

T407016 // AUTUMN 2026 // UNIVERSITÉ DE GENÈVE

COMPARATIVE POLITICS

Complement 01. Concepts, measurement and validity.

TUE 27 OCT 2026

UNI MAIL M2170

DR. JIHYE PARK

NO READING TODAY

Four measures you have already met

SESSION 02 · HANSON AND SIGMAN

Capacity, one latent score built from 21 indicators of extraction, coercion and administration.

SESSION 03 · MUNCK AND VERKUILEN

Nine democracy indices, judged on how they define, measure and combine.

SESSION 04 · LÜHRMANN AND LINDBERG

An autocratization episode, coded as a fall of at least 0.10 on V-Dem's electoral democracy index.

SESSION 06 · LIJPHART, CAREY AND HIX

Gallagher's index of disproportionality and the effective number of parties.

Today turns those readings into a toolkit. **Use it on any number before it goes into a policy brief.**

OPENER · ANY VOLUNTEERS?

You advise a foreign ministry

A donor will cut aid to any country whose governance score falls two years in a row. Can the score be trusted?

01

What is the score supposed to capture?

02

Where do the numbers come from?

03

What else could make the score fall?

MUNCK AND VERKUILEN · THREE CHALLENGES

From concept to score

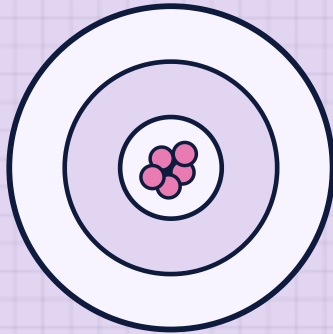


STEP	DEMOCRACY, VANHANEN	DISPROPORTIONALITY, GALLAGHER
Concept	Democracy	How faithfully seats follow votes
Attributes	Competition and participation	Each party's gap between votes and seats
Indicators	Vote share of the largest party, and turnout	Official vote and seat shares
Index	The two attributes multiplied	Square root of half the summed squared gaps

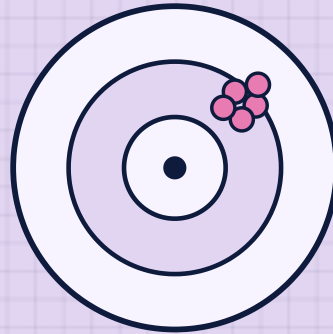
Conceptualization covers the first two steps, measurement the next two and aggregation the last. **A mistake early in the chain carries into every later step.**

TWO STANDARDS

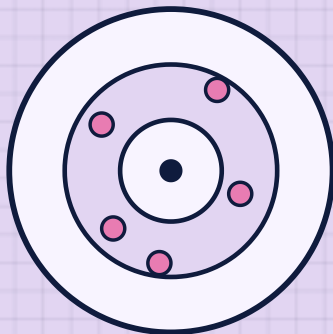
Valid, reliable, both or neither?



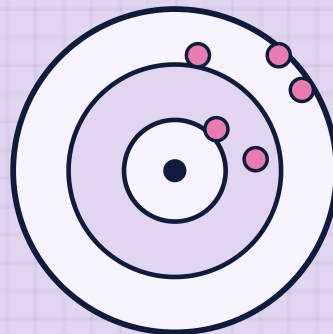
RELIABLE, VALID



RELIABLE, BIASED



NOISY, CENTERED



NOISY, BIASED

- 01 Validity is "the extent to which the proposed measures actually measure what they are supposed to measure."
- 02 Reliability is the prospect that the same data collection process would always produce the same data. Agreement among coders is the usual test.
- 03 "Reliable measures need not be valid ones." In fact, "one way to obtain very reliable measures is to adopt similar biases."

The bullseye is the concept. Each dot is one attempt to score the same case.

MEASUREMENT ERROR

Noise or bias?

RANDOM ERROR

Scatter in both directions

A coder misreads a source, or a survey draws an unlucky sample. The errors cancel out on average, and they make real relationships harder to see.

SYSTEMATIC ERROR

A push in one direction

The error follows some other feature of the case. Collecting more of the same data leaves it in place.

Munck and Verkuilen's examples of systematic error

Reported rapes may track changes in culture more than the number of rapes.

More evidence of corruption may reflect a freer press.

Vanhanen's turnout and largest-party share build bias into his index.

"The record left by history is inherently biased." **Choose indicators less exposed to bias and check them against several sources.**

WHEN INDICATORS DRIFT

The number moves, the concept stays put

ACROSS PLACES

An indicator can mean different things in different countries, so Munck and Verkuilen ask for cross-system equivalence. Police officers per person barely track Capacity ($r = .03$), perhaps because weak rulers hire more of them.

OVER TIME

After 1989 Freedom House moved from a single coder to a team coding components. A shift in the series may come from the method.

UNDER PRESSURE

Once money or reputation rides on a score, governments have reasons to move the indicator itself. Relabeling contracts is quicker than cleaning them up.

Drift is hard to see in the scores themselves. **Ask what else could have moved the number.**

AGGREGATION

Same gaps, different verdicts

TWO INVENTED ELECTIONS • GAPS BETWEEN VOTE AND SEAT SHARES, IN POINTS

ELECTION	GAPS	SUM OF GAPS	GALLAGHER
X	One party +10, two at −5	20	8.7
Y	Five parties +2, five at −2	20	4.5

UNITED STATES, LIJPHART'S TABLE 8.1

Congress and president	4.43 and 46.03
Arithmetic mean	25.2
Geometric mean	14.3

- 01 A plain sum cannot tell a few large distortions from many small ones. Lijphart says it understates how proportional PR is.
- 02 Squaring the gaps makes the large ones count more.
- 03 An arithmetic mean would let the presidential race swamp the legislature, so Lijphart uses the geometric mean.

