POLS2044 WEEK 4
Concepts and Measures

Australian National University
School of Politics & International Relations
Dr. Richard Frank
https://richardwfrank.com/research_design_2022

Last week we discussed many ways of evaluating theories using qualitative approaches. In Week 4 we grapple with the opportunities and difficulties when trying to connect our theoretical concepts to the real world that we can measure (whether with words or numbers). Hopefully you have started to grapple with these issues as you wrote your problem statements in Week 3.

My goals for Week 4:

1. Highlight important considerations when moving from theory to evidence
2. Have you grapple with the challenges and opportunities involved in this process
3. Understand what is involved when we talk about measurement validity
4. Examine a few examples that you would think would be unproblematic
5. Apply what you learned about measurement to a particular topic in tutorial

---

I. Reading notes and questions

There are three readings for this week. I recommend that you start with the Adcock and Collier (2001) article, continue with reading Paxton (2000), and finish with Munck and Verkuilen (2002). The first article is the most important to spend time on, while the second and third articles contributions are more specific but still important. This week is all about figuring out how to connect our causal theories to measured real-world evidence that has demonstrated reliability and validity while minimizing systematic bias.

All three article citations are below and in the course guide. All articles should be downloaded from the ANU library website. See my video (https://youtu.be/Mf6a46uS8Rg) if you have any problems with downloading any of the articles.

Adcock and Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." ***American Political Science Review*** **95(3): 529-546.**

1. The authors start by highlighting "[t]he basic question of measurement validity: Do the observations meaningfully capture the ideas contained in the concepts?" (p. 529). What are the four challenges that can be addressed with greater attention to measurement validity?

2. Really try and work through the path diagram in Figure 1. (p. 531). Can you try and link this process from conceptualization to measurement for your own proposed qualitative research? What about starting to think about qualitative measures? We will be spending more time on the processes described in Figure 1 in lecture and tutorial.

3. "Contextual specificity" is a very fancy (and intimidating) term. However, in essence, what it means is that "differences in context potentially threaten the validity of measurement (p.

534). How might contextual challenges apply to your research? Are there any potential connections to the reasons (Week 3's video of) Professor Tickner believes that identity and culture are so context specific that comparative analysis of these issues is next to impossible?

4. What are the differences between content, criterion, and construct validity? (p. 537) Why are each of these three forms of validity important?

5. Why might too high or too low of a correlation between two variables be a potential problem for convergent validation?

6. (Re)read the quote by Campbell (1977/1988) on page 543, column 2. This quote reminds me of the Ship of Theseus thought experiment (Google or Duck-Duck-Go it if you have not heard about it). Adcock and Collier's (2001) point here is that both causal theories and measurements can be questioned or exchanged for others. However, to make sure that we do not (metaphorically) sink while doing our research, we need to focus on replacing only a small number of planks at a time. Have you thought about the importance of focusing on one element while holding the rest constant when conducting your research or writing?

Paxton, Pamela. 2000. "Women's Suffrage in the Measurement of Democracy: Problems of Operationalization." ***Studies in Comparative International Development*** **35(3): 92–111.**

Adcock and Collier (2001) highlight Paxton's (2000) work as a good example of how democracy measures can falter when they move from concept to measurement. Paxton (2000) focuses on how some democracy measures only include universal male suffrage and how this can lead to theoretical and empirical problems.

7. What are the three areas of research that Paxton (2000) argues are affected by only looking at male suffrage?

8. What are three of the components of democracy that Paxton (2000) describes? Which one does Paxton focus on?

9. What are the empirical implications for Paxton's shift in dates of democratization to include women's suffrage?

Munck, Gerardo L., and Jay Verkuilen. 2002. "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices." Comparative Political Studies 35(1): 5–34.

10. Table 2 and Figure 1 are worth spending time on as they succinctly connect conceptualization, measurement, and aggregation. The previous two readings focus on conceptualization and measurement, and they do not spend much time on aggregation. What does aggregation mean? What forms of aggregation are there?

