Mplus Short Courses
Topic 1

Exploratory Factor Analysis, Confirmatory Factor Analysis, And Structural Equation Modeling For Continuous Outcomes

Linda K. Muthén
Bengt Muthén

www.statmodel.com
08/06/2009

Table Of Contents

Table Of Contents

Mplus Background

- Inefficient dissemination of statistical methods:
  - Many good methods contributions from biostatistics, psychometrics, etc are underutilized in practice
- Fragmented presentation of methods:
  - Technical descriptions in many different journals
  - Many different pieces of limited software
- Mplus: Integration of methods in one framework
  - Easy to use: Simple, non-technical language, graphics
  - Powerful: General modeling capabilities
- Mplus versions
  - V1: November 1998
  - V2: February 2001
  - V3: March 2004
  - V4: February 2006
  - V5: November 2007
  - V5.2: November 2008
- Mplus team: Linda & Bengt Muthén, Thuy Nguyen, Tihomir Asparouhov, Michelle Conn, Jean Maninger

Statistical Analysis With Latent Variables A General Modeling Framework

Statistical Concepts Captured By Latent Variables

Continuous Latent Variables

- Measurement errors
- Factors
- Random effects
- Frailties, liabilities
- Variance components
- Missing data

Categorical Latent Variables

- Latent classes
- Clusters
- Finite mixtures
- Missing data

Statistical Analysis With Latent Variables A General Modeling Framework (Continued)

Models That Use Latent Variables

Continuous Latent Variables

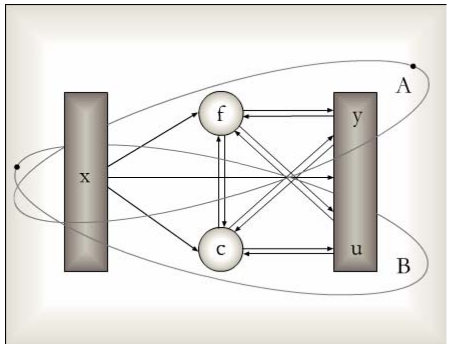
- Factor analysis models
- Structural equation models
- Growth curve models
- Multilevel models

Categorical Latent Variables

- Latent class models
- Mixture models
- Discrete-time survival models
- Missing data models

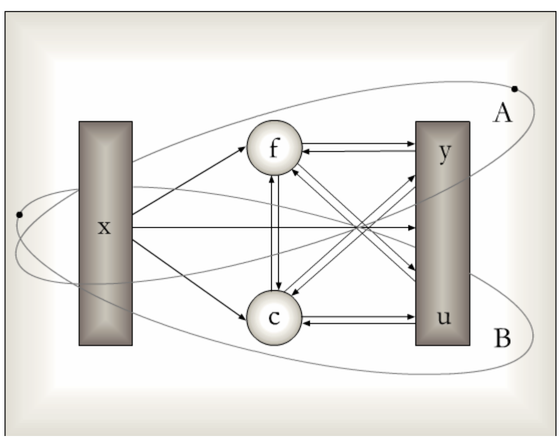
Mplus integrates the statistical concepts captured by latent variables into a general modeling framework that includes not only all of the models listed above but also combinations and extensions of these models.

General Latent Variable Modeling Framework

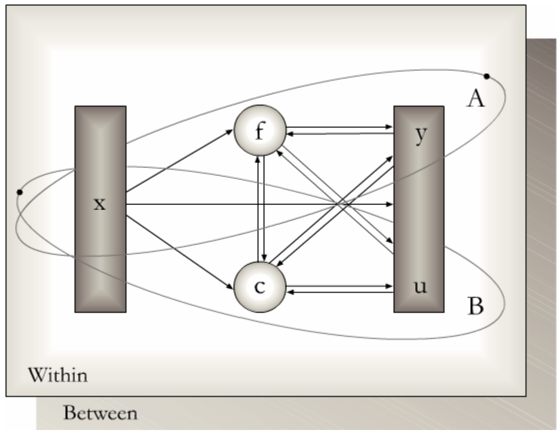


- Observed variables
  - x background variables (no model structure)
  - y continuous and censored outcome variables
  - u categorical (dichotomous, ordinal, nominal) and count outcome variables
- Latent variables
  - f continuous variables
    - – interactions among f's
  - c categorical variables
    - – multiple c's

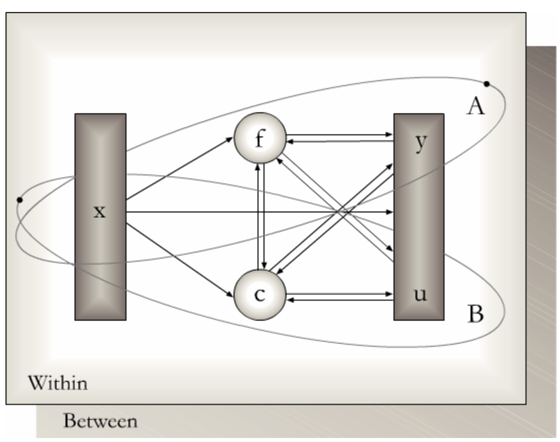
General Latent Variable Modeling Framework



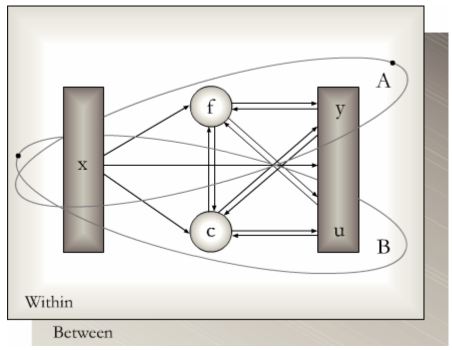
General Latent Variable Modeling Framework



General Latent Variable Modeling Framework



General Latent Variable Modeling Framework



- Observed variables
  - x	background variables (no model structure)
  - y	continuous and censored outcome variables
  - u	categorical (dichotomous, ordinal, nominal) and count outcome variables
- Latent variables
  - f	continuous variables
    – interactions among f's
  - c	categorical variables
    – multiple c's

Mplus

Several programs in one

- Exploratory factor analysis
- Structural equation modeling
- Item response theory analysis
- Latent class analysis
- Latent transition analysis
- Survival analysis
- Growth modeling
- Multilevel analysis
- Complex survey data analysis
- Monte Carlo simulation

Fully integrated in the general latent variable framework

Overview Of Mplus Courses

- **Topic 1.** August 20, 2009, Johns Hopkins University: Introductory - advanced factor analysis and structural equation modeling with continuous outcomes
- **Topic 2.** August 21, 2009, Johns Hopkins University: Introductory - advanced regression analysis, IRT, factor analysis and structural equation modeling with categorical, censored, and count outcomes

- **Topic 3.** March, 2010, Johns Hopkins University: Introductory and intermediate growth modeling
- **Topic 4.** March, 2010, Johns Hopkins University: Advanced growth modeling, survival analysis, and missing data analysis

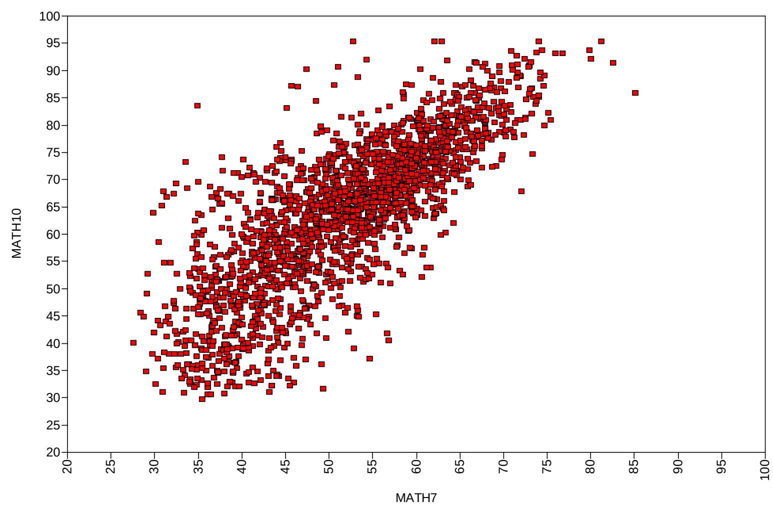
Overview Of Mplus Courses (Continued)

- **Topic 5.** August, 2010, Johns Hopkins University: Categorical latent variable modeling with cross-sectional data
- **Topic 6.** August 2010, Johns Hopkins University: Categorical latent variable modeling with longitudinal data