Session 03 asked whether to add or multiply. **Every rule for combining numbers is a claim about the concept.**

GROUP EXERCISE 01 · DIAGNOSE THE SCORE

A donor ministry asks your group for advice

The ministry wants to reward partner countries that cut corruption. Would this score show which ones did?

The draft score averages five indicators with equal weights. Each is rescaled from 0 to 100 so that a higher value means more corruption.

#	INDICATOR	SOURCE
1	Corruption convictions per 100,000 people	Court records
2	Newspaper stories about corruption per year	Media archive
3	Expert rating of corruption, 0 to 10	A consultancy's country analysts
4	Share of adults who paid a bribe in the past year	National survey
5	Share of public contracts awarded without open tender	Procurement agency

Across partner countries last year, indicators 2 and 3 correlated at .85.

- 01 For each indicator, name one thing besides corruption that could move it. Is that error random or systematic, and which indicators could rise when corruption falls?
- 02 Indicators 2 and 3 agree closely. Does their agreement make the score valid?
- 03 Rebuild the score with at most three indicators. Say how you would combine them and name one check to run before the ministry uses it.

The ministry, score and data are invented.

Four ways to check a measure

CHECK	QUESTION	HANSON AND SIGMAN'S CAPACITY
Face and content	Do the scores look right, and do they cover the whole concept?	Singapore ranks in the top 25, and Somalia, Yemen and the Central African Republic sit near the bottom
Convergent	Does it agree with other measures of the same concept?	$r = .91$ with the World Bank's government effectiveness index
Discriminant	Does it stay apart from different concepts?	Low correlations with conflict, population and oil output, and $r = .79$ with log GDP per capita
Construct	Does it predict what theory says it should?	One point more Capacity goes with about 21 points more undeliverable letters returned

They follow Adcock and Collier (2001). **Each check answers a different doubt, and none of them settles every doubt.**

Why agreement proves little

- 01 Shared sources. Many democracy indices drew on the same precoded data, so a high correlation "may mean that all indices are reflecting the same bias."
- 02 Reliability at most. "One can get very consistent (i.e. reliable) measurements that are not valid," Bollen warned in 1986.
- 03 Aggregate scores. The two Freedom House series line up with the other indices on one dimension and load alone on a second.

"The high degree of correlation among existing democracy indices does not put to rest concerns about their validity."

Agreement counts for more when the other measure rests on a different kind of data. **Hanson and Sigman's census-based Myers index is one example.**

From ideal to cut-off

IDEAL	INDICATOR	GOOD OUTCOME
Representation	Disproportionality	Below 5
Representation	Voter-government distance	Below 5
Accountability	Effective number of parties	Fewer than 3
Accountability	Parties in government	Two or fewer

- 01** A value counts as good when it beats the median of their data.
- 02** Their two measures of representation correlate at only .16.
- 03** Across Lijphart's 36 democracies, disproportionality and the effective number of parties correlate at -0.57.

Lührmann and Lindberg called their own 0.10 cut-off "naturally arbitrary." **Cut-offs are measurement choices too.**

GROUP EXERCISE 02 · TEST THE ACCOUNTABILITY MEASURE

Thursday's reform commission wants a second opinion

Do fewer parties really mean more accountable government?

The commission plans to rank Options A, B and C on accountability with Carey and Hix's two measures, the effective number of parties and the number of parties in government. Carey and Hix justify the measures this way.

CAREY AND HIX, P. 388

The number of parties in the first cabinet "serves as an index of how difficult it is for voters to attribute responsibility for government performance and reward or punish governing parties at the next election accordingly."

CAREY AND HIX, P. 384

Empirical studies indicate "that both prospective and retrospective voting are more effective when coalitions include fewer parties."

- 01 Content. What does accountability mean here, and what does a count of parties leave out? Describe a country that scores well on both measures where voters still cannot remove the government.
- 02 Convergent and discriminant. Name one other measure the party counts should agree with, and one different concept they should not simply track. What does Lijphart's -0.57 do to Carey and Hix's trade-off?
- 03 Construct. If fewer parties do mean more accountability, what should happen to governments after a bad term? Sketch a test and the result that would count against the measure.

TAKE-AWAY

Six questions for any number

- 01 What is the concept, and is it defined too broadly or too narrowly?
- 02 Which indicators stand for it, and what else moves them?
- 03 Is the error random or systematic?
- 04 Does the indicator mean the same thing across countries and years?
- 05 How are the parts combined, and where do the cut-offs sit?
- 06 What validity evidence exists beyond agreement with similar measures?

Complement 02 on 17 November turns to hypotheses. **Every test it covers depends on measures like today's.**

NEXT UP

Before Thursday 29 October

Week 07. Parties and representation.

READ

Katz, Richard S., and Peter Mair. 1995. "Changing Models of Party Organization and Party Democracy: The Emergence of the Cartel Party." *Party Politics* 1(1), 5-28.



Thu 29 Oct

Quiz 3 opens the class. It covers Sessions 05 and 06.

UNI DUF0UR U300



Tue 17 Nov

Complement on hypotheses and how they are evaluated

M2170

PROGRESS SAVED

QUESTIONS?

Office hours Thursdays 14:30 to 16:00 in M4328

JIHYE.PARK@UNIGE.CH

JIHYEPARK.ME/COMPARATIVE-POLITICS-MASPO-FALL2026