12. Despite the issues of conceptualization, measurement, and aggregation discussed in most of this article, Munck and Verkuilen (2002) do recognize that the reviewed indices are still quite highlight correlated. Why is this the case?

13. The authors (p. 29) quote Bollen (1986) in saying that "reliability should not be confused with validity." What sort of validity do you think they are talking about? Hint, look at the forms of validity highlighted in question 4 above.

Lecture PART 1: INTRODUCTION & RECAP

Weeks 1-3 recap

1. Using the scientific method
   Be curious about the world and write your research down.
2. Theorise using causal inference
   Our theories need to cross causal story hurdles.
3. Linking theories to evidence using qualitative analysis
   Thick description possible with a narrow case focus.

Developing theories—summary

Offer an answer to an interesting research question.
Solve an interesting puzzle.
Identify interesting variation (across time or space)
Move from a specific event to more general theories
Drop the proper nouns
Use a new Y
Use a new X
Add a new Z
Use the literature
Make sure the theory can be disproven.

Building on the literature

What causes might be missed/overlooked?
Can theories be used elsewhere?
What are future implications?
Does it apply at a different unit of analysis?

What are political science methods?

Political science methodology provides tools for answering questions about how and why the world works the way it does.
This involves the analysis of descriptive indicators for causal inference.
The goal is not normative judgement of what is "good" or "bad" or how the world should operate.
There is often a U-shaped relationship between the number of cases and the number of published studies. (Ragin et al. 1994)

Graph of various methodological approaches depending on the number of variables and cases chosen

Four hurdles to establishing causality

1. Is there a credible mechanism connecting X and Y?
2. Can we rule out Y causing X (endogeneity)?
3. Is there covariation between X and Y?

4. Have we controlled for potential spuriousness (Z)?

A list of a bunch of qualitative study examples
Why should we care about measurement?

We can fall in love with stories we tell ourself about the world; however, if these stories remain untested assumptions we have no idea whether they are true or false.

Today's motivating questions

How can we link solid causal theories to real-world evidence?
How can we be sure this evidence has measurement validity?

Motivating puzzle

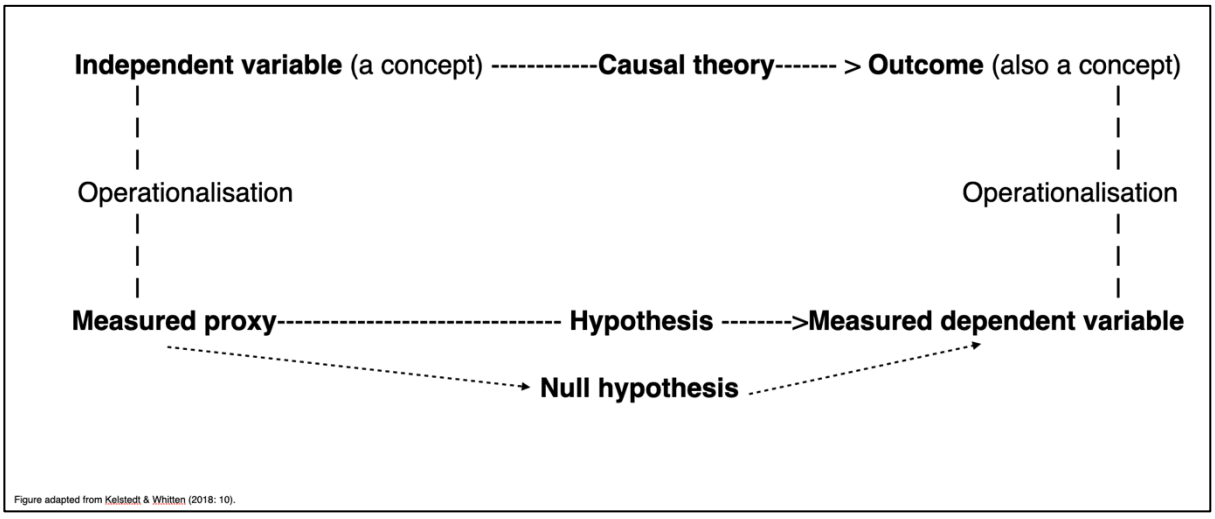
Most people use real-world data without thinking about how they are generated and whether they capture what they think they do.

