“Ghost Chasing”:
Demystifying Latent Variables and SEM

Andrew Ainsworth
University of California, Los Angeles

Topics

- "Ghost Chasing" and Latent Variables
- What is SEM?
- SEM elements and Jargon
- Example Latent Variables
- SEM Limitations

"Ghost Chasing"

- Psychologists are in the business of Chasing "Ghosts"
  - Measuring "Ghosts"
  - "Ghost" diagnoses
  - Exchanging one "Ghost" for another "Ghost"
- Latent (AKA "Ghost") Variables
  - Anything we can't measure directly
  - We must rely on measurable indicators

What is a Latent Variable?

- An operationalization of data as an abstract construct
  - A data reduction method that uses “regression like” equations
  - Take many variables and explain them with a one or more “factors”
  - Correlated variables are grouped together and separated from other variables with low or no correlation

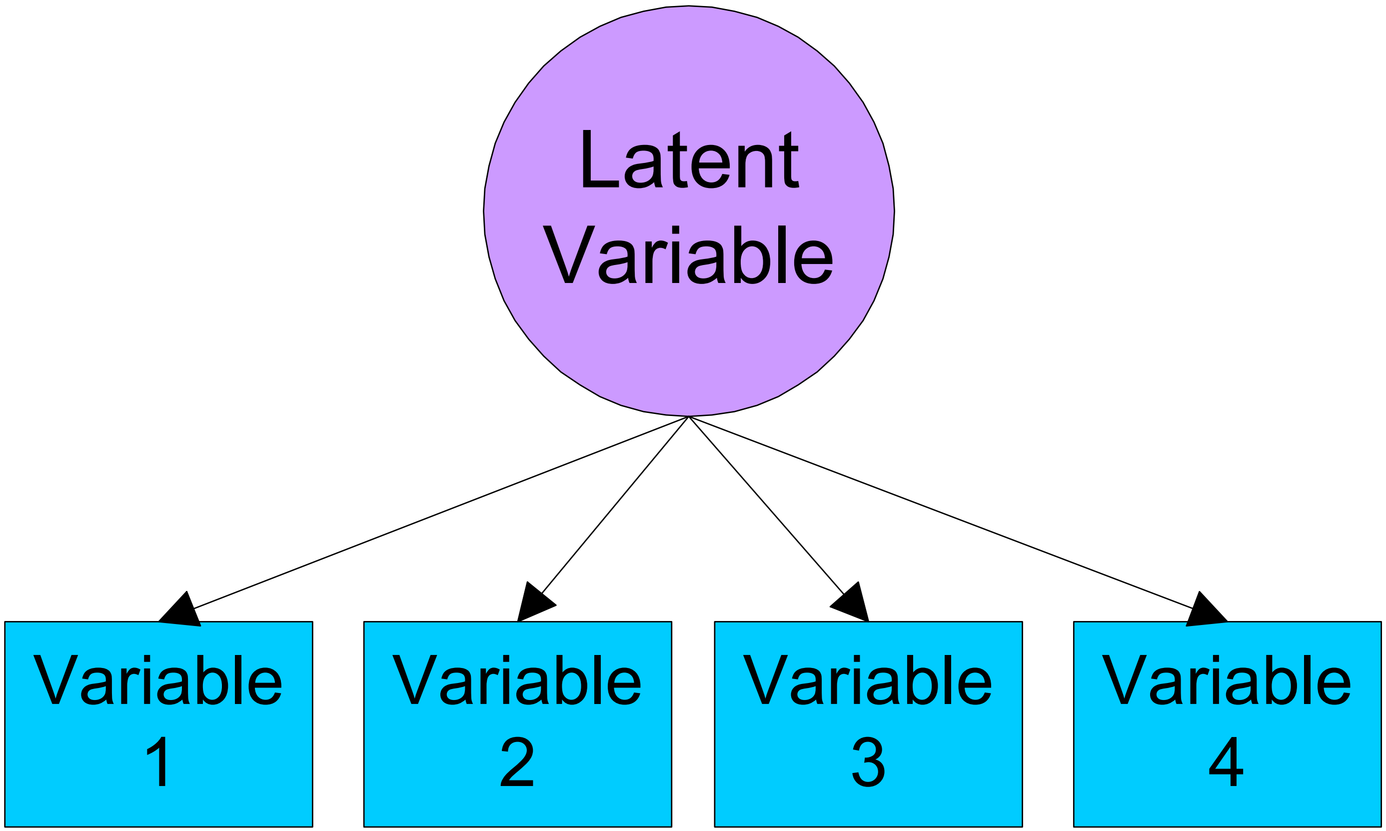
Establishing Latent Variables

- Exploratory Factor Analysis
  - Summarizing data by grouping correlated variables
  - Investigating sets of measured variables for underlying constructs
  - Often done near the onset of research and/or scale construction