- **Topic 7.** March, 2011, Johns Hopkins University: Multilevel modeling of cross-sectional data
- **Topic 8.** March 2011, Johns Hopkins University: Multilevel modeling of longitudinal data

Regression Analysis

LSAY Math Regression



Regression Analysis

Regression model:

$$y_i = \alpha + \beta\, x_i + \varepsilon_i\,, \qquad (1)$$

$$E(\varepsilon_i \,|\, x_i) = E(\varepsilon_i) = E(\varepsilon) = 0 \text{ (}x\text{ and }\varepsilon\text{ uncorrelated)}, \qquad (2)$$

$$V(\varepsilon_i \,|\, x_i) = V(\varepsilon_i) = V(\varepsilon) \text{ (constant variance)}. \qquad (3)$$

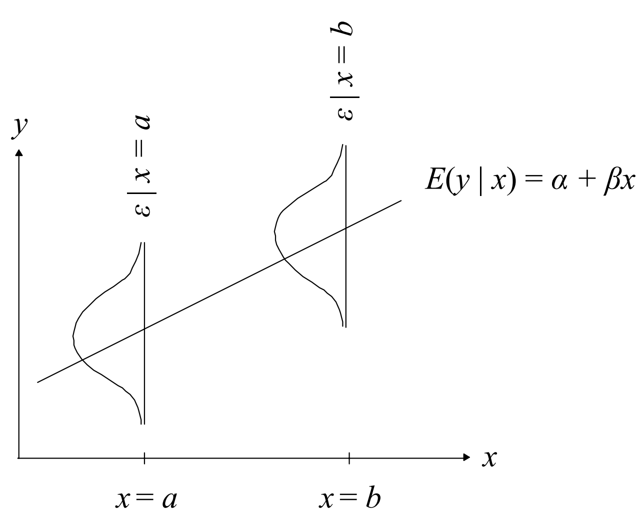
For inference and ML estimation, we also assume ε normal.

The model implies

$E(y \,|\, x) = \alpha + \beta\, x$ (conditional expectation function)

$V(y \,|\, x) = V(\varepsilon)$ (homoscedasticity)

Regression Analysis (Continued)



Regression Analysis (Continued)

Population formulas:

$$y_i = \alpha + \beta\, x_i + \varepsilon_i\,, \tag{1}$$

$$E(y) = E(\alpha) + E(\beta\, x) + E(\varepsilon)$$

$$= \alpha + \beta\, E(x) \tag{2}$$

$$V(y) = V(\alpha) + V(\beta\, x) + V(\varepsilon)$$

$$= \beta^2\, V(x) + V(\varepsilon) \tag{3}$$

$$Cov(y, x) = E[y - E(y)]\,[x - E(x)] = \beta\, V(x) \tag{4}$$

$$R^2 = \beta^2\, V(x) \,/\, (\beta^2\, V(x) + V(\varepsilon)) \tag{5}$$

$$Stdyx\ \beta = \beta\,\frac{SD(x)}{SD(y)} \tag{6}$$

Regression Analysis (Continued)

The model has 3 parameters: α, β, and $V(\varepsilon)$
Note: $E(x)$ and $V(x)$ are not model parameters

Formulas for ML and OLS parameter estimates based on a random sample

$$\hat{\beta} = s_{yx} \,/\, s_{xx}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\, \bar{x}$$

$$\hat{V}(\varepsilon) = s_{yy} - \hat{\beta}^2 s_{xx}$$

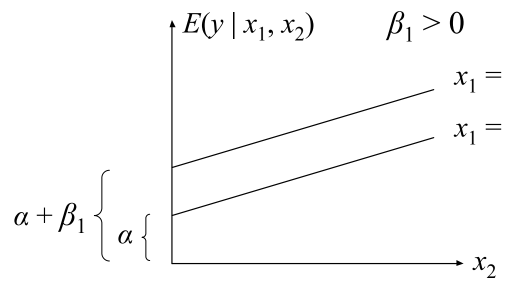
Prediction

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}\, x_i$$

Regression Analysis (Continued)

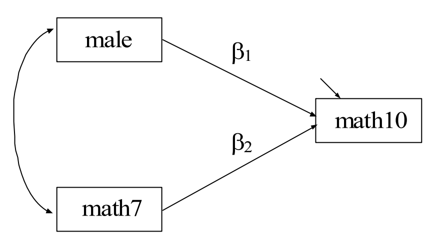
x_1 0/1 dummy variable (e.g. gender), x_2 continuous variable

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$
$$E(y \mid x_1 = 0, x_2) = \alpha + \beta_2 x_2$$
$$E(y \mid x_1 = 1, x_2) = \underbrace{\alpha + \beta_1}_{\text{intercept}} + \beta_2 x_2$$



Analogous to ANCOVA

Regression Of LSAY Math10 On Gender And Math7



Parameter estimates are produced for the intercept, the two slopes, and the residual variance.

Note: Means, variances and covariance for male and math7 are not part of the model

Input For Regression Of Math10 On Gender And Math7

```
TITLE:      Regressing math10 on math7 and gender
DATA:       FILE = dropout.dat;
            FORMAT = 11f8 6f8.2 1f8 2f8.2 10f2;
VARIABLE:   NAMES ARE id school gender mothed fathed fathsei ethnic
            expect pacpush pmpush homeres math7 math8 math9 math10
            math11 math12 problem esteem mathatt clocatn dlocatn
            elocatn flocatn glocatn hlocatn ilocatn jlocatn
            klocatn llocatn;
            MISSING = mothed (8) fathed (8)  fathsei (996 998)
                 ethnic (8) homeres (98) math7-math12 (996 998);
            USEVAR = math7 math10 male;
DEFINE:     male = gender - 1; ! male is a 0/1 variable created from
                          ! gender = 1/2 where 2 is male
MODEL:      math10 ON male math7;
OUTPUT:     TECH1 SAMPSTAT STANDARDIZED;
PLOT:       TYPE = PLOT1;
```


Output Excerpts For Regression Of Math10 On Gender And Math7

Estimated Sample Statistics

Means

	MATH10	MATH7	MALE
1	62.423	50.378	0.522

Covariances

	MATH10	MATH7	MALE
MATH10	186.926		
MATH7	109.826	103.950	
MALE	-0.163	-0.334	0.250

Correlations

	MATH10	MATH7	MALE
MATH10	1.000		
MATH7	0.788	1.000	
MALE	-0.024	-0.066	1.000

Output Excerpts For Regression Of Math10 On Gender And Math7 (Continued)

Model Results

```
                   Estimates     S.E.  Est./S.E.      Std    StdYX

MATH10 ON
               MALE    0.763    0.374     2.037    0.763    0.028
              MATH7    1.059    0.018    57.524    1.059    0.790

Intercepts
             MATH10    8.675    0.994     8.726    8.675    0.635

Residual Variances
             MATH10   70.747    2.225    31.801   70.747    0.378

R-SQUARE

  Observed Variable   R-Square

             MATH10    0.622
```


Further Readings On Regression Analysis

Agresti, A. & Finlay, B. (1997). Statistical methods for the social sciences. Third edition. New Jersey: Prentice Hall.

Amemiya, T. (1985). Advanced econometrics. Cambridge, Mass.: Harvard University Press.

Hamilton, L.C. (1992). Regression with graphics. Belmont, CA: Wadsworth.

Johnston, J. (1984). Econometric methods. Third edition. New York: McGraw-Hill.

Lewis-Beck, M. S. (1980). Applied regression: An introduction. Newbury Park, CA: Sage Publications.

Moore, D.S. & McCabe, G.P. (1999). Introduction to the practice of statistics. Third edition. New York: W.H. Freeman and Company.

Pedhazur, E.J. (1997). Multiple regression in behavioral research. Third Edition. New York: Harcourt Brace College Publishers.

Path Analysis

Path Analysis

Used to study relationships among a set of observed variables

- Estimate and test direct and indirect effects in a system of regression equations
- Estimate and test theories about the absence of relationships

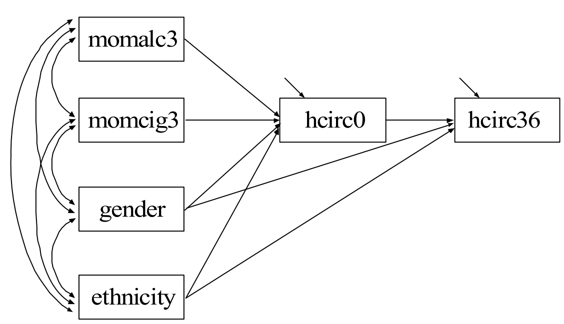
Maternal Health Project (MHP) Data

The data are taken from the Maternal Health Project (MHP). The subjects were a sample of mothers who drank at least three drinks a week during their first trimester plus a random sample of mothers who used alcohol less often.

