

An Embarrassingly Brief Introduction to Psychological Measurement

Ben Babcock
University of Minnesota

PLEASE, don't let this (bad?) lecture be your only exposure to psychological measurement!

Enroll in Psy 5862
Psychological Measurement:
Theory and Methods

Offered every Fall semester!!!

Adventure... With Mishaps

"Sing to me of the man, Muse, the man of twists and turns
driven time and again off course, once he had plundered
the hallowed heights of Troy.

Many cities of men he saw and learned their minds,
many pains he suffered, heartsick on the open sea,
fighting to save his life and bring his comrades home."

Where Does the Story Begin?

The story begins *in medias res*. This is Latin for “_____” (Amer. Her. College Dictionary, 3rd ed.).

Jumping into statistics without discussing where the numbers came from is starting *in medias res*. We have work to do first.

What We Should Worry About First

Are the numbers any good?

Are these experimental or observational data?

Was the experiment any good?

Do we have good measures (reliability/validity)?

Do the numbers mean anything?

What scale are the numbers in?

Experimental vs. Observational Data

This distinction matters for the purposes of conclusions.

Experimental control = Easier claim for causality

Observational data \neq Strong causal claims

STATISTICAL PROCEDURES ARE
NOT MAGICAL TRUTH
GENERATING MACHINES!



Was the Experiment Any Good?

Even if your data are from an experiment, the numbers will not be any good if the experiment's design was rubbish.

Sampling procedure? Random sampling?

Random assignment?

Sound methods and measures?

Confounds and internal validity?

Good Measures: Validity and Reliability

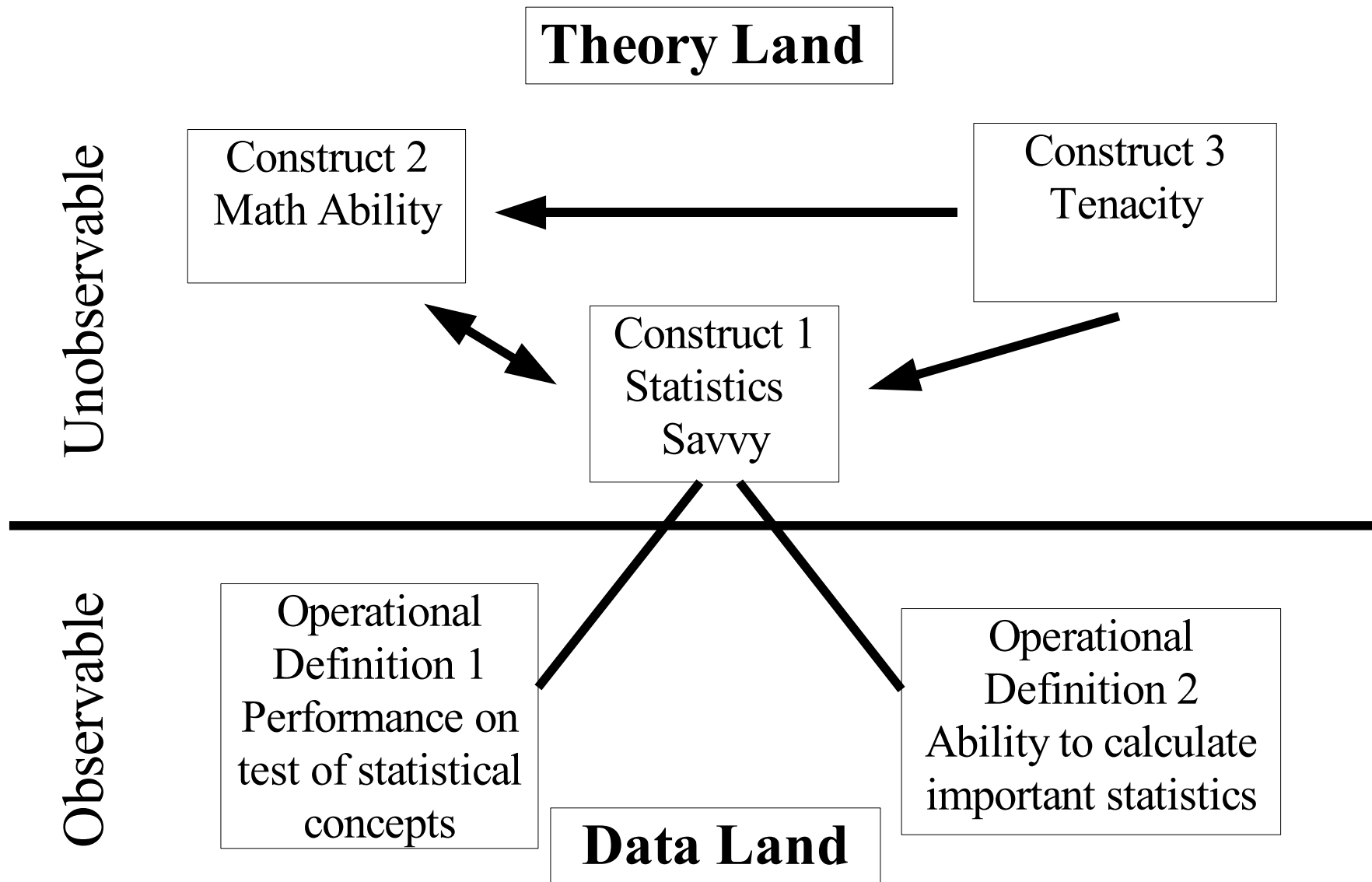
Think about measurement like a target where the bullseye represents the construct (unobservable thing) you actually want to measure.