Establishing Latent Variables

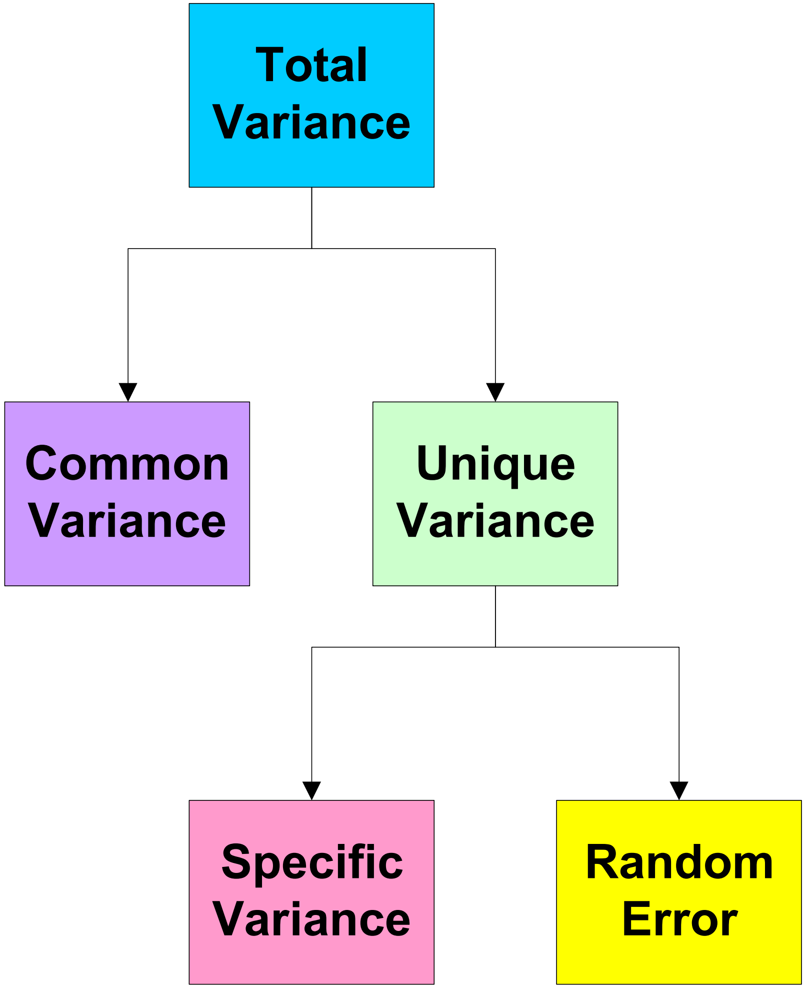
- Confirmatory Factor Analysis
  - Testing whether proposed constructs influence measured variables
  - When factor structure is known or at least theorized
  - Often done when relationships among variables are known

Conceptualizing Latent Variables



Conceptualizing Latent Variables

- Latent variables – representation of the variance shared among the variables
  - common variance without error or specific variance



What is SEM?

- SEM – Structural Equation Modeling
- Also Known As
  - CSA – Covariance Structure Analysis
  - Causal Models
  - Simultaneous Equations
  - Path Analysis
  - Confirmatory Factor Analysis
  - Latent Variable Modeling

SEM in a nutshell

- Combination of factor analysis and regression
  - Tests relationships variables
  - Specify models that explain data with few parameters
  - Flexible - Works with continuous and discrete variables
  - Significance testing and model fit

Goals in SEM

- Hypothesize a model that:
  - Has a number of parameters less than the number of unique Variance/Covariance entries (i.e. (p*(p+1))/2)
  - Has an implied covariance matrix that is not significantly different from the sample covariance matrix
  - Allows us to estimate population parameters that make the sample data the most likely