Mothers were measured at the fourth and seventh month of pregnancy, at delivery, and at 8, 18, and 36 months postpartum. Offspring were measured at 0, 8, 18 and 36 months.

Variables for the mothers included: demographic, lifestyle, current environment, medical history, maternal psychological status, alcohol use, tobacco use, marijuana use, and other illicit drug use. Variables for the offspring included: head circumference, height, weight, gestational age, gender, and ethnicity.

Data for the analysis include mother's alcohol and cigarette use in the third trimester and the child's gender, ethnicity, and head circumference both at birth and at 36 months.



Input For Maternal Health Project Path Analysis

```
TITLE:      Maternal health project path analysis

DATA:       FILE IS headalln.dat;
            FORMAT IS 1f8.2 47f7.2;

VARIABLE:   NAMES ARE id weight0 weight8 weight18 weigh36
            height0 height8 height18 height36 hcirc0 hcirc8
            hcirc18 hcirc36 momalc1 momalc2 momalc3 momalc8
            momalc18 momalc36 momcig1 momcig2 momcig3 momcig8
            momcig18 momcig36 gender eth momht gestage age8
            age18 age36 esteem8 esteem18 esteem36 faminc0
            faminc8 faminc18 faminc36 momdrg36 gravid sick8
            sick18 sick36 advp advm1 advm2 advm3;

            MISSING = ALL (999);

            USEV = momalc3 momcig3 hcirc0 hcirc36 gender eth;

            USEOBS = id NE 1121 AND NOT (momalc1 EQ 999 AND
                     momalc2 EQ 999 AND momalc3 EQ 999);
```


Input For Maternal Health Project Path Analysis (Continued)

```
DEFINE:     hcirc0  = hcirc1/10;
            hcirc36 = hcirc36/10;
            momalc3 = log(momalc3 +1);

MODEL:      hcirc36 ON hcirc0 gender eth;
            hcirc0 ON momalc3 momcig3 gender eth;

OUTPUT:     SAMPSTAT STANDARDIZED;
```

Output Excerpts Maternal Health Project Path Analysis

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                                1.781
             Degrees of Freedom                       2
             P-Value                              .4068

RMSEA (Root Mean Square Error Of Approximation)
             Estimate                              .000
             90 Percent C.I.                       .000  0.079
             Probability RMSEA <= .05              .774
```


Output Excerpts Maternal Health Project Path Analysis (Continued)

Model Results

```
                    Estimates      S.E.  Est./S.E.     Std     StdYX

  HCIRC36   ON
     HCIRC0            .415       .036    11.382      .415      .439
     GENDER            .762       .107     7.146      .762      .270
     ETH              -.094       .107     -.879     -.094     -.033

  HCIRC0   ON
     MOMALC3          -.500       .239    -2.090     -.500     -.084
     MOMCIG3          -.013       .005    -2.604     -.013     -.108
     GENDER            .495       .118     4.185      .495      .166
     ETH               .578       .125     4.625      .578      .194
```

Output Excerpts Maternal Health Project Path Analysis (Continued)

```
Residual Variances
   HCIRC0           2.043      .119    17.107    2.043      .920
   HCIRC36          1.385      .087    15.844    1.385      .697

Intercepts
   HCIRC0          33.729      .112   301.357   33.729    22.629
   HCIRC36         35.338     1.227    28.791   35.338    25.069
```

R-Square

```
Observed
Variable         R-Square

HCIRC0               .080
HCIRC36              .303
```


The MODEL INDIRECT Command

MODEL INDIRECT is used to request indirect effects and their standard errors. Delta method standard errors are computed as the default.

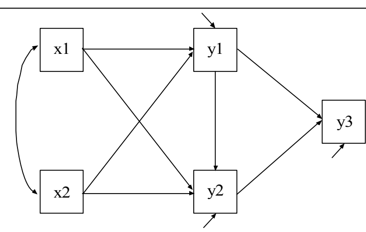
The BOOTSTRAP option of the ANALYSIS command can be used to obtain bootstrap standard errors for the indirect effects.

The STANDARDIZED option of the OUTPUT command can be used to obtain standardized indirect effects.

The MODEL INDIRECT Command (Continued)

The CINTERVAL option of the OUTPUT command can be used to obtain confidence intervals for the indirect effects and the standardized indirect effects. Three types of 95% and 99% confidence intervals can be obtained: symmetric, bootstrap, or bias-corrected bootstrap confidence intervals. The bootstrapped distribution of each parameter estimate is used to determine the bootstrap and bias-corrected bootstrap confidence intervals. These intervals take non-normality of the parameter estimate distribution into account. As a result, they are not necessarily symmetric around the parameter estimate.

The MODEL INDIRECT Command (Continued)



MODEL INDIRECT has two options:

- IND – used to request a specific indirect effect or a set of indirect effects
- VIA – used to request a set of indirect effects that includes specific mediators

MODEL INDIRECT

