

SC705: Advanced Statistics
Instructor: Natasha Sarkisian
Class notes: Learning to use LISREL and PRELIS

LISREL, AMOS, EQS, and MPlus are four popular statistical packages for doing SEM. LISREL (LInear Structural RELations) popularized SEM in sociology and the social sciences and is still the package of reference in most articles about structural equation modeling, even though AMOS is becoming more popular because it makes it easier to specify models (because of its user-friendly graphical interface).

Typically, analyses using LISREL involved two types of syntax files (input files):

PRELIS syntax files, with an extension .PR2

LISREL syntax files, with an extension .LS8

Recently, LISREL also introduced the SIMPLIS language. SIMPLIS syntax is .SPL

Modern versions of LISREL also have good interactive PRELIS facilities (and you can save what you do as a PRELIS syntax). Newer versions of LISREL also allows interactive model specification. Three ways you can do that:

- LISREL Project option – what you do through pull-down menus gets recorded in LISREL language (in a .LPJ file)
- SIMPLIS Project option – what you do through pull-down menus gets recorded in LISREL language (in a .SPJ file)
- Path Diagram option – the model is specified through a combination of pull-down menus and path diagram and recorded in SIMPLIS language. The diagrams are stored in .PTH files.

We will focus primarily on learning how to write PRELIS and LISREL syntax file, even though we will examine some interactive options later on. See Byrne, “Using LISREL, PRELIS, and SIMPLIS”, pp.43-87.

Notation and Matrices:

Handout: “Using LISREL, PRELIS, and SIMPLIS”, pp.11-17 from Byrne, Barbara M., 1998, Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming. Mahwah, NJ: Erlbaum.

Summary of Matrices, Greek Notation, and Program Codes

Greek Letter	Matrix	Matrix Elements	Program Code	Matrix Type
Measurement Model				
Lambda-X	Λ_x	λ_x	LX	Regression
Lambda-Y	Λ_y	λ_y	LY	Regression
Theta delta	θ_δ	θ_δ	TD	Var/cov
Theta epsilon	θ_ϵ	θ_ϵ	TE	Var/cov
Structural Model				
Gamma	Γ	γ	GA	Regression
Beta	B	β	BE	Regression
Phi	ϕ	ϕ	PH	Var/cov
Psi	Ψ	ψ	PS	Var/cov
Xi (or Ksi)	---	ξ	---	Vector
Eta	---	η	---	Vector
Zeta	---	ζ	---	Vector

Var/cov = variance-covariance

(Table from Byrne 1998, "Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS")

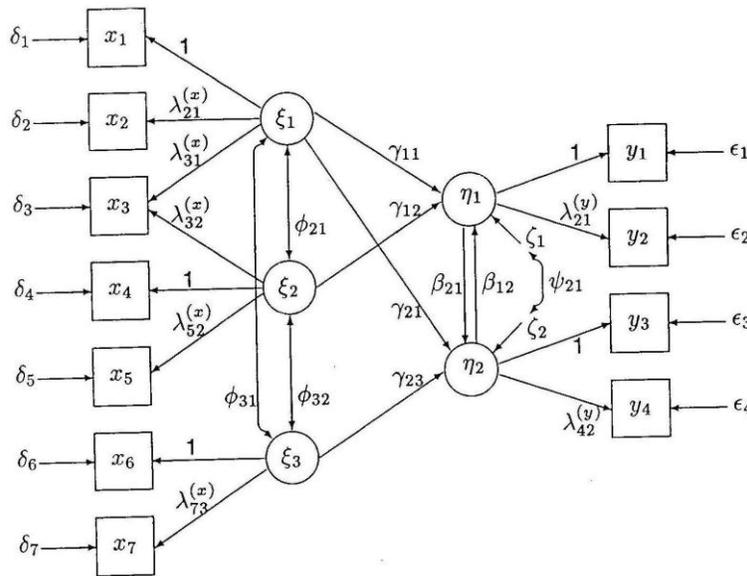


Figure 1.1 Path Diagram for Hypothetical Model

Equations and Matrices:

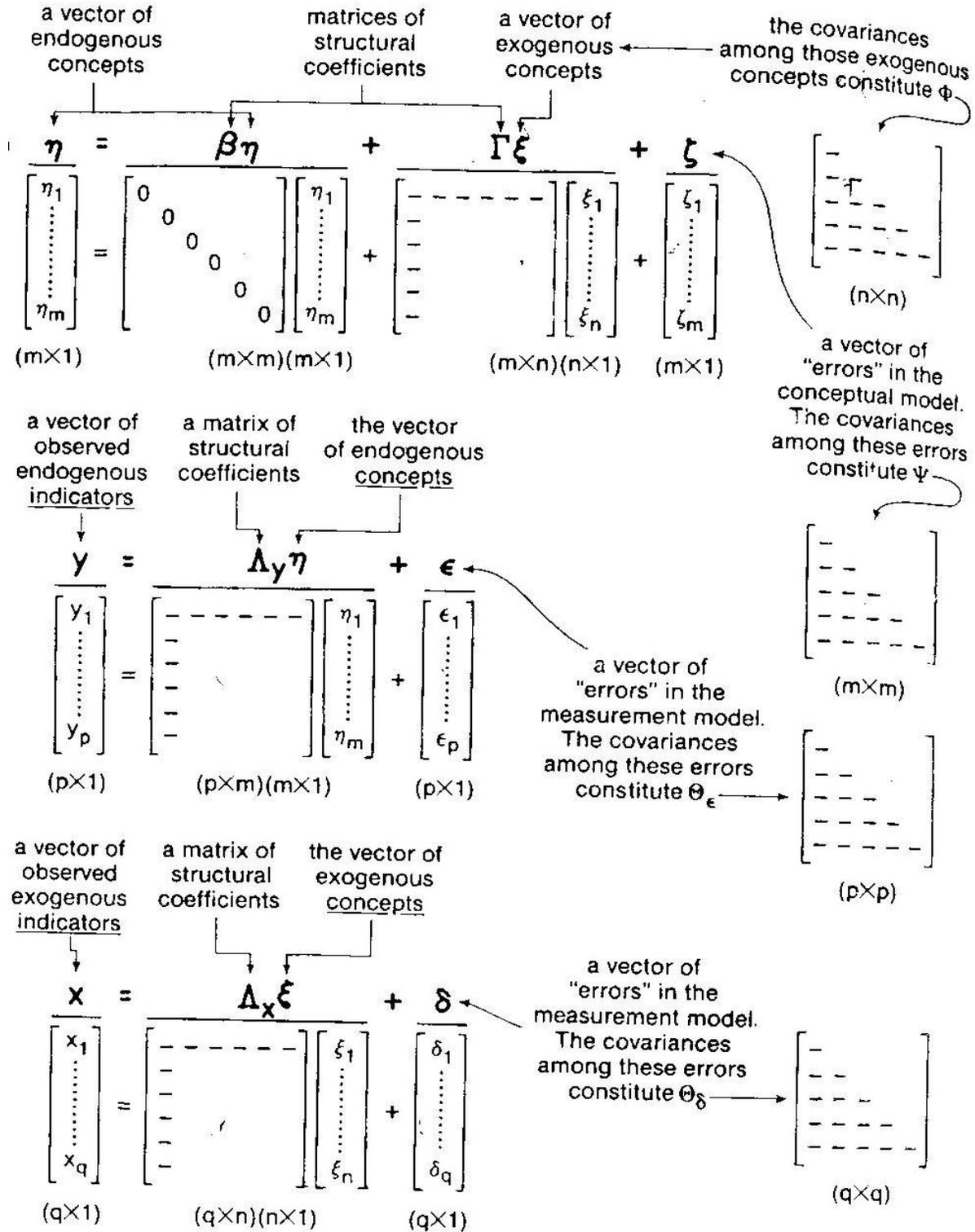
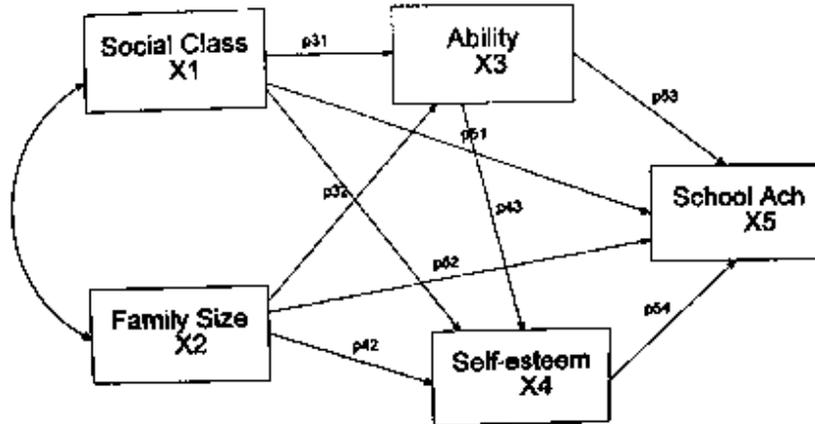


FIG. 15.1. Summary of the general structural equation model (Hayduk, 1987).