Important Matrices

- s matrix
  - Sample Covariances
  - The data
- $\sigma(\theta)$ matrix
  - Model Implied Covariances
- Residual Covariance Matrix

	Item$_1$	**Item$_2$**	**Item$_3$**	**Item$_4$**
Item$_1$	$\mathbf{s^2_{11}}$	s^2_{12}	s^2_{13}	s^2_{14}
Item$_2$	s^2_{21}	$\mathbf{s^2_{22}}$	s^2_{23}	s^2_{24}
Item$_3$	s^2_{31}	s^2_{32}	$\mathbf{s^2_{33}}$	s^2_{34}
Item$_4$	s^2_{41}	s^2_{42}	s^2_{43}	$\mathbf{s^2_{44}}$

	Item$_1$	**Item$_2$**	**Item$_3$**	**Item$_4$**
Item$_1$	σ^2_{11}	σ^2_{12}	σ^2_{13}	σ^2_{14}
Item$_2$	σ^2_{21}	σ^2_{22}	σ^2_{23}	σ^2_{24}
Item$_3$	σ^2_{31}	σ^2_{32}	σ^2_{33}	σ^2_{34}
Item$_4$	σ^2_{41}	σ^2_{42}	σ^2_{43}	σ^2_{44}

SEM Jargon

- Measurement
  - The part of the model that relates measured variables to latent factors
  - The measurement model is the factor analytic part of SEM
- Structure
  - This is the part of the model that relates variable or factors to one another (prediction)
  - If no factors are in the model then only path model exists between measured variables

SEM Jargon

- Model Specification
  - Creating a hypothesized model that you think explains the relationships among multiple variables
  - Converting the model to multiple equations
- Model Estimation
  - Technique used to calculate parameters
  - E.G. - Ordinary Least Squares (OLS), Maximum Likelihood (ML), etc.

SEM Jargon

- Model Identification
  - Rules for whether a model can be estimated
  - For example, For a single factor:
    - At least 3 indicators with non-zero loadings
    - no correlated errors
    - Fix either the Factor Variance or one of the Factor Loadings to 1

SEM Jargon

- Model Evaluation
  - Testing how well a model fits the data
  - Just like with other analyses (e.g. ANOVA) we look at squared differences
    - SEM looks at the squared difference between the s and $\sigma(\theta)$ matrices
    - While weighting the squared difference depending on the estimation method (e.g. OLS, ML, etc.)

$$\text{pick a } \sigma(\theta) \xrightarrow{\text{min}} Q = \left(s - \sigma(\theta)\right)' W \left(s - \sigma(\theta)\right)$$

SEM Jargon

- Model Evaluation
  - Even with well fitting model you need to test significance of predictors
    - Each parameter is divided by its SE to get a Z-score which can be evaluated
    - SE values are calculated as part of the estimation procedure

Conventional SEM diagrams

- □ = measured variable
- ○ = latent variable
- ⇨ = regression weight or factor loading
- ⬄ = covariance

