Structural Equation Modeling: What is a Latent Variable?

by [guest](https://www.theanalysisfactor.com/author/guest/) [6 Comments](https://www.theanalysisfactor.com/what-is-a-latent-variable/#comments)

*By Manolo Romero Escobar*

What is a latent variable?

*“The many, as we say, are seen but not known, and the ideas are known but not seen” (Plato, The Republic)*

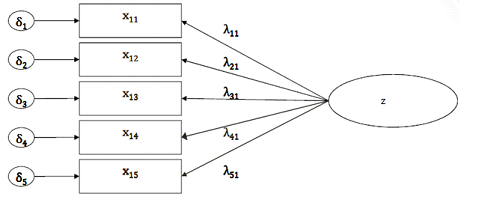
My favourite image to explain the relationship between latent and observed variables comes from the “Myth of the Cave” from Plato’s The Republic.  In this myth a group of people are constrained to face a wall.  The only things they see are shadows of objects that pass in front of a fire behind them.

As researchers we try to measure our constructs as best as we can.  Often we can see and measure the shadows of the constructs–the “objects” of our inquiry– but we can’t directly observe or measure  the constructs themselves.

So we infer these constructs,  which are unobserved, hidden, or latent, from the data we collect on related variables we can observe and directly measure.

Latent refers to the fact that even though these variables were not measured directly in the research design they are the ultimate goal of the project.

The nature of the latent variable is intrinsically related to the nature of the indicator variables used to define them. In the most usual case, we structure the model so that the indicators are “effects” of the latent variable, like in the case of the common factor analysis.



The idea is that the value of the latent variable caused people to respond as they did on the observed indicators.

On a technical note, estimation of a latent variable is done by analyzing the variance and covariance of the indicators. The measurement model of a latent variable with effect indicators is the set of relationships (modeled as equations) in which the latent variable is set as the predictor of the indicators.

This diagram could be written as a set of 5 regression models.

These relationships are not given by the data, but are modeled by the analyst/researcher based on theory and previous research. It can be understood as an extension of GLM ([see previous posts on SEM](https://www.theanalysisfactor.com/why-need-knowledge-of-multiple-regression-to-understand-sem/)) in which the predictor is a latent variable and the outcomes are the indicators.

And of course, this measurement model could be used in a much larger [SEM](https://www.theanalysisfactor.com/structural-equation-modeling-first-step-confirmatory-factor-analysis-2/) in which this latent variable z was either a predictor or outcome of other variables.

One last quote, this one from one of the founders of [factor analysis](https://www.theanalysisfactor.com/factor-analysis-1-introduction/):

*“For having executed our experiment and calculated the correlation, we must then remember that the latter does not represent the mathematical relation between the two sets of objects compared, but only between the two sets of measurements which we have derived from the former by more or less fallible processes” (Spearman, 1904)*

**References**

Bollen, K. and Lennox, R. (1991). Conventional Wisdom on Measurement: A Structural Equation Perspective. Psychological Bulletin, 110(2):305-314. doi:10.1037/0033-2909.110.2.305

Spearman, C. (1904). “General Intelligence”, objectively determined and measured. The American Journal of Psychology, 15(2), 201-292. doi:10.2307/1412107