Path analysis in LISREL

We will use the following example of data (N=100) and model (from Maruyama 1998, p.57).



	X1	X2	X3	X4	X5
X1	1.00				
X2	-.33	1.00			
X3	.39	-.33	1.00		
X4	.14	-.14	.19	1.00	
X5	.43	-.28	.67	.22	1.00

The issue of model identification

We will deal with this issue throughout our discussion of SEM – it’s a very complex issue. The basic idea is that we need to consider whether we have sufficient data to find the best solution for the model we specified. There are three possible situations, and in broad terms, these are related to the amount of information we have, although this can in fact get very complicated when we consider complex models (e.g., nonrecursive models, or models for panel data) or when we have data problems (e.g., multicollinearity). In a simple situation, we want to calculate the number of data points available (the number of variances and covariances, $p*(p+1)/2$) and the number of parameters that we are estimating (these should include variance for each exogenous variable, disturbance term for each endogenous variable, as well as paths among variables).

1. Underidentified models – the number of parameters exceeds the number of data points
2. Just-identified models – the number of parameters equals the number of data points
3. Overidentified models – the number of parameters is less than the number of data points.

We always strive to have an overidentified model (with degrees of freedom > 0).

It is also useful to know that a model with no paths among variables is called the *null model*, and a model with all possible paths among variables is a *saturated model* (saturated models are just-identified).

For our example, model identification is assessed as follows: $p(p+1)/2$ – number of variances and covariances in the original matrix. Here that’s $5*6/2=15$. The number of paths on the diagram is 10, and we also have one variance per exogenous variable, and a disturbance term for the endogenous one. That adds up to 15. So the model is just-identified.

Matrix formula:

$$\eta = B\eta + \Gamma\xi + \zeta$$

Equations:

$$\begin{aligned} \eta_1 &= 0*\eta_1 + 0*\eta_2 + 0*\eta_3 + \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1 \\ \eta_2 &= \beta_{21}*\eta_1 + 0*\eta_2 + 0*\eta_3 + \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \zeta_2 \\ \eta_3 &= \beta_{31}*\eta_1 + \beta_{32}*\eta_2 + 0*\eta_3 + \gamma_{31}\xi_1 + \gamma_{32}\xi_2 + \zeta_3 \end{aligned}$$

Matrix equation:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix}$$

Other matrices involved:

Φ (2x2 matrix of variances and covariances of exogenous variables ξ) – since class and size linked with double-headed arrow, means we allow the covariance; therefore, all three elements of this matrix are estimated. Therefore, LISREL default (symmetric, free) is what we need.

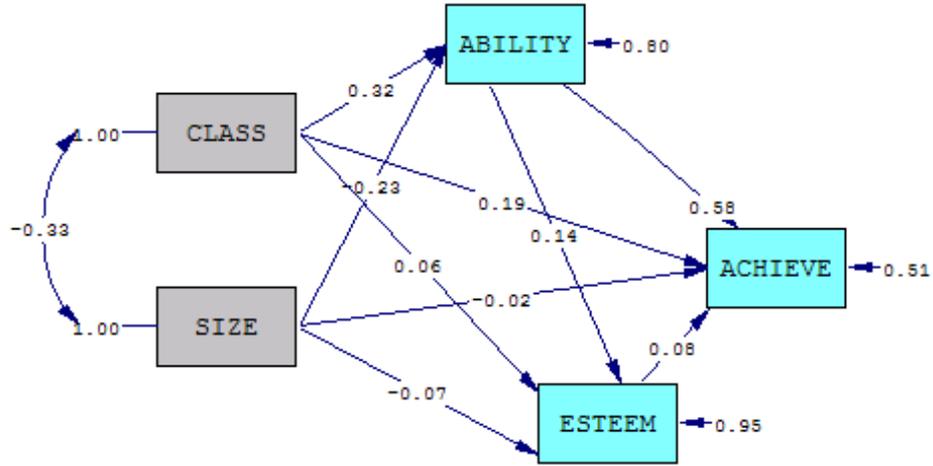
Ψ (3x3 matrix of variances and covariances of disturbance terms ζ) – since we don't allow disturbance terms to covary, the default (diagonal, free) is what we need.