```
    y3 IND y1 x1;       !x1 -> y1 -> y3
    y3 IND y2 x2;       !x2 -> y2 -> y3
    y3 IND x1;          !x1 -> y1 -> y3
                        !x1 -> y2 -> y3
                        !x1 -> y1 -> y2 -> y3
    y3 VIA y2 x1;       !x1 -> y2 -> y3
                        !x1 -> y1 -> y2 -> y3
```

Further Readings On Path Analysis

MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G. & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7, 83-104.

MacKinnon, D.P., Lockwood, C.M. & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. Multivariate Behavioral Research, 39, 99-128.

Shrout, P.E. & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. Psychological Methods, 7, 422-445.

Measurement Errors And Multiple Indicators Of Latent Variables

Measurement Error

- Attenuation in correlations
- Measurement error in independent variables – attenuation in regression slopes
- Measurement error in dependent variables – increased standard errors
- Single indicator of a latent variable – known amount of measurement error can be specified
- Multiple indicators of a latent variable – measurement error can be estimated

X With Measurement Error

Regressing on the true η

$$y_i = \alpha + \beta\,\eta_i + \varepsilon_i$$

x measures η measures with error

$$x_i = \eta_i + \delta_i$$

$V(x) = V(\eta) + V(\delta)$. Reliability$(x) = V(\eta) \,/\, (V(\eta) + V(\delta))$

Regressing on x

$$y_i = \alpha^* + \beta^* x_i + \varepsilon_i$$

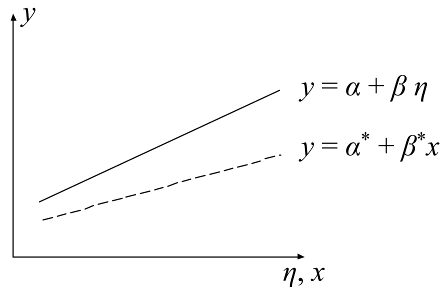
$$\beta^* = \frac{Cov(y,x)}{V(x)} = \frac{Cov(y,\eta)}{V(\eta)+V(\delta)} < \beta$$

Attenuated slope

X With Measurement Error (Continued)

An example:

$$\beta = 0.8$$
$$V(x) = V(\eta) + V(\delta)$$
$$= 1 + 0.43$$
$$\text{Reliability } (x) = 1/(1 + 0.43) = 0.7$$
$$\rightarrow \beta^* = 0.56$$



Measurement Error In A Single Indicator

$x_i = v + \lambda\, \eta_i + \varepsilon_i$

With $\lambda = 1$, $V(y) = \psi + \theta$ and reliability = $\psi / V(y)$

$V(y)$ is estimated as the sample variance, which means that reliability * sample variance = ψ and θ = (1– reliability) * sample variance.

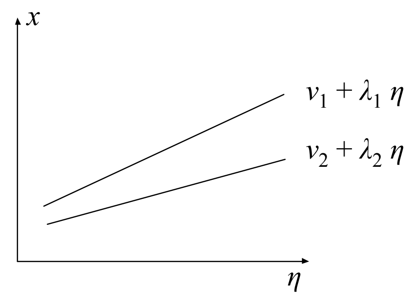
In Mplus: f BY y@1;
y@a;

where a = θ.

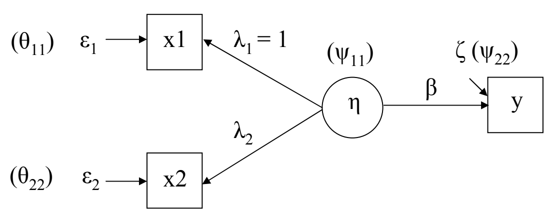
Multiple Indicators Of A Latent Variable

$$x_{1i} = v_1 + \lambda_1 \eta_i + \delta_{1i}$$

$$x_{2i} = v_2 + \lambda_2 \eta_i + \delta_{2i}$$



Multiple Indicators Of An Exogenous Latent Variable

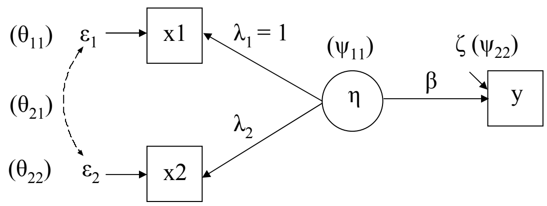


Examples: Alcohol consumption during pregnancy
Dietary fat intake
Blood pressure

β gives the correct picture, free of measurement error (and the influence of collinearity)

$(\beta = Cov(y_1, x_2) / Cov(x_2, x_1))$

Multiple Indicators Of An Exogenous Latent Variable (Continued)



Hypothetical example 1 (β = 0.5)

Reliability(x) = 0.5
$\lambda_1 = \lambda_2 = 1$, $\psi_{11} = 0.5$, $\theta_{11} = \theta_{22} = 0.5$
$\psi_{22} = 0.75$, $R^2(\text{y}) = 0.25$
$\theta_{21} = 0.10$ (corr = 0.20)

$$\beta^* = \frac{0.25}{0.5 + 0.1} \quad \text{(why? See end of day)}$$
$$= 0.41$$

Hypothetical example 2 (β = 0.5)

Reliability(x) = 0.8
Change to $\psi_{11} = 0.8$

$$\beta^* = \frac{0.40}{0.8 + 0.04} = 0.48$$

Factor Analysis

Factor Analysis

Factor analysis is a statistical method used to study the dimensionality of a set of variables. In factor analysis, latent variables represent unobserved constructs and are referred to as factors or dimensions.

- Exploratory Factor Analysis (EFA)

  Used to explore the dimensionality of a measurement instrument by finding the smallest number of interpretable factors needed to explain the correlations among a set of variables – exploratory in the sense that it places no structure on the linear relationships between the observed variables and on the linear relationships between the observed variables and the factors but only specifies the number of latent variables

Factor Analysis (Continued)

- Confirmatory Factor Analysis (CFA)

  Used to study how well a hypothesized factor model fits a new sample from the same population or a sample from a different population – characterized by allowing restrictions on the parameters of the model

Applications Of Factor Analysis

- Personality and cognition in psychology
  - Child Behavior Checklist (CBCL)
  - MMPI
- Attitudes in sociology, political science, etc.
- Achievement in education
- Diagnostic criteria in mental health

The Factor Analysis Model

The factor analysis model expresses the variation and covariation in a set of observed continuous variables y (j = 1 to p) as a function of factors η (k = 1 to m) and residuals ε (j = 1 to p). For person i,

$$y_{i1} = \nu_1 + \lambda_{11}\,\eta_{i1} + \lambda_{12}\,\eta_{i2} + \ldots + \lambda_{1k}\,\eta_{ik} + \ldots + \lambda_{1m}\,\eta_{im} + \varepsilon_{i1}$$

.
.
.

$$y_{ij} = \nu_j + \lambda_{j1}\,\eta_{i1} + \lambda_{j2}\,\eta_{i2} + \ldots + \lambda_{jk}\,\eta_{ik} + \ldots + \lambda_{jm}\,\eta_{im} + \varepsilon_{ij}$$

.
.
.

$$y_{ip} = \nu_p + \lambda_{p1}\,\eta_{i1} + \lambda_{p2}\,\eta_{i2} + \ldots + \lambda_{pk}\,\eta_{ik} + \ldots + \lambda_{pm}\,\eta_{im} + \varepsilon_{ip}$$

The Factor Analysis Model (Continued)

where

ν_j are intercepts

λ_{jk} are factor loadings

η_{ik} are factor values

ε_{ij} are residuals with zero means and correlations of zero with the factors

The Factor Analysis Model (Continued)

In matrix form,

$$\boldsymbol{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i\,,$$

where

$\boldsymbol{\nu}$	is the vector of intercepts ν_j,
$\boldsymbol{\Lambda}$	is the matrix of factor loadings λ_{jk},
$\boldsymbol{\Psi}$	is the matrix of factor variances/covariances, and
$\boldsymbol{\Theta}$	is the matrix of residual variances/covariances

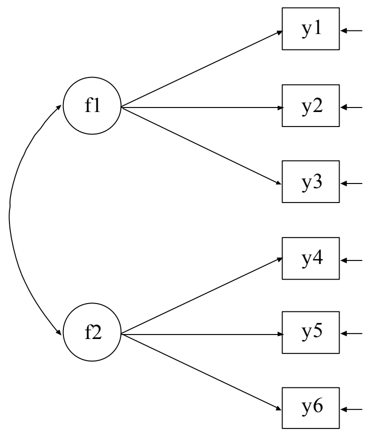
with the population covariance matrix of observed variables $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\, \boldsymbol{\Psi}\, \boldsymbol{\Lambda}' + \boldsymbol{\Theta}.$$

Factor Analysis Terminology

- Factor pattern: Λ
- Factor structure: $\Lambda * \Psi$, correlations between items and factors
- Heywood case: $\theta_{jj} < 0$
- Factor scores: $\hat{\eta}_i$
- Factor determinacy: quality of factor scores; correlation between η_i and $\hat{\eta}_i$

A Two-Factor Model



- Squares or rectangles represent observed variables
- Circles or ovals represent factors or latent variables
- Uni-directional arrows represent regressions or residuals
- Bi-directional arrows represent correlations/covariances

Formulas For The Path Diagram

$$y_{i1} = \nu_1 + \lambda_{11} f_{i1} + 0\, f_{i2} + \varepsilon_{i1}$$
$$y_{i2} = \nu_2 + \lambda_{21} f_{i1} + 0\, f_{i2} + \varepsilon_{i2}$$
$$y_{i3} = \nu_3 + \lambda_{31} f_{i1} + 0\, f_{i2} + \varepsilon_{i3}$$
$$y_{i4} = \nu_4 + 0\, f_{i1} + \lambda_{42} f_{i2} + \varepsilon_{i4}$$
$$y_{i5} = \nu_5 + 0\, f_{i1} + \lambda_{52} f_{i2} + \varepsilon_{i5}$$
$$y_{i6} = \nu_6 + 0\, f_{i1} + \lambda_{62} f_{i2} + \varepsilon_{i6}$$

Elements of $\Sigma = \Lambda\, \Psi\, \Lambda' + \Theta$:

Variance of $y_1 = \sigma_{11} = \lambda_{11}^2\, \psi_{11} + \theta_{11}$

Covariance of $y_1, y_2 = \sigma_{21} = \lambda_{11}\, \psi_{11}\, \lambda_{21}$

Covariance of $y_1, y_4 = \sigma_{41} = \lambda_{11}\, \psi_{21}\, \lambda_{42}$

Recommendations For Using Factor Analysis In Practice

Issues

- History of EFA versus CFA
- Can hypothesized dimensions be found?
  - Validity of measurements

A Possible Research Strategy For Instrument Development

1. Pilot study 1
   - Small n, EFA
   - Revise, delete, add items

Recommendations For Using Factor Analysis In Practice (Continued)

2. Pilot study 2
   - Small n, EFA
   - Formulate tentative CFA model
3. Pilot study 3
   - Larger n, CFA
   - Test model from Pilot study 2 using random half of the sample
   - Revise into new CFA model
   - Cross-validate new CFA model using other half of data
4. Large scale study, CFA
5. Investigate other populations

Exploratory Factor Analysis

Exploratory Factor Analysis (EFA)

Used to explore the dimensionality of a measurement instrument by finding the smallest number of interpretable factors needed to explain the correlations among a set of variables – exploratory in the sense that it places no structure on the linear relationships between the observed variables and the factors but only specifies the number of latent variables

- Find the number of factors
- Determine the quality of a measurement instrument
    - Identify variables that are poor factor indicators
    - Identify factors that are poorly measured

Holzinger-Swineford Data

The data are taken from the classic 1939 study by Karl J. Holzinger and Frances Swineford. Twenty-six tests intended to measure a general factor and five specific factors were administered to seventh and eighth grade students in two schools, the Grant-White School (n = 145) and Pasteur School (n = 156). Students from the Grant-White School came from homes where the parents were American-born. Students from the Pasteur School came from the homes of workers in factories who were foreign-born.

Data for the analysis include nineteen test intended to measure four domains: spatial ability, verbal ability, speed, and memory. Data from the 145 students from the Grant-White School are used.

Holzinger-Swineford Variables

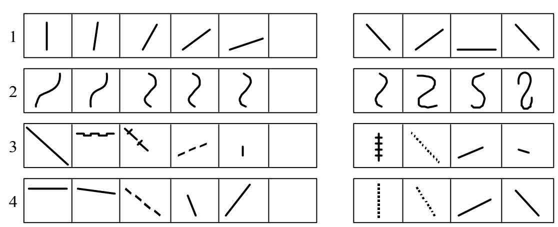
- SPATIAL TESTS
  - Visual perception test
  - Cubes
  - Paper form board
  - Lozenges
- VERBAL TESTS
  - General information
  - Paragraph comprehension
  - Sentence completion
  - Word classification
  - Word meaning

Holzinger-Swineford Variables (Continued)

- SPEED TESTS
  - Add
  - Code
  - Counting groups of dots
  - Straight and curved capitals
- MEMORY
  - Word recognition
  - Number recognition
  - Figure recognition
  - Object-number
  - Number-figure
  - Figure-word

Examples Of Holzinger-Swineford Variables

Test 1 Visual-Perception Test



Test 5 General Information

