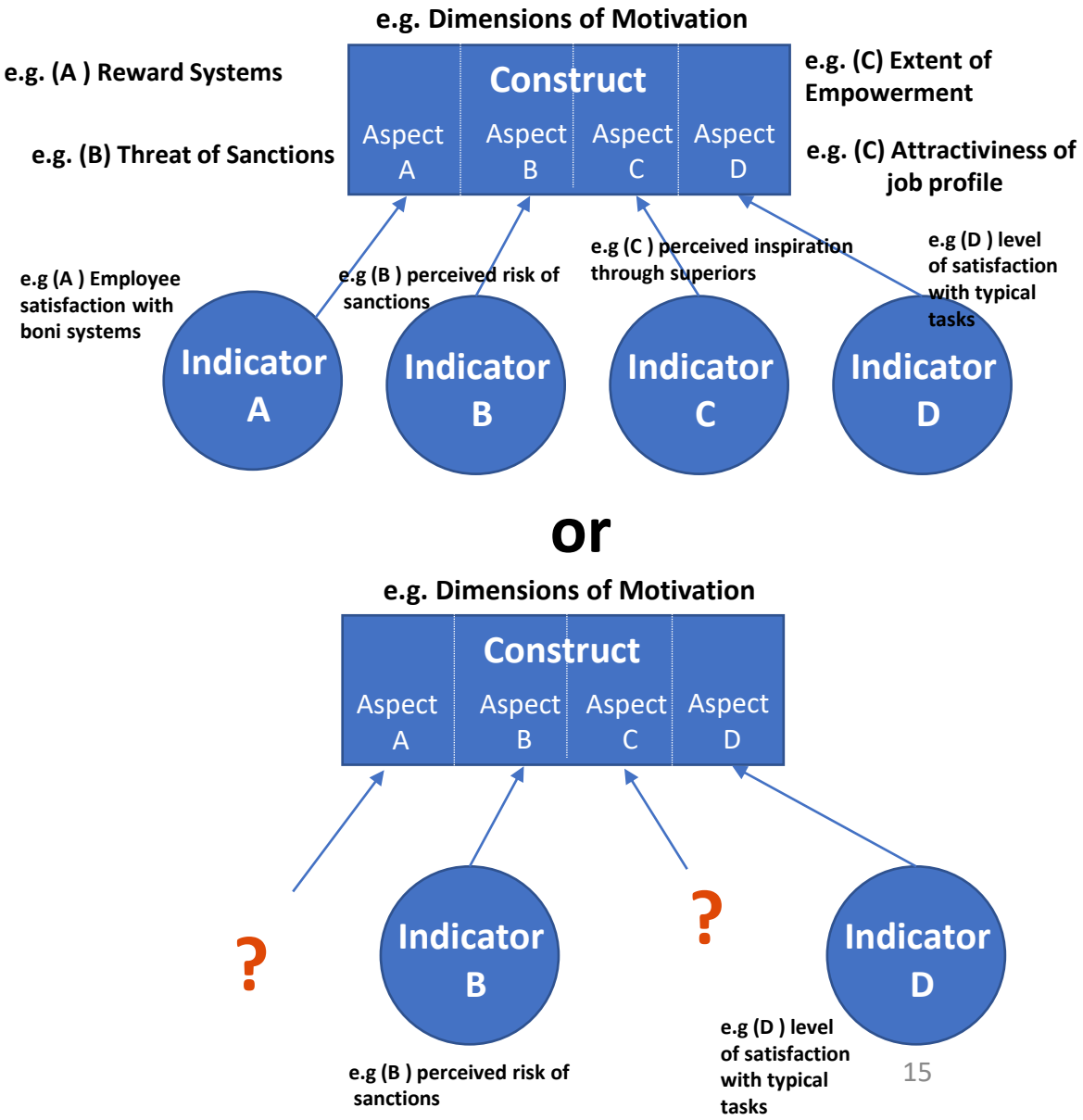


Understanding Internal Validity (Content)