Validity:

Reliability:

Construct Validity: What is a Construct?



Good Measures: Construct Validity

Construct validity answers the question, “does your instrument measure the construct you think it is measuring?” Do your data agree with (your) theory?

A construct is an unobservable idea or concept about how the world works. Constructs relate to one another to form _____.

An operational definition is how we attempt to tie theory down to data. We define a way to indirectly observe the construct.

Good Measures: Construct Validity

Construct Validity

Methods for CV:

- Content Validity

- Criterion-Related Validity
 - Concurrent

 - Predictive

Good Measures: Reliability

Think about a group of scores on a test. Observed scores X contain two theoretical (and unknown) parts:

“true score” T and “error” e .

$$X = T + e$$

$$\text{Observed} = \text{True} + \text{error}$$

True score: The average score for a person if the person took the test many (infinite) times.

Error: Random process that takes scores away from the true score. Error is uncorrelated with true score or other error terms.

Reliability is the proportion of variance in observed scores accounted for by _____. It is the squared correlation between _____ and _____. It ranges from 0 to 1.

Reliability

Why do we want the reliability of our scores to be high?

Let's correlate the observed and true scores to and square it to find the reliability.

The BIG PROBLEM:

Estimating Reliability: Two Samples

If you can test people two times and you have two parallel versions of your test, the reliability is simply the _____ between the two versions of the test.

$$\rho = r_{X_1 X_2}$$

where ρ is the symbol for reliability
1 and 2 are the forms of the test and
 X are the observed test scores.

Problems:

Estimating Reliability: One Sample

Method 1: Split-Half Method

1. Administer a test one time to a group of people.
2. Split the test in half by some method.
3. Correlate the scores on the two halves.
4. “Adjust” the reliability up to estimate what it might have been with a longer test* with the Spearman-Brown prophecy equation.

$$r_{1+2} = \frac{2r_{12}}{1+r_{12}}$$

where 1 and 2 indicate the halves of the test.

*Indicates:

Estimating Reliability: One Sample

Method 2: Cronbach's α

Imagine if you split the test in every way possible and took the average of the split-half reliabilities. Cronbach's α does this. The statistic is a lower bound reliability estimate.

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n V_i}{V_X} \right)$$

where n is the number of test items

V_i is the variance of a single item

and V_X is the variance of the entire test.

Don't worry about this calculation. Just know that it is a lower bound estimate of reliability.

Food for Thought

Can test scores have validity without reliability?

Can test scores have reliability without validity?

PLEASE, don't let this (atrocious, painful, lamentable?) lecture be your only exposure to psychological measurement!

Enroll in Psy 5862
Psychological Measurement:
Theory and Methods

Offered every Fall semester!!!