```
In each sentence below you have four choices for the last
word, but only one is right. From the last four words of
each sentence, select the right one and underline it.
EXAMPLE: Men see with their ears, nose, eyes, mouths.
----------------------------------------------------------
1. Pumpkins grow on bushes, trees, vines, shrubs.
2. Coral comes from reefs, mines, trees, tusks.
3. Sugar cane grows mostly in Montana, Texas, Illinois,
New York
```

Examples Of Holzinger-Swineford Variables (Continued)

Test 17 Object-Number

Name______________

Date______________

Here is a list of objects. Each one has a number. Study the list so that you can remember the number of each object.

Object	Number
apple	29
brush	71
candy	58
chair	44
cloud	53
dress	67
flour	15
grass	32
heart	86

After each object, write the number that belongs to it.

Object	Number
pupil	
chair	
house	
sugar	
flour	
river	
apple	
match	
train	

Sample Correlations For Holzinger-Swineford Data

	VISUAL	CUBES	PAPER	LOZENGES	GENERAL
VISUAL					
CUBES	.326				
PAPER	.372	.190			
LOZENGES	.449	.417	.366		
GENERAL	.328	.275	.309	.381	
PARAGRAP	.342	.228	.260	.328	.622
SENTENCE	.309	.159	.266	.287	.654
WORDC	.326	.156	.334	.380	.574
WORDM	.317	.195	.260	.347	.720
ADDITION	.104	.066	.128	.075	.314
CODE	.306	.151	.248	.181	.342
COUNTING	.308	.168	.198	.239	.210
STRAIGHT	.487	.248	.389	.373	.343
WORDR	.130	.082	.250	.161	.261
NUMBERR	.223	.135	.186	.205	.219
FIGURER	.419	.289	.307	.289	.177
OBJECT	.169	.011	.128	.139	.213
NUMBERF	.364	.264	.259	.353	.259
FIGUREW	.267	.110	.155	.180	.196

Sample Correlations For Holzinger-Swineford Data (Continued)

	PARAGRAP	SENTENCE	WORDC	WORDM	ADDITION
SENTENCE	.719				
WORDC	.520	.633			
WORDM	.714	.685	.537		
ADDITION	.209	.254	.297	.179	
CODE	.360	.248	.294	.287	.468
COUNTING	.104	.198	.290	.121	.587
STRAIGHT	.314	.356	.405	.272	.418
WORDR	.286	.233	.243	.250	.157
NUMBERR	.249	.157	.170	.213	.150
FIGURER	.288	.201	.299	..236	.137
OBJECT	.276	.251	.271	.285	.301
NUMBERF	.167	.176	.258	.213	.320
FIGUREW	.251	.241	.261	.277	.199

Sample Correlations For Holzinger-Swineford Data (Continued)

	CODE	COUNTING	STRAIGHT	WORDR	NUMBERR
COUNTING	.422				
STRAIGHT	.527	.528			
WORDR	.324	.130	.193		
NUMBERR	.238	.163	.138	.387	
FIGURER	.314	.128	.277	.382	.313
OBJECT	.357	.278	.191	.372	.346
NUMBERF	.346	.347	.325	.199	.318
FIGUREW	.290	.108	.252	.219	.183

	FIGURER	OBJECT	NUMBERF	FIGUREW
OBJECT	.339			
NUMBERF	.355	.452		
FIGUREW	.254	.327	.358	

EFA Model Estimation

Estimators

In EFA, a correlation matrix is analyzed.

- ULS – minimizes the residuals, observed minus estimated correlations
  - Fast
  - Not fully efficient
- ML – minimizes the differences between matrix summaries (determinant and trace) of observed and estimated correlations
  - Computationally more demanding
  - Efficient

EFA Model Indeterminacies And Rotations

A model that is identified has only one set of parameter values. To be identified, an EFA model must have m^2 restrictions on factor loadings, variances, and covariances. There are an infinite number of possible ways to place the restrictions. In software, restrictions are placed in two steps.

Step 1 – Mathematically convenient restrictions

- m(m+1)/2 come from fixing the factor variances to one and the factor covariances to zero
- m(m-1)/2 come from fixing (functions of) factor loadings to zero
  - ULS – $\Lambda'\Lambda$ diagonal
  - ML – $\Lambda' \Theta^{-1} \Lambda$ diagonal
  - General approach – fill the upper right hand corner of lambda with zeros

EFA Model Indeterminacies And Rotations (Continued)

Step 2 – Rotation to interpretable factors

Starting with a solution based on mathematically convenient restrictions, a more interpretable solution can be found using a rotation. There are two major types of rotations: orthogonal (uncorrelated factors) and oblique (correlated factors).

- Do an orthogonal rotation to maximize the number of factor loadings close to one and close to zero
- Do an oblique rotation of the orthogonal solution to obtain factor loadings closer to one and closer to zero

New EFA Features In Mplus Version 5

- Several new rotations including Quartimin and Geomin
- Standard errors for rotated loadings and factor correlations
- Non-normality robust standard errors and chi-square tests of model fit
- Modification indices for residual correlations
- Maximum likelihood estimation with censored, categorical, and count variables
- Exploratory factor analysis for complex survey data (stratification, clustering, and weights)
  TYPE = COMPLEX EFA # #;
- Exploratory factor mixture analysis with class-specific rotations
  TYPE = MIXTURE EFA # #;
- Two-level exploratory factor analysis for continuous and categorical variables with new rotations and standard errors, including unrestricted model for either level
  TYPE = TWOLEVEL EFA # # UW # # UB;

Determining The Number Of Factors That Explain The Correlations Among Variables

Descriptive Values

- Eigenvalues
- Residual Variances

Tests Of Model Fit

- RMSR – average residuals for the correlation matrix – recommend to be less than .05

Determining The Number Of Factors That Explain The Correlations Among Variables (Continued)

- Chi-Square – tests that the model does not fit significantly worse than a model where the variables correlate freely – p-values greater than .05 indicate good fit

  H_0: Factor model
  H_1: Unrestricted correlations model
  If $p < .05$, H_0 is rejected
  Note: We want large p

- RMSEA – function of chi-square – test of close fit – value less than .05 recommended

$$RMSEA = \sqrt{\max[(\chi^2 / n\,d - 1/n), 0]}\sqrt{G}$$

where d is the number of degrees of freedom of the model and G is the number of groups.

Steps In EFA

- Carefully develop or use a carefully developed set of variables that measure specific domains

- Determine the number of factors
  - Descriptive values
    - Eigenvalues
    - Residual variances
  - Tests of model fit
    - RMSR
    - Chi-square
    - RMSEA

Steps In EFA (Continued)

- Interpret the factors

- Determine the quality of the variables measuring the factors
  - Size loadings
  - Cross loadings

- Determine the quality of the factors
  - Number of variables that load on the factor
  - Factor determinacy – correlation between the estimated factor score and the factor

- Eliminate poor variables and factors and repeat EFA steps

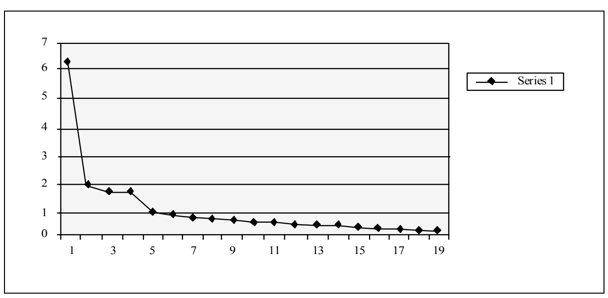
Input For Holzinger-Swineford EFA