In LISREL, we'll implement it with X and Y instead of η and ξ because we don't have latent variables (but other matrices are the same).

Syntax:

```
DA NI=5 NO=100 MA=KM
LA
CLASS SIZE ABILITY ESTEEM ACHIEVE
KM SY
1.00
-.33 1.00
.39 -.33 1.00
.14 -.14 .19 1.00
.43 -.28 .67 .22 1.00
SE
3 4 5 1 2
MO NX=2 NY=3 BE=FU, FI GA=FU, FR
FR BE 2 1 BE 3 1 BE 3 2
PD
OU
```

Path Diagram:



Chi-Square=0.00, df=0, P-value=1.00000, RMSEA=0.000

Output :

DA NI=5 NO=100 MA=KM

Number of Input Variables 5
 Number of Y - Variables 3
 Number of X - Variables 2
 Number of ETA - Variables 3
 Number of KSI - Variables 2
 Number of Observations 100

DA NI=5 NO=100 MA=KM

Correlation Matrix

	ABILITY	ESTEEM	ACHIEVE	CLASS	SIZE
ABILITY	1.00				
ESTEEM	0.19	1.00			
ACHIEVE	0.67	0.22	1.00		
CLASS	0.39	0.14	0.43	1.00	
SIZE	-0.33	-0.14	-0.28	-0.33	1.00

DA NI=5 NO=100 MA=KM

Parameter Specifications

BETA			
	ABILITY	ESTEEM	ACHIEVE
ABILITY	0	0	0
ESTEEM	1	0	0
ACHIEVE	2	3	0

GAMMA		
	CLASS	SIZE
ABILITY	4	5
ESTEEM	6	7
ACHIEVE	8	9

	PHI	
	CLASS	SIZE
	-----	-----
CLASS	10	
SIZE	11	12

	PSI		
	ABILITY	ESTEEM	ACHIEVE
	-----	-----	-----
	13	14	15

Note: Above, in the parameter specification section, we can see how LISREL counts the number of parameters in the structural model: it counts the path coefficients (non-fixed elements of beta and gamma matrices), but it also counts the variances and covariances for exogenous variables (elements of phi matrix) and the variances and covariances of disturbance terms (the elements of psi matrix).

DA NI=5 NO=100 MA=KM

Number of Iterations = 0

LISREL Estimates (Maximum Likelihood)

BETA

	ABILITY	ESTEEM	ACHIEVE
	-----	-----	-----
ABILITY	- -	- -	- -
ESTEEM	0.14 (0.11) 1.29	- -	- -
ACHIEVE	0.58 (0.08) 7.04	0.08 (0.07) 1.10	- -

GAMMA

	CLASS	SIZE
	-----	-----
ABILITY	0.32 (0.10) 3.27	-0.23 (0.10) -2.34
ESTEEM	0.06 (0.11) 0.55	-0.07 (0.11) -0.68
ACHIEVE	0.19 (0.08) 2.33	-0.02 (0.08) -0.21

Covariance Matrix of Y and X

	ABILITY	ESTEEM	ACHIEVE	CLASS	SIZE
	-----	-----	-----	-----	-----
ABILITY	1.00				
ESTEEM	0.19	1.00			
ACHIEVE	0.67	0.22	1.00		
CLASS	0.39	0.14	0.43	1.00	
SIZE	-0.33	-0.14	-0.28	-0.33	1.00

PHI

	CLASS	SIZE
CLASS	1.00 (0.14) 6.96	
SIZE	-0.33 (0.11) -3.09	1.00 (0.14) 6.96

PSI

Note: This matrix is diagonal.

