# PLSC 308: Introduction to Political Research

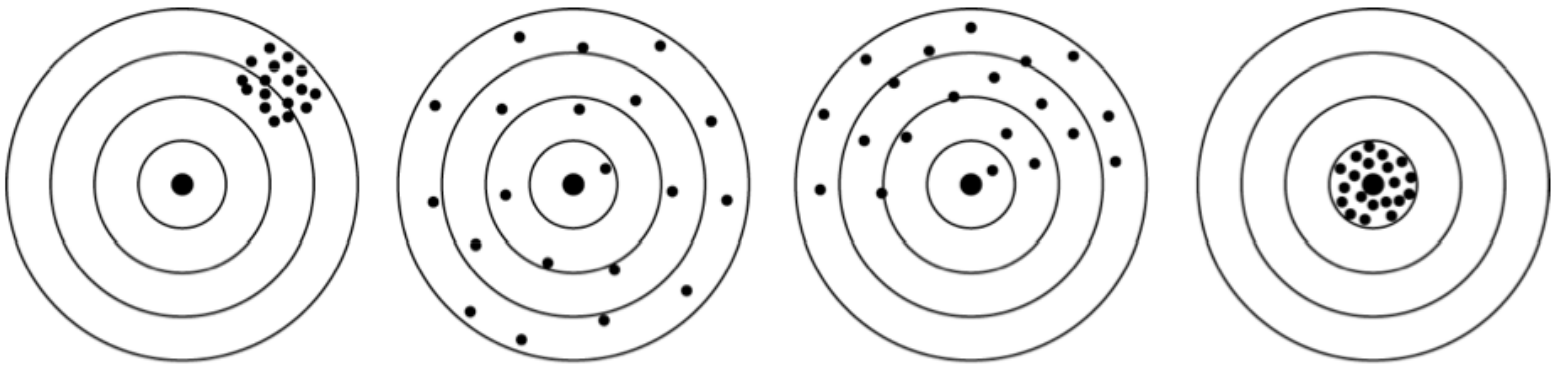
Christopher Zorn

February 11, 2016

# Measurement: Desiderada

- Reliability
- Validity

# Obligatory Reliability-Validity Graphic



# Reliability

> *"...whether a particular (measurement) technique, applied repeatedly to the same object, would yield the same result each time."*
>
> *– Babbie (1995, 124)*

Threats to reliability:

- Vague conceptualization / operationalization
- Variation in instruments / assessors
- Precision of instrument
- Change over time / space