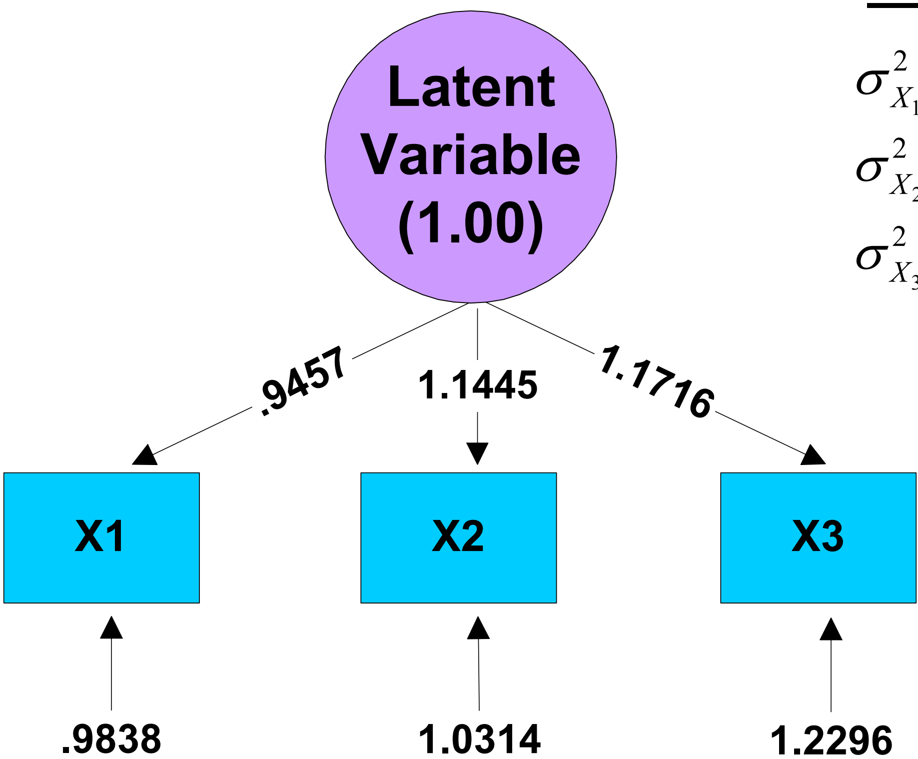
Sample Variance/Covariance Matrix

	X1	X2	X3
X1	1.8782	1.0824	1.1080
X2	1.0824	2.3414	1.3409
X3	1.1080	1.3409	2.6023

Basic Tracing Rules for a Latent Variable

- Once parameters are estimated
- Calculating the Implied Covariance Matrix
- Rules for Implied Variance
  - Common Variance – trace a path from a variable back to itself, multiplying parameters
  - Add to it the unique variance of that DV
- Rules for covariance between variables
  - Trace path from any variable to another, multiplying parameters

Implied Covariance Matrix



Variances

$$\sigma^2_{X_1} = .9457(1)(.9457) = .8944 + .9838 = 1.8782$$

$$\sigma^2_{X_2} = 1.1445(1)(1.1445) = 1.3099 + 1.0314 = 2.3413$$

$$\sigma^2_{X_3} = 1.1716(1)(1.1716) = 1.3726 + 1.2296 = 2.6022$$

Covariances

$$\sigma_{X_1X_2} = .9457(1)(1.1445) = 1.0824$$

$$\sigma_{X_1X_3} = .9457(1)(1.1716) = 1.1080$$

$$\sigma_{X_2X_3} = 1.1445(1)(1.1716) = 1.3409$$

	X1	X2	X3
X1	1.8782	1.0824	1.1080
X2	1.0824	2.3413	1.3409
X3	1.1080	1.3409	2.6022

Residual Matrix

$$\begin{vmatrix} 1.8782 & 1.0824 & 1.1080 \\ 1.0824 & 2.3414 & 1.3409 \\ 1.1080 & 1.3409 & 2.6023 \end{vmatrix}_{s} - \begin{vmatrix} 1.8782 & 1.0824 & 1.1080 \\ 1.0824 & 2.3413 & 1.3409 \\ 1.1080 & 1.3409 & 2.6022 \end{vmatrix}_{\sigma(\theta)} = \begin{vmatrix} 0 & 0 & 0 \\ 0 & .0001 & 0 \\ 0 & 0 & .0001 \end{vmatrix}_{residual}$$

Function Min and Chi-Square

$$Q = (s - \sigma(\theta))' W (s - \sigma(\theta)) =$$

$$Q = \left[\underbrace{\begin{vmatrix} 1.8782 & 1.0824 & 1.1080 \\ 1.0824 & 2.3414 & 1.3409 \\ 1.1080 & 1.3409 & 2.6023 \end{vmatrix}}_{s} - \underbrace{\begin{vmatrix} 1.8782 & 1.0824 & 1.1080 \\ 1.0824 & 2.3413 & 1.3409 \\ 1.1080 & 1.3409 & 2.6022 \end{vmatrix}}_{\sigma(\theta)} \right]' * \left[\underbrace{\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}}_{W} \right]$$

$$* \left[\underbrace{\begin{vmatrix} 1.8782 & 1.0824 & 1.1080 \\ 1.0824 & 2.3414 & 1.3409 \\ 1.1080 & 1.3409 & 2.6023 \end{vmatrix}}_{s} - \underbrace{\begin{vmatrix} 1.8782 & 1.0824 & 1.1080 \\ 1.0824 & 2.3413 & 1.3409 \\ 1.1080 & 1.3409 & 2.6022 \end{vmatrix}}_{\sigma(\theta)} \right] =$$