ABILITY	ESTEEM	ACHIEVE
0.80 (0.12) 6.96	0.95 (0.14) 6.96	0.51 (0.07) 6.96

Squared Multiple Correlations for Structural Equations

ABILITY	ESTEEM	ACHIEVE
0.20	0.05	0.49

Squared Multiple Correlations for Reduced Form

ABILITY	ESTEEM	ACHIEVE
0.20	0.03	0.21

Reduced Form

	CLASS	SIZE
ABILITY	0.32 (0.10) 3.27	-0.23 (0.10) -2.34
ESTEEM	0.11 (0.11) 0.99	-0.11 (0.11) -0.99
ACHIEVE	0.38 (0.10) 3.95	-0.15 (0.10) -1.62

Goodness of Fit Statistics

Degrees of Freedom = 0
 Minimum Fit Function Chi-Square = 0.0 (P = 1.00)
 Normal Theory Weighted Least Squares Chi-Square = 0.00 (P = 1.00)

The Model is Saturated, the Fit is Perfect !

Measurement Model

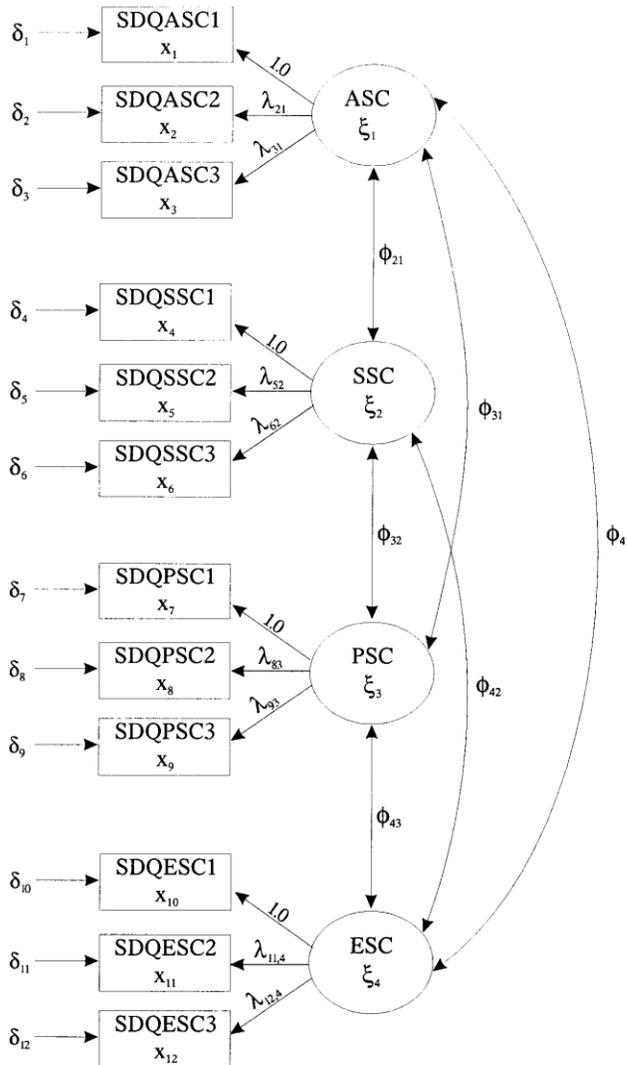
Simple path analysis ignores the possibility of measurement error – it assumes that each variable is measured perfectly. Measurement errors are less problematic for the endogenous variables – they become incorporated into the disturbance terms, so they don't affect the actual regression coefficients, although they do affect the proportion of variance explained (they would, however, affect the standardized regression coefficients because the total variance is affected).

The errors of measurement for exogenous variables will affect the regression coefficients, however, and therefore they are more problematic. Sometimes it is possible to incorporate measurement error based on known reliability for the measure, but it is also problematic if we are not very sure about that reliability estimate. One can do sensitivity analyses to see how various estimates of reliability affect the structural model results. But a better way to deal with measurement error is to have multiple indicators and to specify a measurement model, so for now, we'll focus on that.