Video excerpt from Princess Bride (1987)
https://youtu.be/D9MS2y2YU_o

LECTURE PART 2: Moving from concept to measure

Moving from theory to test

What is the difference between a theory and a hypothesis?



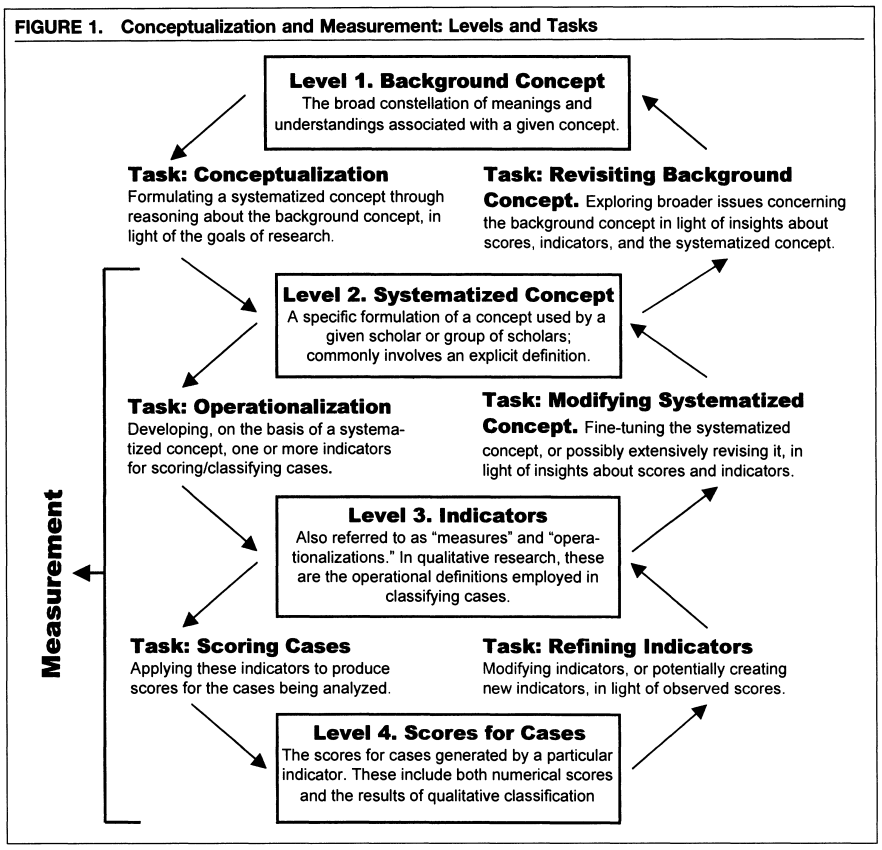
The null hypothesis is akin to the legal principle of the presumption of innocence.

Slide of tweet saying how the lack of data is making it hard to build early warning alerts for climate change.

From theory to measurement

Causal theories are relationships between concepts.
Measurement is trying to observe these concepts (or their proxies) in the real world.

Adcock and Collier (2001: 531) path diagram



In other words:

Step 1: Conceptualise
Step 2: Operationalise
Step 3: Code

Where is Canberra's best coffee shop?

1: Conceptualise coffee shop characteristics
2: Operationalise these characteristics
3: Code as many coffee shops as possible

Conceptual/coding challenges

Are observations heterogeneous or homogenous?

Short excerpt from This is Spinal Tap (1984)

https://youtu.be/KOO5S4vxi0o

Any gap between theory and measurement here?

Causal threat: measurement error

Measurement error is the difference between the measured value and the true value of something.

It includes both a random component and the potential of a systematic component

Causal Threat: Systematic error

Video about former President Trump's net worth (https://youtu.be/0XxoFV_ktzk)

Type 1 and Type 2 errors

False positives and false negatives

Random error

Erik Gartzke. 1999. "War Is in the Error Term." *International Organization* 53(3): 567-587.