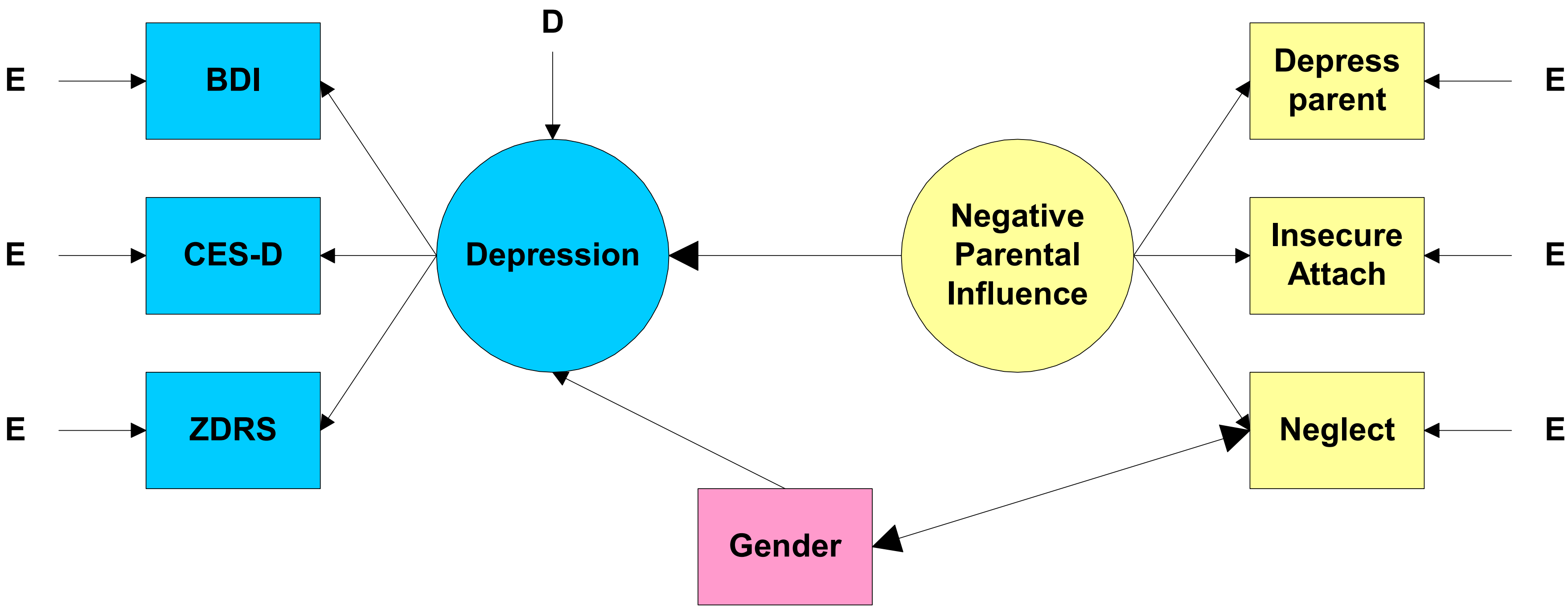
$$Q = .00000008$$

$$\chi^2(?) = Q * (492 - 1) = .00000008 * 491 = .00000982$$

$$df_{\chi^2} = \text{(\#unique VC elements)-(\# of estimated parameters)}$$

$$df_{\chi^2} = 6 - 6 = 0$$

Full Measurement Diagram



SEM limitations

- SEM is a confirmatory approach
  - You need to have established theory about the relationships
  - Exploratory methods (e.g. model modification) can be used on top of the original theory
  - SEM is not causal; experimental design = cause

SEM limitations

- SEM ⇨ correlational but, can be used with experimental data
  - Mediation and manipulation can be tested
- SEM ⇨ very fancy technique but it does not make up for a bad methods

SEM limitations

- Biggest limitation is sample size
  - It needs to be large to get stable estimates of the covariances/correlations
  - @ 200 subjects for small to medium sized model
  - A minimum of 10 subjects per estimated parameter
  - Also affected by effect size and power

Take Home Messages

- You're a "Ghost Chaser" and didn't know it
- Latent Variables are "Ghosts"
- SEM – method for getting closer to studying the "ghosts" directly
- SEM is complicated but it is accessible to you if you need to use it

Thank You!!