```
TITLE:      EFA on 19 variables from Holzinger and Swineford
            (1939)

DATA:       FILE IS holzall.dat;
            FORMAT IS f3,2f2,f3,2f2/3x,13(1x,f3)/3x,11(1x,f3);

VARIABLE:   NAMES ARE id female grade agey agem school visual
            cubes paper lozenges general paragrap sentence wordc
            wordm addition code counting straight wordr numberr
            figurer object numberf figurew deduct numeric
            problemr series arithmet;

            USEV ARE visual cubes paper lozenges general
            paragrap sentence wordc wordm addition code counting
            straight wordr numberr figurer object numberf
            figurew;

            USEOBS IS school EQ 0;

ANALYSIS:   TYPE=EFA 1 8; ESTIMATOR = ML;
```


Determine The Number Of Factors

Examine The Eigenvalues

- Number greater than one
- Scree plot



Determine The Number Of Factors (Continued)

Examine The Fit Measures And Residual Variances
(ML, n = 145)

Factors	Chi-Square x^2 df p	RMSEA	RMSR	Negative Res. Var.
1	469.81 (152) .000	.120	.1130	no
2	276.44 (134) .000	.086	.0762	no
3	188.75 (117) .000	.065	.0585	no
4	110.34 (101) .248	.025	.0339	no
5	82.69 (86) .581	.000	.0280	no
6	no. conv.			
7	no. conv.			
8	no. conv.			

Interpret The Factors

- Examine factor loadings for the set of possible solutions
- Determine if factors support theory

Output Excerpts Holzinger-Swineford EFA Using 19 Variables

Promax Rotated Loadings – 3 Factor Solution

	SPATIAL/ MEMORY 1	SPEED 2	VERBAL 3
VISUAL	**.740**	-.087	-.002
CUBES	**.522**	-.118	-.008
PAPER	**.508**	-.028	.058
LOZENGES	**.650**	-.153	.092
GENERAL	.043	.084	**.745**
PARAGRAP	.090	-.066	**.803**
SENTENCE	-.052	.046	**.851**
WORDC	.144	.136	**.547**
WORDM	.061	-.092	**.853**
ADDITION	-.257	**.923**	.073
CODE	.223	**.482**	.054
COUNTING	.112	**.728**	-.149
STRAIGHT	.389	.405	.013

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 3 Factor Solution

	SPATIAL/ MEMORY 1	SPEED 2	VERBAL 3
WORDR	**.284**	.063	.128
NUMBERR	**.374**	.038	.022
FIGURER	**.666**	-.072	-.063
OBJECT	.214	.270	.086
NUMBERF	**.534**	.237	-.146
FIGUREW	**.302**	.090	.094

Promax Factor Correlations

	1	2	3
1	1.000		
2	.536	1.000	
3	.539	.379	1.000

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3	SPEED 4
VISUAL	**.713**	.027	.008	.005
CUBES	**.541**	-.051	.007	-.050
PAPER	**.466**	.047	.070	.022
LOZENGES	**.650**	-.028	.106	-.062
GENERAL	.094	-.043	**.749**	.083
PARAGRAP	.040	.107	**.791**	-.092
SENTENCE	.002	-.050	**.846**	.052
WORDC	.155	.014	**.550**	.146
WORDM	.022	.078	**.840**	-.107
ADDITION	-.203	.108	.081	**.785**
CODE	.087	.289	.055	.419
COUNTING	.179	-.024	-.132	**.760**
STRAIGHT	.479	-.094	.033	.486

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3	SPEED 4
WORDR	-.037	**.551**	.098	-.052
NUMBERR	.062	**.532**	-.006	-.064
FIGURER	.368	.504	-.086	-.141
OBJECT	-.205	**.736**	.042	.119
NUMBERF	.275	.446	-.154	.178
FIGUREW	.082	**.376**	.080	.019

Promax Factor Correlations

	1	2	3	4
1	1.000			
2	.468	1.000		
3	.468	.421	1.000	
4	.360	.388	.325	1.000

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Varimax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3	SPEED 4
VISUAL	**.666**	.194	.183	.143
CUBES	**.487**	.072	.117	.042
PAPER	**.455**	.170	.191	.126
LOZENGES	**.608**	.135	.241	.068
GENERAL	.230	.133	**.743**	.183
PARAGRAP	.195	.244	**.772**	.038
SENTENCE	.158	.119	**.808**	.146
WORDC	.267	.174	**.589**	.242
WORDM	.180	.219	**.806**	.021
ADDITION	-.062	.189	.177	**.754**
CODE	.191	.367	.197	.486
COUNTING	.224	.110	.034	**.748**
STRAIGHT	.489	.103	.206	.545

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Varimax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3	SPEED 4
WORDR	.077	**.522**	.184	.064
NUMBERR	.144	**.506**	.103	.054
FIGURER	.398	.524	.081	.021
OBJECT	-.036	**.673**	.155	.229
NUMBERF	.326	.484	.034	.293
FIGUREW	.160	**.392**	.173	.118

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 5 Factor Solution

	SPATIAL 1	2	VERBAL 3	MEMORY 4	SPEED 5
VISUAL	**.613**	.211	.006	.050	.011
CUBES	**.552**	-.044	.029	-.028	-.044
PAPER	**.399**	.187	.058	.057	.021
LOZENGES	**.696**	-.070	.129	-.018	-.051
GENERAL	.137	-.094	**.771**	-.042	.096
PARAGRAP	-.006	.131	**.772**	.110	-.100
SENTENCE	-.010	.083	**.826**	-.049	.049
WORDC	.149	.061	**.543**	.018	.145
WORDM	.050	-.075	**.845**	.081	-.097
ADDITION	-.185	.032	.095	.113	**.765**
CODE	-.009	.291	.028	.295	.413
COUNTING	.167	.098	-.112	-.014	**.744**
STRAIGHT	.374	.474	-.013	-.124	.497

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 5 Factor Solution

	SPATIAL 1	2	VERBAL 3	MEMORY 4	SPEED 5
WORDR	-.085	.098	.082	**.552**	-.060
NUMBERR	.071	-.094	.010	**.543**	-.059
FIGURER	.286	.144	-.101	.533	-.150
OBJECT	-.160	-.163	.056	**.745**	.124
NUMBERF	.358	-.256	-.126	.502	.195
FIGUREW	.074	.004	.075	**.386**	.026

Promax Factor Correlations

	1	2	3	4	5
1	1.000				
2	.206	1.000			
3	.415	.287	1.000		
4	.425	.335	.424	1.000	
5	.305	.035	.275	.343	1.000

Output Excerpts Using 19 Variables
Quartimin Rotated Loadings

	SPATIAL	MEMORY	VERBAL	SPEED
VISUAL	**0.646**	0.076	0.092	0.050
CUBES	**0.488**	-0.010	0.064	-0.018
PAPER	**0.422**	0.077	0.128	0.053
FLAGS	**0.585**	0.017	0.178	-0.021
GENERAL	0.058	-0.049	**0.773**	0.093
PARAGRAP	0.019	0.088	**0.810**	-0.079
SENTENCE	-0.028	-0.064	**0.860**	0.056
WORDC	0.121	0.014	**0.584**	0.159
WORDM	0.000	0.058	**0.855**	-0.095
ADDITION	-0.196	0.093	0.100	**0.769**
CODE	0.084	0.283	0.100	0.431
COUNTING	0.149	-0.001	-0.081	**0.761**
STRAIGHT	0.418	-0.051	0.105	0.507

Output Excerpts Using 19 Variables
Quartimin Rotated Loadings (Continued)

	SPATIAL	MEMORY	VERBAL	SPEED
WORDR	-0.006	**0.517**	0.124	-0.034
NUMBERR	0.086	**0.509**	0.028	0.041
FIGURER	0.366	0.505	-0.023	-0.100
OBJECT	-0.150	**0.683**	0.065	0.131
NUMBERF	0.274	0.447	-0.092	0.207
FIGUREW	0.091	**0.361**	0.113	0.037

QUARTIMIN FACTOR CORRELATIONS

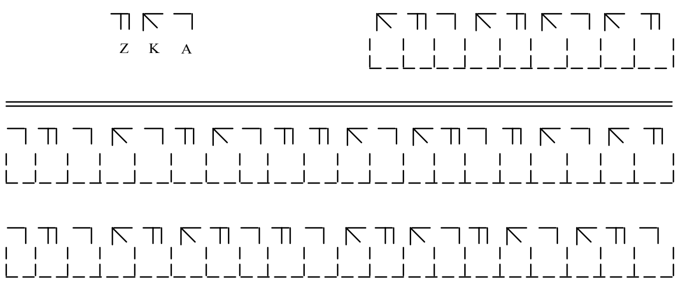
SPATIAL	1.000			
MEMORY	0.289	1.000		
VERBAL	0.371	0.377	1.000	
SPEED	0.266	0.323	0.290	1.000

Determine The Quality Of The Variables

Examine Cross Loadings

Four variables have cross loadings:

- Code (Speed) – loads on Memory and Speed factors
  - Requires matching letters to a set of figures



Determine The Quality Of The Variables (Continued)

- Straight (Speed) – loads on Spatial and Speed factors
  - Requires deciding if a letter consists of entirely straight lines or has curved lines

T	F	G	O	E	Y	Q	D	N	L	P	V	M	J	O

T	Z	F	S	H	J	K	L	D	T	F	U	C	T	U

E	H	K	B	L	S	J	R	R	O	P	T	K	E	M

N	T	D	C	U	S	Z	Y	H	O	T	L	F	C	N

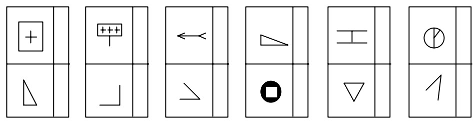
U	D	Q	J	F	D	U	H	L	V	P	E	S	U	Y

D	J	L	H	T	D	C	J	Q	P	R	E	J	L	D

Determine The Quality Of The Variables (Continued)

- Figure (Memory) – loads on Spatial and Memory
  - Requires remembering a set of figures