Random measurement error

"People are not very good at understanding randomness. There's much more chance out there than we think there is. While we are seeking for patterns and explanations as we look backward, we're not giving a fair shot to the explanation that many things are really just random events."
— Lisa Goldberg

Pareidolia—Seeing things that are not there.

New Hampshire's Old Man of the Mountain

Challenges to measurement

Conceptual clarity—Do we know what we want to measure?

Operational reliability—Are the measures repeatable and consistent?

Conceptual validity—Does the measure accurately measure the concept we are trying to measure?

Validity

Face validity—On its face does a measure appear to be measuring what it says it is measuring?

Example of North Korea's constitution

Content validity—Does a measure capture all of the systemised concept? Is anything missing? Is anything there that should not be?

Example: Polity V does not include measures of participation

Criterion validity—Does a measure correlate with criterion (i.e. ground truth) variables?

Examples of "Le Grand K" and political polling

Construct validity—Do measures behave the way you theoretically expect in the wild?

Example of the relationship between governance and democracy

Validity and reliability

Can you think of a valid but unreliable measure?
A reliable but invalid measure?

LECTURE PART 3: A few examples

Example 1: GDP

Definition from the World Bank

(https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?view=chart)

CNBC video about the GDP definition

(https://youtu.be/iLom1WlqwS0)

Joseph Stiglitz article in Scientific American "GDP is the wrong tool for measuring what matters"

Bhutan's Gross National Happiness Index

(https://ophi.org.uk/policy/gross-national-happiness-index/#:~:text=The%20phrase%20'gross%20national%20happiness,approach%20towards%20notions%20of%20progress)

Population Definition from the World Bank

(https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?view=chart)

Australia's definition of population

(https://population.gov.au/population-topics/topic-population-measurement)

Article example of looking at Sweden's population according to a subnational grid

Example of PRIO-GRID 2.0's grid population data

Example: Democracy

When is a country a democracy?

How is democracy a latent or unobservable concept?

V-DEM's conceptual scheme

Munck & Verkuilen's (2002: 8) process

Table 2
A Framework for the Analysis of Data: Conceptualization, Measurement, and Aggregation

Challenge	Task	Standard of Assessment
Conceptualization	Identification of attributes	Concept specification: Avoid maximalist definitions (the inclusion of theoretically irrelevant attributes) or minimalist definitions (the exclusion of theoretically relevant attributes)
	Vertical organization of attributes by level of abstraction	Conceptual logic: Isolate the "leaves" of the concept tree and avoid the problems of redundancy and conflation
Measurement	Selection of indicators	Validity: Use multiple indicators and establish the cross-system equivalence of these indicators; use indicators that minimize measurement error and can be crosschecked through multiple sources
		Reliability
	Selection of measurement level	Validity: Maximize homogeneity within measurement classes with the minimum number of necessary distinctions
		Reliability
	Recording and publicizing of coding rules, coding process, and disaggregate data	Replicability
Aggregation	Selection of level of aggregation	Validity: Balance the goal of parsimony with the concern with underlying dimensionality and differentiation
	Selection of aggregation rule	Validity: Ensure the correspondence between the theory of the relationship between attributes and the selected rule of aggregation
		Robustness of aggregate data
	Recording and publicizing of aggregation rules and aggregate data	Replicability

Note focus on:
1. Conceptualisation
2. Measurement (including reliability, validity, and replicability)
3. Aggregation (often overlooked)

