Construct Validity: Credibility When Measurement Is Cheap

Melissa Dell Ashesh Rambachan

NBER Methods Lecture 2026

Slides posted on the NBER Methods Lecture website.

Construct Validity

How should a theoretically meaningful concept be operationalized?

This stage involves **conceptual uncertainty**: whether a proposed operationalization captures the construct required for a given research question.

Economists increasingly measure **novel, latent, and high-dimensional concepts** that are not directly verifiable.

AI Changes the Costs of Assessing Construct Validity

1. More candidate definitions

AI can make it cheaper to generate alternative measures using different:

- rubrics, codebooks, dictionaries, and prompts
- model-based operationalizations

LLMs can also help surface ambiguous edge cases that reveal where the construct remains underspecified.

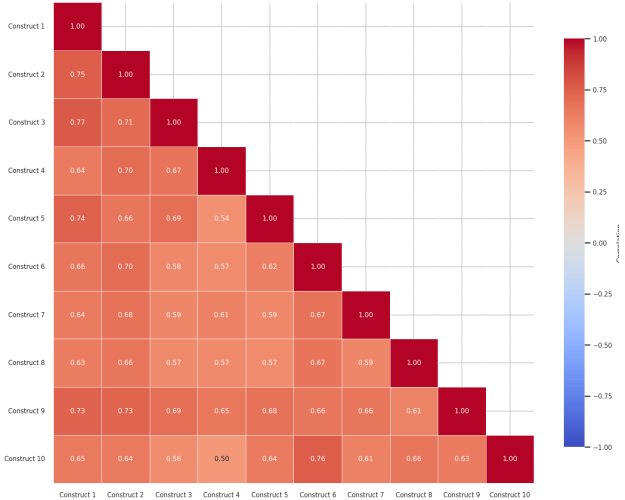
2. More evidence for validation

AI can make it cheaper to construct measures to assess construct validity:

- theoretically relevant correlates
- contrasts across groups
- other observable implications of the construct

This expands the evidence available for assessing whether an operationalization captures the intended construct.

AI and Constructs: Economic Policy Uncertainty



AI Raises Important Construct Validity Concerns

The same cost reductions that make construct validation easier also risk moving the problem of searching across regression specifications (Sala-I-Martin, 1997) to **searching across representations of the underlying reality**.

A further concern: construct validity is **use-contingent**. A measure validated as a passive description may lose validity once it is embedded in decisions and agents re-optimize.

Because AI makes some measures **cheap to deploy in real-time at scale**, it may exacerbate this problem: the construct definition analogue of Goodhart's law.

Construct Validity as Partial Identification

How to formalize and microfound construct validity for high-dimensional, latent concepts is largely an open question.

I will briefly outline one potential approach: formalizing the **nomological network** of Cronbach and Meehl (1955) as a "partial identification" problem.

The approach is useful when researchers have reasonable priors about how the construct should behave.

What the Formalization Does—and Does Not—Do

What it does

Once the network of predictions is stated, the framework assesses which candidate measures are consistent with its implications, when direct validation is not possible.

The partially identified set is defined relative to:

- a researcher-specified **candidate menu** of measures; and
- a researcher-specified set of **moment restrictions**.

Both require justification from theory, prior research, or substantive context.

What it does not do

It does not imply that the nomological network of predictions is itself a causal account of measurement (Borsboom, Mellenbergh and van Heerden, 2004).

Candidate Measurement Functions

Let X denote high-dimensional reality. A researcher considers a **menu** of candidate measurement functions:

$$\mathcal{M} = m_j(\cdot)_{j \in \mathcal{J}},$$

where each candidate maps the same underlying reality into a candidate measure $m_j(X)$.

Both the measures and downstream estimands must be made comparable:

- candidate measures may differ in scale, sign, support, and discreteness and need to be normalized appropriately;
- the downstream estimand $\beta(m)$ should be defined consistently across candidates, e.g. an effect per standard deviation or a rank association.