```
Put a check mark ( √ ) in the space after each
figure that was on the study sheet. Do not put
a check after any figure that you have not
studied.
```



Determine The Quality Of The Variables (Continued)

- Numberf (Memory) – loads on Spatial and Memory
  - Requires remembering a figure and associating it with a number

Here is a list of numbers. Each has a figure, or picture, with it. Study the list so that you can remember the figure that belongs with each number.

Number	Figure
52	
74	
12	

After each number draw the figure that belongs with it

Number	Figure
17	
65	
37	

Deleting Four Items That Have Cross Loadings

Output Excerpts Holzinger-Swineford EFA Using 15 Variables

Promax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	SPEED 3	VERBAL 4
VISUAL	**.590**	.040	.078	.034
CUBES	**.566**	-.089	.007	-.012
PAPER	**.419**	.104	.029	.056
LOZENGES	**.734**	-.012	-.014	.028
GENERAL	.128	-.037	.050	**.739**
PARAGRAP	.031	.108	-.118	**.792**
SENTENCE	-.041	-.044	.043	**.878**
WORDC	.132	.008	.158	**.568**
WORDM	.043	.060	-.109	**.826**
ADDITION	-.161	.087	**.698**	.127
COUNTING	.200	-.012	**.841**	-.147
WORDR	.000	**.613**	-.066	.023
NUMBERR	.133	**.585**	-.044	-.104
OBJECT	-.127	**.646**	.144	.019
FIGUREW	.066	**.350**	-.004	.096

Output Excerpts Holzinger-Swineford EFA Using 15 Variables (Continued)

Note that factor structure is maintained and that speed has only two indicators

Promax Factor Correlations

	1	2	3	4
1	1.000			
2	.386	1.000		
3	.258	.355	1.000	
4	.495	.478	.309	1.000

Output Excerpts Holzinger-Swineford EFA Using 15 Variables (Continued)

Estimated Error Variances

VISUAL	CUBES	PAPER	LOZENGES	GENERAL
.576	.714	.738	.452	.346

PARAGRAP	SENTENCE	WORDC	WORDM	ADDITION
.306	.274	.488	.279	.444

COUNTING	WORDR	NUMBERR	OBJECT	FIGUREW
.257	.635	.657	.541	.809

Tests Of Model Fit

Chi-square	48.636 (51) .5681
RMSEA	.000
RMSR	.0275

Deleting A Factor With Only Two Items

Output Excerpts Holzinger-Swineford EFA Using 13 Variables

Promax Rotated Loadings – 3 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3
VISUAL	**0.577**	0.061	0.035
CUBES	**0.602**	-0.114	-0.039
PAPER	**0.434**	0.115	0.033
LOZENGES	**0.765**	-0.032	-0.010
GENERAL	0.152	-0.029	**0.728**
PARAGRAP	0.009	0.080	**0.777**
SENTENCE	-0.060	-0.015	**0.891**
WORDC	0.149	0.065	**0.572**
WORDM	0.015	0.037	**0.816**
WORDR	-0.023	**0.611**	0.010
NUMBERR	0.116	**0.573**	-0.114
OBJECT	-0.127	**0.678**	0.043
FIGUREW	0.081	**0.351**	0.076

Output Excerpts Holzinger-Swineford EFA Using 13 Variables (Continued)

Promax Factor Correlations

	1	2	3
1	1.000		
2	0.436	1.000	
3	0.540	0.486	1.000

Tests Of Model Fit

Chi-square	39.027 (42) .6022
RMSEA	0.000
RMSR	0.0301

Practical Issues Related To EFA

Choice Of Variables – results can be influenced by the set of variables used.

- EFA requires a set of variables that has been carefully developed to measure certain domains, not just any set of variables.
- Number of factors can be influenced by the number of variables per factor.
- Similar number of variables per factor – at least four or five variables per factor is recommended.

Sample Size

- Advantages of large sample size
  - Sample correlations have smaller sampling variability—closer to population values
  - Reduces Heywood cases, negative residual variances

Practical Issues Related To EFA (Continued)

- Several observations per estimated parameter are recommended
- Advantages of small sample size
  - Can avoid heterogeneity
  - Can avoid problems with sensitivity of chi-square

Size Of Factor Loadings – no general rules

Elimination Of Factors/Variables

- Drop variables that poorly measure factors
- Drop factors that are poorly measured

Maximum Number Of Factors That Can Be Extracted

$a \leq b$ where a = number of parameters to be estimated (H_0)
b = number of variances/covariances (H_1)

$$a = \underbrace{p\,m}_{\Lambda} + \underbrace{m\,(m+1)/2}_{\Psi} + \underbrace{p}_{\Theta} - m^2$$

$b = p\,(p + 1)/2$

where p = number of observed variables
m = number of factors

Example: $p = 5$ which gives $b = 15$
$m = 1$: $a = 10$
$m = 2$: $a = 14$
$m = 3$: $a = 17$

Even if $a \leq b$, it may not be possible to extract m factors due to Heywood cases.

Sample Size

- Stability of sample correlations
  - $V(r) = (1 - \rho^2)^2/n$
  - Example: $\rho = 0.5$, $s.d. = 0.1$, $n = 56$
- Stability of estimates
  - n larger than the number of parameters
  - Example: 5 dimensions hypothesized, 5 items per dimension, number of EFA parameters = 140, n = 140-1400 in order to have 1-10 observations per parameter
- Monte Carlo studies (Muthén & Muthén, 2002)

Further Readings On EFA

Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research, 36, 111-150.

Cudeck, R. & O'Dell, L.L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. Psychological Bulletin, 115, 475-487.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4, 272-299.

Gorsuch, R.L. (1983). Factor analysis. 2nd edition. Hillsdale, N.J.: Lawrence Erlbaum.

Kim, J.O. & Mueller, C.W. (1978). An introduction to factor analysis: what it is and how to do it. Sage University Paper series on Quantitative Applications in the Social Sciences, No 07-013. Beverly Hills, CA: Sage.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC: American Psychological Association.

Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA)

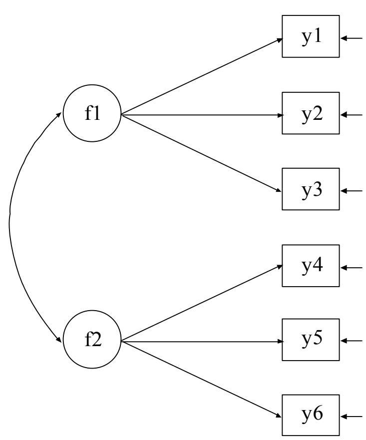
Used to study how well a hypothesized factor model fits a new sample from the same population or a sample from a different population. CFA is characterized by allowing restrictions on factor loadings, variances, covariances, and residual variances.

- See if factor models fits a new sample from the same population – the confirmatory aspect
- See if the factor models fits a sample from a different population – measurement invariance
  - Study the properties of individuals by examining factor variances, and covariances
    - Factor variances show the heterogeneity in a population
    - Factor correlations show the strength of the association between factors

Confirmatory Factor Analysis (CFA) (Continued)

- Study the behavior of new measurement items embedded in a previously studied measurement instrument
- Estimate factor scores
- Investigate an EFA more fully

A Two-Factor CFA Model



- Squares or rectangles represent observed variables
- Circles or ovals represent factors or latent variables
- Uni-directional arrows represent regressions or residuals
- Bi-directional arrows represent correlations/covariances

The CFA Model

The CFA model is the same as the EFA model with the exception that restrictions can be placed on factor loadings, variances, covariances, and residual variances resulting in a more parsimonious model. In addition residual covariances can be part of the model.

Measurement Parameters – describe measurement characteristics of observed variables

- Intercepts
- Factor loadings
- Residual variances

The CFA Model (Continued)

Structural Parameters – describe characteristics of the population from which the sample is drawn