Munck & Verkuilen's (2002: 13) process

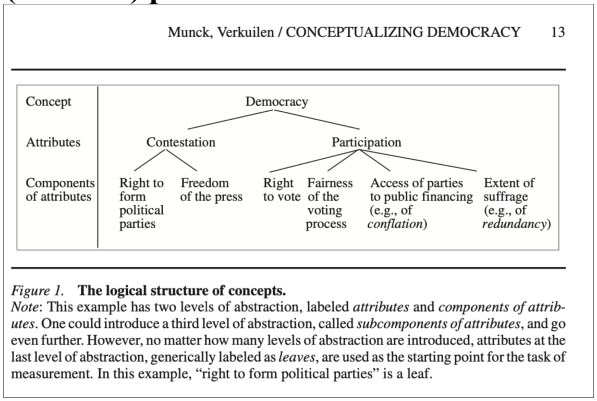


Figure 1. **The logical structure of concepts.**
Note: This example has two levels of abstraction, labeled *attributes* and *components of attributes*. One could introduce a third level of abstraction, called *subcomponents of attributes*, and go even further. However, no matter how many levels of abstraction are introduced, attributes at the last level of abstraction, generically labeled as *leaves*, are used as the starting point for the task of measurement. In this example, "right to form political parties" is a leaf.

Example 4: Kenya's election

Competing for 50% of the vote

Map of political disorder in Kenya (ACLED)

Source: ACLED (https://acleddata.com/2022/08/09/kenyas-political-violence-landscape-in-the-lead-up-to-the-2022-elections/)

August 14 SABC video from Kenya

(https://youtu.be/54NES--dFlw)

Example of competing explanations for government removal of national security forces

More prosaically

A tweet from the Victorian Election Violence UK database showing common place names.

Example 5: Natural resources and civil conflict onset

Taken from Humphreys (2005)

Mechanisms connecting natural resources & civil conflict onset

Mechanisms
1. Greedy rebels
2. Greedy outsiders
3. Grievances
4. Feasibility
5. Weak states

6. Sparse networks

Proxies
1. Diamond production
2. State instability
3. State instability x autocracy
4. Oil reserves (per capita)
5. Oil production (per capita)
6. share of agriculture (% GDP)

Measurement and transparency

Is reporting data a real measure of governmental transparency? Or capacity?

Today's motivating questions

How can we link solid causal theories to real-world evidence?

How can we be sure this evidence has measurement validity?

III. WEEK 4 TUTORIALS

The focus of today's tutorials is on applying the readings and lecture material to your own research.

Remember, students will only receive credit for completing the questions below on Wattle if either (1) the tutors also record of you having attended the tutorial or (2) you have a documented medical reason for not being able to attend tutorial this week.

Part 1: Individual work (~five minutes)

The first part of class is trying to get us to think about an underlying concept and how we might measure it in the world. For this week, the concept is: corruption. This is a tough one because those that are corrupt have every incentive to hide their corruption from anyone who could punish them for it either legally or reputationally.

1. Write five questions or prompts that you could use when interviewing subjects about their experience with governmental corruption.

Submit your questions to Wattle's link before you move on to the group work.

Part 2: Small group work (~30 minutes)

Next come together in groups of three students. Read off your questions/prompts to each other.

There is no need to upload anything to Wattle for this or anything below.

2. Are there any notable similarities or difference between your responses?

3. How did your questions map onto your underlying assumptions about the relevant concepts and possible measurements of "corruption"?

Now go to Transparency International's (TI) most recent report on global corruption (https://www.transparency.org/en/cpi/2021).

Look at some of the country rankings and then the methodology section (page 15).

4. Is this measurement methodology clear?

5. Do you think it is methodologically valid and reliable?

Maybe we are being unfair. If you poke around their website a bit, you can find and download their "Technical Methodology Note" which has more detail.

6. Does this level of detail help clarify things?

7. What do you think about TI's aggregate efforts?

Finally, go to the World Bank's Enterprise Surveys website which focuses on their corruption surveys (https://www.enterprisesurveys.org/en/data/exploretopics/corruption).

Try sorting the summary table by indicator 12 ("Percent of firms identifying corruption as a major constraint")

8. Are there any notable similarities or differences between this ranking and that of TI?

9. More generally, do you think these questions link more (or less) directly to your conceptualisation of corruption in your questions you started with for question 1 above?

Part 3: The whole tutorial group (whatever time remains)

10. What were the main takeaways of your individual and small group activities conceptualising, measuring, and aggregating corruption?

11. What connections can you draw between the lecture and tutorial discussion of concepts and measures to your own research?

12. What questions, concerns, or takeaways did you have after completing your problem statements last week?