Nomological Restrictions

Let V denote auxiliary observables implicated by the nomological network:

$$V = (Z, W, G, Q).$$

- $Z = (Z_1, \ldots, Z_K)$: related quantities
- $W = (W_1, \ldots, W_L)$: distinct quantities
- G: groups that should differ on the construct
- Q: covariates used for conditional or partial associations

The network implies restrictions $r = 1, \ldots, R$ on any candidate measure $m(X)$; *e.g.*,:

$$\underbrace{\mathrm{Corr}\big(m(X), Z_k \mid Q\big) \geq c_k}_{\text{convergent validity}} \qquad \underbrace{\big|\mathrm{Corr}\big(m(X), W_\ell \mid Q\big)\big| \leq d_\ell}_{\text{discriminant validity}} \qquad \underbrace{\frac{\mathbb{E}[m(X) \mid G = 1] - \mathbb{E}[m(X) \mid G = 0]}{\sigma_m} \geq \delta}_{\text{known-groups validity}}$$

The Admissible Set of Measures

Collect the thresholds into

$$\theta = (c_1, \ldots, c_K, d_1, \ldots, d_L, \delta).$$

Each nomological restriction can be written as a population moment inequality:

$$\mathbb{E}\left[g_r\big(m(X), V; \theta_r\big)\right] \geq 0, \qquad r \in \mathcal{R}.$$

Given a fixed candidate menu $\mathcal{M}$, define the admissible subset:

$$\mathcal{M}^*_{\mathcal{R},\theta} = \left\{ m \in \mathcal{M} \ : \ \mathbb{E}\left[g_r\big(m(X), V; \theta_r\big)\right] \geq 0 \quad \forall r \in \mathcal{R} \right\}.$$

The admissible set contains the candidate measures consistent with all stated restrictions and is a population object that encodes *conceptual* uncertainty.

Construct-Validity Profiles

For each candidate measure m, report the slack in each restriction:

$$s_r(m;\theta_r) = \mathbb{E}\left[g_r\big(m(X), V;\theta_r\big)\right], \qquad r \in \mathcal{R}.$$

Positive values indicate satisfied restrictions; negative values indicate violations.

Reporting the full profile helps show:

- which restrictions drive admissibility;
- how conclusions change under alternative thresholds; and
- whether conclusions are robust or fragile.

Thresholds should be grounded in prior evidence, benchmarks, expert elicitation, or substantive context.

The Construct-Identified Set

Ultimately, the researcher cares about a downstream population parameter:

$$\beta(m),$$

such as a regression coefficient, treatment effect, or structural parameter defined using the measure $m(X)$.

Given the admissible set $\mathcal{M}^*_{\mathcal{R},\theta}$, define:

$$\mathcal{B}^*_{\mathcal{R},\theta} = \left\{\beta(m) : m \in \mathcal{M}^*_{\mathcal{R},\theta}\right\}.$$

Inference for the Construct-Identified Range

For scalar targets, the goal is to estimate a confidence region for:

$$[L, U] = \left[\inf_{m \in \mathcal{M}^*_{\mathcal{R},\theta}} \beta(m), \sup_{m \in \mathcal{M}^*_{\mathcal{R},\theta}} \beta(m) \right].$$

Valid inference must account for two sources of sampling uncertainty:

- uncertainty in $\widehat{\beta}(m)$ for each candidate measure;
- uncertainty in the estimated moment inequalities that determine which measures are classified as admissible.

These sources are coupled when the measures determining the plug-in endpoints are also near the admissibility boundary.

Estimation is beyond today's scope, but existing partial-identification and endpoint-inference methods can be combined to estimate a valid, if conservative, confidence region.

How Credible Are the Auxiliary Measures?

Validation through a nomological network depends on auxiliary observables V.

But validation cannot recurse forever.

The practical requirement is more modest. Researchers should:

- state how auxiliary observables are **defined and constructed**;
- report how they are **validated**, when feasible;
- employ at least some well-accepted measures, **constructed by a different method**; (Campbell and Fiske, 1959)
- distinguish restrictions based on **well-anchored measures** from exploratory ones.

A sparse set of well-measured restrictions is more informative than a dense network of poorly justified ones.

Bibliography

Borsboom, Denny, Gideon J. Mellenbergh, and Jaap van Heerden. 2004. "The Concept of Validity." *Psychological Review*, 111(4): 1061–1071.

Campbell, Donald T, and Donald W Fiske. 1959. "Convergent and discriminant validation by the multitrait-multimethod matrix." *Psychological bulletin*, 56(2): 81.

Cronbach, Lee J, and Paul E Meehl. 1955. "Construct validity in psychological tests." *Psychological bulletin*, 52(4): 281.

Sala-I-Martin, Xavier X. 1997. "I Just Ran Two Million Regressions." *The American Economic Review*, 178–183.