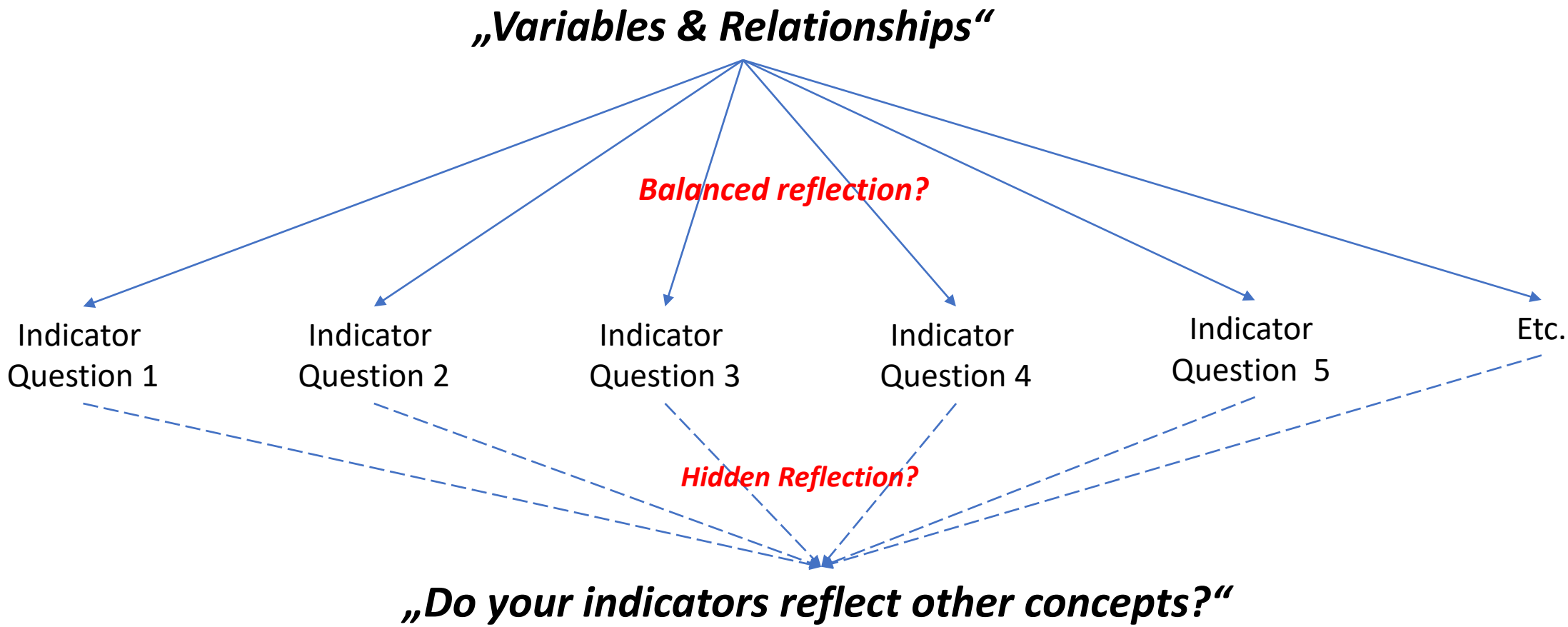
Face Validity



Sample Validity



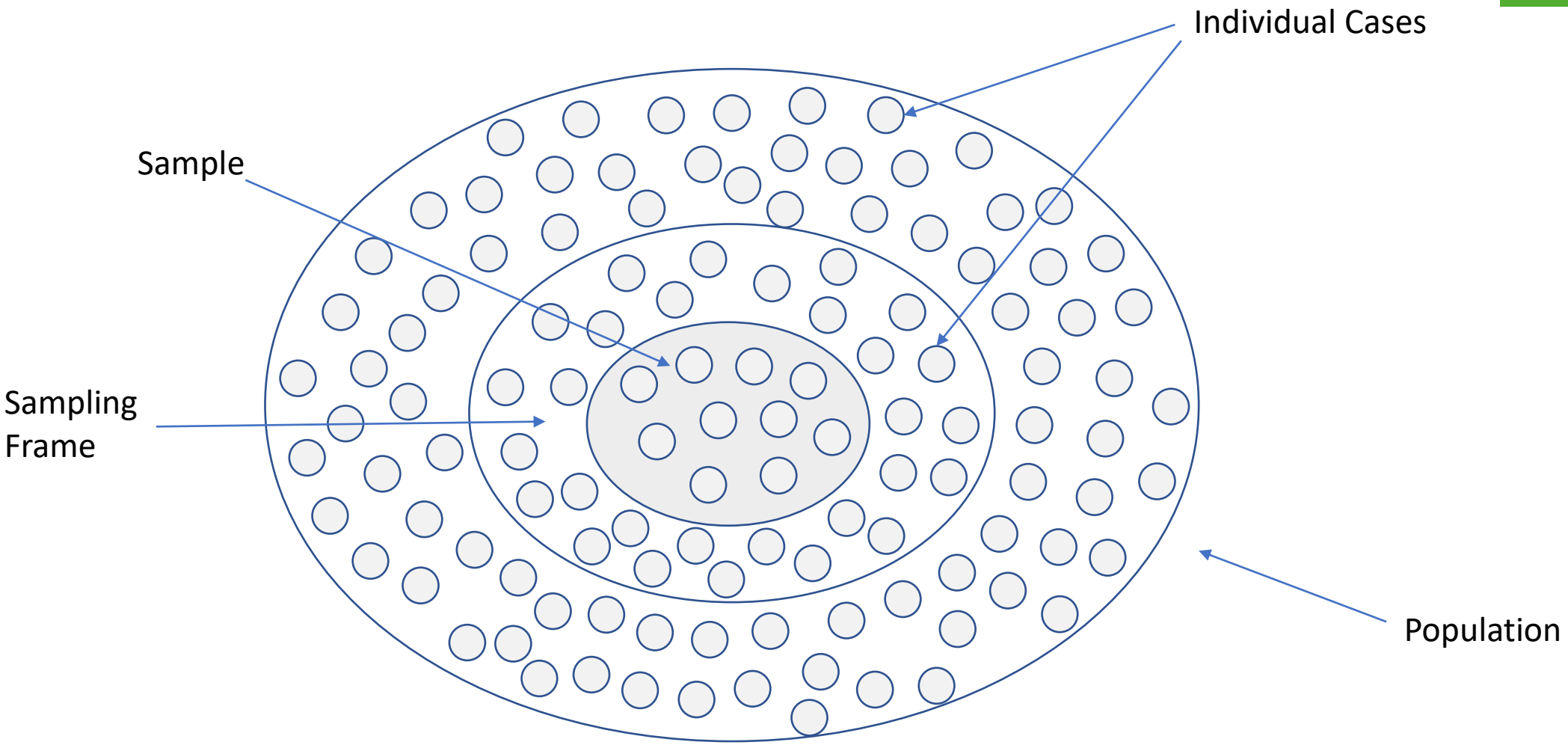
Group assignment: Determining Content Validity of your Data Collection Tool



Population and sampling

1. Population
   - Selection
   - Definition
2. Sample
   - Part of the population we collect data from
   - In quantitative research: should be representative
3. Sampling frame
   - Address list
   - Yellow pages, telephone book
   - Representative? Omissions should be random!
     - Incompleteness
     - Out of date

Population and Sampling



Probability vs. Non-Probability Sampling

Non-Probability Sampling

- **Quota Sampling:** Participants are chosen based on predetermined characteristics. The Distribution of characteristics should be the same like in the target population
- **Snowball Sampling:** A few participants are generated an asked to generate more participants
- **Convenience Sampling:** Participants are chosen based on their availability and readiness to participate
- **Judgmental Sampling:** Participants are chosen based on the judgment of the researcher

Probability Sampling

- **Simple-random Sampling:** Participants are chosen by stochastic means, while every subject has the same probability to be chosen
- **Systematic Sampling:** Every *n*th case is chosen from the sampling frame list. Sampling frame list needs to be randomly created.
- **Stratified Random Sampling:** Population is stratified, that is divided into sub-groups with peculiar characteristics, out of each a random choice is taken
- **Cluster Sampling:** Population is divided into clusters (e.g. according regions or locations), while in this cluster random choice is being applied.

GISMA
BUSINESS SCHOOL

... thank you for your attention!