References

Bollen, K. (1989) *Structural Equations with Latent Variables*. New York : Wiley.

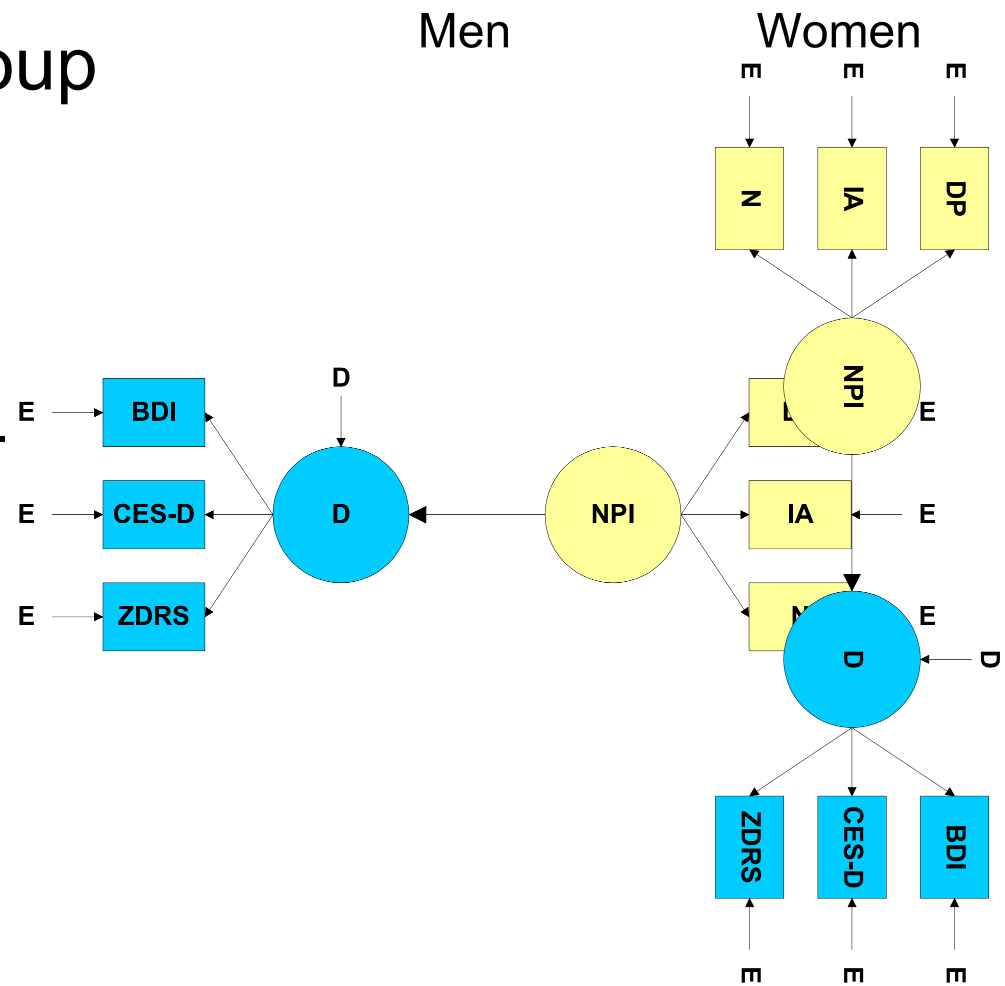
Comrey, A. L., & Lee, H. B. (1992). *A First Course in Factor Analysis* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.

Ullman, J. B. (2001). Structural Equation Modeling. In B. G. Tabachnik & L. S. Fidell (Eds.), *Using Multivariate Statistics* (4th ed.): Allyn and Bacon.

Some SEM advanced questions

- Are there group differences?
  - Multigroup models
  - e.g. Men vs. Women



Some SEM advanced questions

- Can change in responses be tracked over time?
  - Latent Growth Curve Analysis

Latent Growth Model