While the structural model (path analysis portion of SEM) is based on regression, the measurement model is based on Confirmatory Factor Analysis (CFA). Note that there are some major differences between CFA and typical Exploratory Factor Analysis that many of you might be familiar with:

- EFA is atheoretical, CFA is based on theory
- In EFA, all indicators are related to all latent variables, only the strength of these correlations differs. In CFA, only some indicators are related to each of the latent variables; typically they do not overlap (i.e. each indicator is linked to only one latent variable, although there are exceptions).
- Related to the previous point, EFA models are always underidentified and therefore multiple solutions are possible; all are equally good, and the best solution is usually selected on the basis of producing a desirable structure of loadings (i.e. that each indicator has high loading for only one latent variable, and only weak loadings for the other variables). CFA models, in contrast, should be just-identified or overidentified.
- In EFA, the latent variables (factors) are usually assumed uncorrelated with each other (so called Principal Components Analysis). CFA, in contrast, is based on common factor analysis, and the factors are not considered orthogonal – they are, using the factor analysis terminology, oblique.
- Another difference between the PCA used in EFA and the common factor analysis used in CFA is in the utilization of variances and covariances – the PCA models redistribute all variance in the data across factor loadings, while the common factor analysis models partition the variance into common variance and residual variance.

When estimating a measurement model, we first need to specify the model based on theory (i.e. specify which indicators measure which latent variables).



(Diagram from: Byrne 1998, p.27)

Note that the latent variables are all connected with double-ended arrows: CFA model typically allows all latent variables to covary.

We also need to decide on the reference indicator – i.e. one path per indicator should be selected as a reference indicator and set to 1 to identify the scale for the latent variable. Alternatively, we could allow all paths to be estimated freely but set the variance of each latent variable to 1 (i.e., we can either have a latent variable that is measured in units of one of the indicators, or we can standardize it).

The issue of identification for the measurement model is similar to that for the path model – we need to count the number of variances and covariances $p*(p+1)/2$, and we need to compare that to the number of estimated parameters.

Measurement model using LISREL

Let's estimate a measurement model with two latent variables, academic ability measured by two test scores, X1 and X2, and peer popularity, measured by choices of seating, choices during schoolwork, and playground choices (X3, X4, and X5). The number of cases N=100. Here's the correlation matrix:

	X1	X2	X3	X4	X5
X1	1.00				
X2	.28	1.00			
X3	.16	.10	1.00		
X4	.03	.04	.52	1.00	
X5	.15	.05	.59	.36	1.00

Formulas and equations:

$$X = \Lambda_x \xi + \delta$$

$$x_1 = 1*\xi_1 + 0*\xi_2 + \delta_1$$

$$x_2 = \lambda_{21}*\xi_1 + 0*\xi_2 + \delta_2$$

$$x_3 = 0*\xi_1 + 1*\xi_2 + \delta_3$$

$$x_4 = 0*\xi_1 + \lambda_{42}*\xi_2 + \delta_4$$

$$x_5 = 0*\xi_1 + \lambda_{52}*\xi_2 + \delta_5$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ 0 & 1 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix}$$

Other matrices:

Θ_δ (5x5 matrix of variances and covariances of measurement errors δ) – measurement errors vary but they do not covary. Therefore, we want to have this matrix to be diagonal and free – so LISREL default is what we need.

Φ (2x2 matrix of variances and covariances of exogenous variables ξ) – in a pure measurement model, we allow all latent variables to covary (all have to be connected by double-headed arrows). Therefore, all three elements of this matrix are estimated -- LISREL default (symmetric, free) is what we need.

DA NI=5 NO=100 MA=KM

LA

SCORE1 SCORE2 SEAT SCHOOL PLAY

KM SY

1.00
 .28 1.00
 .16 .10 1.00
 .03 .04 .52 1.00
 .15 .05 .59 .36 1.00

MO NX=5 NK=2 LX=FU, FI

LK

ABILITY PEER

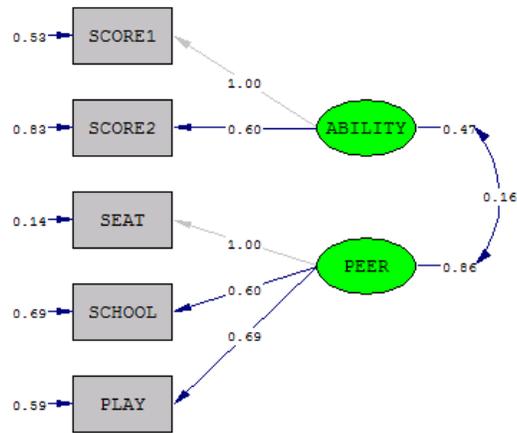
FR LX 2 1 LX 4 2 LX 5 2

VA 1.0 LX 1 1 LX 3 2

PD

OU

Output:



Chi-Square=1.08, df=4, P-value=0.89774, RMSEA=0.000

DA NI=5 NO=100 MA=KM

Number of Input Variables 5
 Number of Y - Variables 0
 Number of X - Variables 5
 Number of ETA - Variables 0
 Number of KSI - Variables 2
 Number of Observations 100

DA NI=5 NO=100 MA=KM

Correlation Matrix

	SCORE1	SCORE2	SEAT	SCHOOL	PLAY
SCORE1	1.00				
SCORE2	0.28	1.00			
SEAT	0.16	0.10	1.00		
SCHOOL	0.03	0.04	0.52	1.00	
PLAY	0.15	0.05	0.59	0.36	1.00

DA NI=5 NO=100 MA=KM

Parameter Specifications

LAMBDA-X

	ABILITY	PEER
	-----	-----
SCORE1	0	0
SCORE2	1	0
SEAT	0	0
SCHOOL	0	2
PLAY	0	3

PHI

	ABILITY	PEER			
	-----	-----			
ABILITY	4				
PEER	5	6			
THETA-DELTA					
	SCORE1	SCORE2	SEAT	SCHOOL	PLAY
	-----	-----	-----	-----	-----
	7	8	9	10	11

DA NI=5 NO=100 MA=KM

Number of Iterations = 10

LISREL Estimates (Maximum Likelihood)

LAMBDA-X

	ABILITY	PEER
	-----	-----
SCORE1	1.00	- -
SCORE2	0.60 (0.63) 0.95	- -
SEAT	- -	1.00
SCHOOL	- -	0.60 (0.14) 4.24
PLAY	- -	0.69 (0.15) 4.52

PHI

	ABILITY	PEER
	-----	-----
ABILITY	0.47 (0.50) 0.92	
PEER	0.16 (0.10) 1.57	0.86 (0.21) 4.03

THETA-DELTA

SCORE1	SCORE2	SEAT	SCHOOL	PLAY
0.53	0.83	0.14	0.69	0.59
(0.50)	(0.21)	(0.16)	(0.11)	(0.11)
1.08	3.90	0.85	6.02	5.23

Squared Multiple Correlations for X - Variables

SCORE1	SCORE2	SEAT	SCHOOL	PLAY
0.47	0.17	0.86	0.31	0.41

Goodness of Fit Statistics

Degrees of Freedom = 4
 Minimum Fit Function Chi-Square = 1.09 (P = 0.90)
 Normal Theory Weighted Least Squares Chi-Square = 1.08 (P = 0.90)
 Estimated Non-centrality Parameter (NCP) = 0.0
 90 Percent Confidence Interval for NCP = (0.0 ; 1.72)

Minimum Fit Function Value = 0.011
 Population Discrepancy Function Value (F0) = 0.0
 90 Percent Confidence Interval for F0 = (0.0 ; 0.017)
 Root Mean Square Error of Approximation (RMSEA) = 0.0
 90 Percent Confidence Interval for RMSEA = (0.0 ; 0.066)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.93
 Expected Cross-Validation Index (ECVI) = 0.26
 90 Percent Confidence Interval for ECVI = (0.26 ; 0.28)
 ECVI for Saturated Model = 0.30
 ECVI for Independence Model = 0.99

Chi-Square for Independence Model with 10 Degrees of Freedom = 88.07
 Independence AIC = 98.07
 Model AIC = 23.08
 Saturated AIC = 30.00
 Independence CAIC = 116.10
 Model CAIC = 62.73
 Saturated CAIC = 84.08

Normed Fit Index (NFI) = 0.99
 Non-Normed Fit Index (NNFI) = 1.09
 Parsimony Normed Fit Index (PNFI) = 0.40
 Comparative Fit Index (CFI) = 1.00
 Incremental Fit Index (IFI) = 1.03
 Relative Fit Index (RFI) = 0.97

Critical N (CN) = 1208.77

Root Mean Square Residual (RMR) = 0.021
 Standardized RMR = 0.021
 Goodness of Fit Index (GFI) = 1.00
 Adjusted Goodness of Fit Index (AGFI) = 0.98
 Parsimony Goodness of Fit Index (PGFI) = 0.27