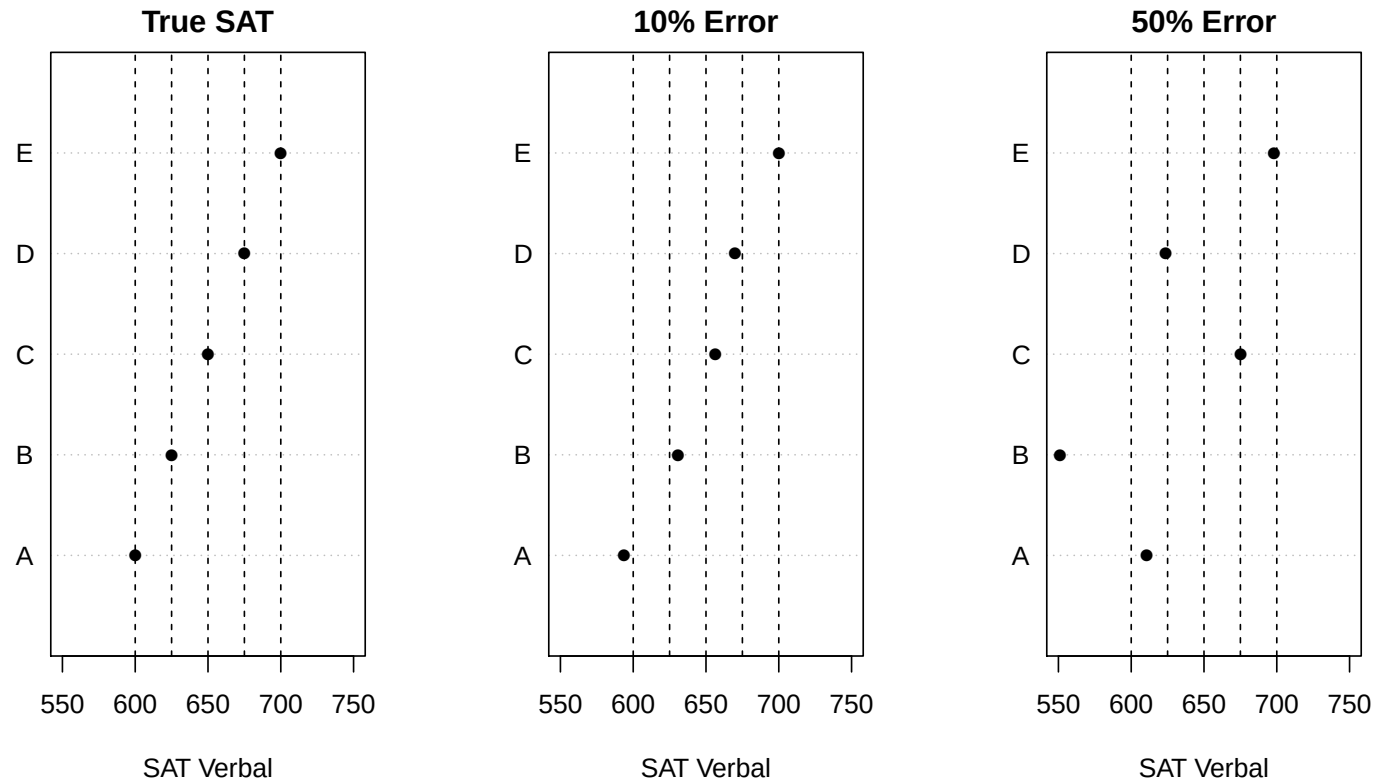
# Assessing Reliability: Methods

- Test-Retest: Measure the same thing(s) twice...
  - Pros: Most valid  robust
  - Cons: *Cost...*
- Alternative-Form: Test-retest with different measures
  - Pros: Tests different indicators; can be less costly
  - Cons: No direct measure of reliability
- Split-Half: Alternative-form on "split" data
  - Pros: Much less costly
  - Cons: Only suggestive

# Unreliability: Consequences

Unreliability $\longrightarrow$ Low "signal to noise" ratio...

SAT example:



# Validity

*"A valid measure accurately represents the concept it is supposed to measure."*

*– Kellstedt and Witten (2009, 94)*

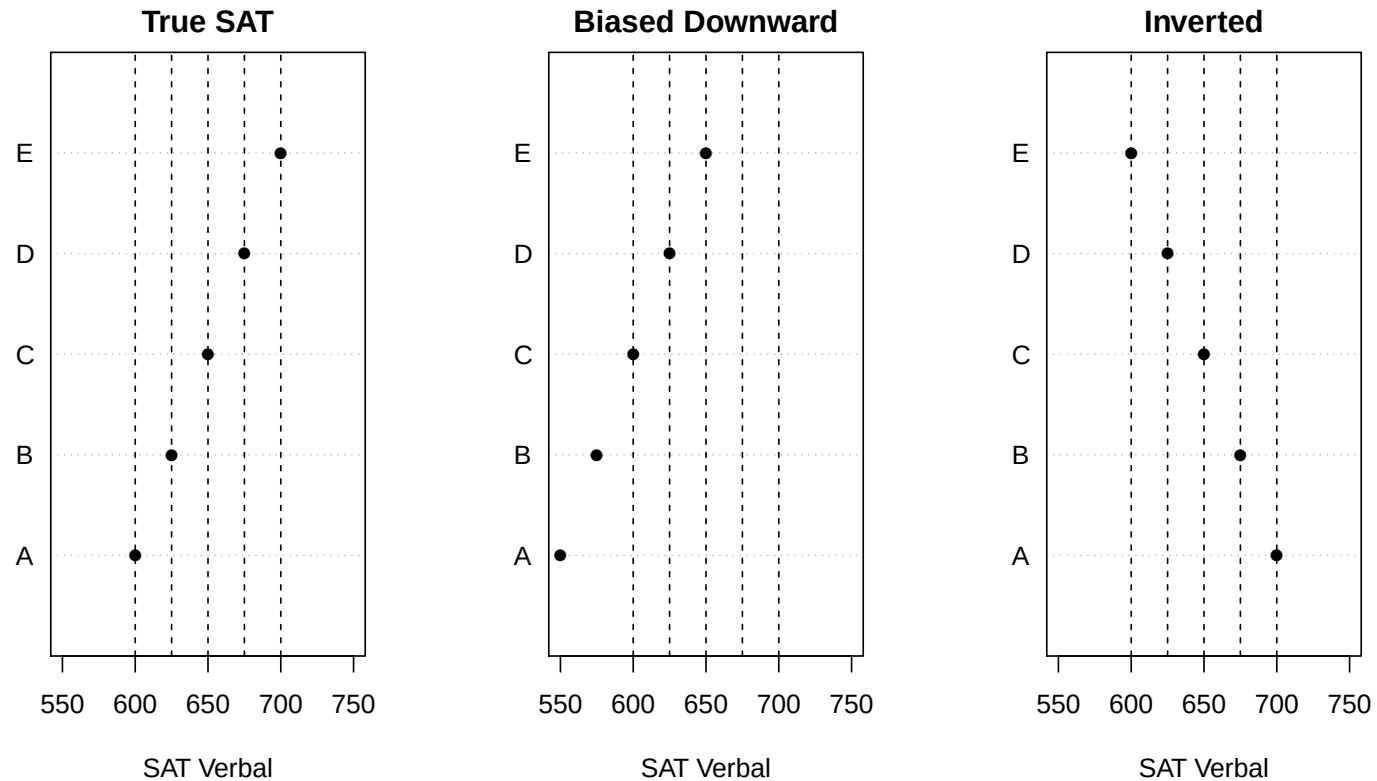
# Flavors of Validity

- Face Validity: The measurement subjectively seems to reflect the concept.
- Content Validity: The measurement contains/captures all facets/dimensions of the concept.
- Construct Validity
  - *Convergent*: The measurement is associated with other measures of the same or related concepts.
  - *Discriminant*: The measurement is not associated with other measures of putatively unrelated concepts.

# Lack of Validity: Consequences

Invalidity $\longrightarrow$ "Bias" in measurement vis-a-vis concept...

SAT example:



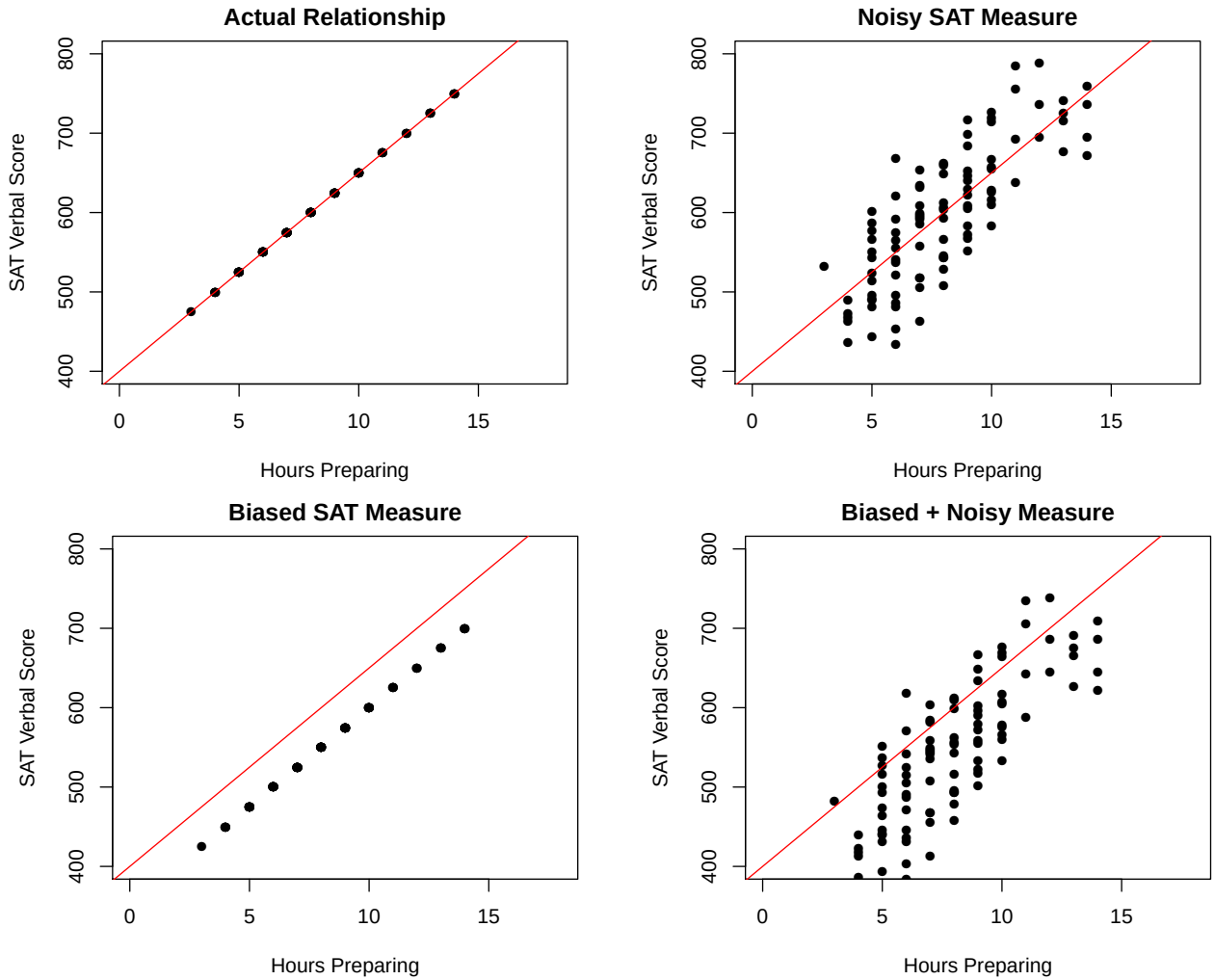
# Summary: Consequences

In general:

- Lack of reliability harms the precision of our observed relationships.
- Lack of validity biases the relationships we observe.

# Consequences, Illustrated

SAT redux:



# Measurement: Multiple Indicators

The Good:

- Capture different *dimensions* of a concept (social vs. economic vs. foreign policy liberalism)
- Possibility to build *scales / indices* (see below)

The Bad:

- Demand careful thought about interrelationships
- Costly (time, effort, etc.)

# Scales / Indices

Combine multiple separate measures into a single summary index.

*Additive* index: Individual measures...

- ...need to be components of the larger concept
- ...should all be equally-strongly related to the larger concept.
- ...need to be measured or converted to common units / scale

# Scale Example: "Size"

| Name | Height (in) | Weight (lbs) | Shoe | Hat |
|---|---|---|---|---|
| Aziz | 70 | 155 | 9.5 | 7.0 |
| Brendan | 64 | 150 | 11.5 | 7.125 |
| Cam | 68 | 185 | 11 | 7.5 |
| Dingxiang | 74 | 210 | 12 | 7.375 |

Things:

- Necessity of rescaling...
- Which (if any) are more or less related to the core concept?

# Practical Things: Naming Variables

- Make variable names descriptive (`CaseID`, `name`, etc.... not `Var1`, `Var2`...)
- Give directional names where appropriate (`female`, not `gender`; `conservatism`, not `ideology`, etc.)
- Include units if possible (`PopInMillions` vs. `Population`; `PercentUnionized` vs. `Unionized`, etc.)
- Adopt a consistent naming convention for variables (e.g., `.Lag` for lagged values, etc.)