- Factor means
- Factor variances
- Factor covariances

Metric Of Factors – needed to determine the scale of the latent variables

- Fix one factor loading to one
- Fix the factor variance to one

CFA Model Identification

Necessary Condition For Identification

$a \leq b$ where a = number of parameters to be estimated in H_0
b = number of variances/covariances in H_1

Sufficient Condition For Identification

Each parameter can be solved for in terms of the variances and covariances

CFA Model Identification (Continued)

Practical Way To Check

- Program will complain if a parameter is most likely not identified.
- If a fixed or constrained parameter has a modification index of zero, it will not be identified if it is free.

Models Known To Be Identified

- One factor model with three indicators
- A model with two correlated factors each with two indicators

CFA Modeling Estimation And Testing

Estimator

In CFA, a covariance matrix is analyzed.

- ML – minimizes the differences between matrix summaries (determinant and trace) of observed and estimated variances/covariances
- Robust ML – same estimates as ML, standard errors and chi-square robust to non-normality of outcomes and non-independence of observations

Chi-square test of model fit

Tests that the model does not fit significantly worse than a model where the variables correlate freely – p-values greater than or equal to .05 indicate good fit

H_0: Factor model
H_1: Free variance-covariance model
If $p < .05$, H_0 is rejected
Note: We want large p

CFA Modeling Estimation And Testing (Continued)

Model fit indices (cutoff recommendations for good fit based on Yu, 2002 / Hu & Bentler, 1999; see also Marsh et al, 2004)

- CFI – chi-square comparisons of the target model to the baseline model – greater than or equal to .96/.95
- TLI – chi-square comparisons of the target model to the baseline model – greater than or equal to .95/.95
- RMSEA – function of chi-square, test of close fit – less than or equal to .05 (not good at n=100)/.06
- SRMR – average correlation residuals – less than or equal to .07 (not good with binary outcomes)/.08
- WRMR – average weighted residuals – less than or equal to 1.00 (also good with non-normal and categorical outcomes – not good with growth models with many timepoints or multiple group models)

Degrees Of Freedom For Chi-Square Testing Against An Unrestricted Model

The p value of the χ^2 test gives the probability of obtaining a χ^2 value this large or larger if the H_0 model is correct (we want high p values).

Degrees of Freedom:
(Number of parameters in H_1) – (number parameters in H_0)

Number of H_1 parameters with an unrestricted $\boldsymbol{\Sigma}$: $p\ (p + 1)/2$

Number of H_1 parameters with unrestricted $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:
$p + p\ (p + 1)/2$

A degrees of freedom example – EFA

- $p\ (p + 1)/2 - (p\ m + m\ (m + 1)/2 + p - m^2)$
  Example: if $p = 5$ and $m = 2$, then $df = 1$

Chi-Square Difference Testing Of Nested Models

- When a model H_a imposes restrictions on parameters of model H_b, H_a is said to be nested within H_b
- To test if the nested model H_a fits significantly worse than H_b, a chi-square test can be obtained as the difference in the chi-square values for the two models (testing against an unrestricted model) using as degrees of freedom the difference in number of parameters for the two models
- The chi-square difference is the same as 2 times the difference in log likelihood values for the two models
- The chi-square theory does not hold if H_a has restricted any of the H_b parameters to be on the border of their admissible parameter space (e.g. variance = 0)

CFA Model Modification

Model modification indices are estimated for all parameters that are fixed or constrained to be equal.

- Modification Indices – expected drop in chi-square if the parameter is estimated
- Expected Parameter Change Indices – expected value of the parameter if it is estimated
- Standardized Expected Parameter Change Indices – standardized expected value of the parameter if it is estimated

Model Modifications

- Residual covariances
- Factor cross loadings

Factor Scores

Factor Score

- Estimate of the factor value for each individual based on the model and the individual's observed scores
- Regression method (Empirical Bayes; EAP)
- Estimated factor scores do not have the same relations as true scores

Factor Determinacy

- Measure of how well the factor scores are estimated
- Correlation between the estimated score and the true score
- Ranges from 0 to 1 with 1 being best

Uses Of Factor Scores

- Rank people on a dimension
- Create percentiles
- Proxies for latent variables
    - Independent variables in a model – not as dependent variables

Technical Aspects Of Maximum-Likelihood Estimation And Testing

ML Estimation

The ML estimator chooses parameter values (estimates) so that the likelihood of the sample is maximized. Normal theory ML assumes multivariate normality for $\boldsymbol{y}_i$ and n i.i.d. observations,

$$logL = -c - n / 2 \log |\boldsymbol{\Sigma}| - 1 / 2 \boldsymbol{A}, \qquad (1)$$

where $c = n / 2 \log (2\pi)$ and

$$A = \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}) \qquad (2)$$

$$= trace\, [\boldsymbol{\Sigma}^{-1} \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{y}_i - \boldsymbol{\mu})'] \qquad (3)$$

$$= n\ trace\, [\boldsymbol{\Sigma}^{-1} (\boldsymbol{S} + (\bar{\boldsymbol{y}} - \boldsymbol{\mu})(\bar{\boldsymbol{y}} - \boldsymbol{\mu})']. \qquad (4)$$

ML Estimation (Continued)

This leads to the ML fitting function to be minimized with respect to the parameters

$$F_{ML}\,(\boldsymbol{\pi}) = 1/2\;[ln\,|\boldsymbol{\Sigma}| + trace\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{T}) - ln\,|\boldsymbol{S}| - p], \tag{5}$$

where

$$\boldsymbol{T} = \boldsymbol{S} + (\bar{\boldsymbol{y}} - \boldsymbol{\mu})\,(\bar{\boldsymbol{y}} - \boldsymbol{\mu})'. \tag{6}$$

When there is no mean structure, $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}$, and

$$F_{ML}\,(\boldsymbol{\pi}) = 1/2\;[ln\,|\boldsymbol{\Sigma}| + trace\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}) - ln\,|\boldsymbol{S}| - p]. \tag{7}$$

Model Testing

The standard H_1 model considers an unrestricted mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Under this model $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}$ and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}$, which gives the maximum-likelihood value

$$logL_{H_1} = -c - n\,/\,2\;log\,|\boldsymbol{S}| - n\,/\,2\;p, \tag{8}$$

Note that

$$F_{ML}\,(\boldsymbol{\pi}) = -logL/n + logL_{H_1}/n, \tag{9}$$

Letting $\hat{\pi}$ denote the ML estimate under H_0, the value of the likelihood-ratio χ^2-test of model fit for H_0 against H_1 is therefore obtained as $2\;n\;F_{ML}\,(\hat{\boldsymbol{\pi}})$

Model Fit With Non-Normal Continuous Outcomes

- Non-normality robust chi-square testing
  - A robust goodness-of-fit test (cf. Satorra & Bentler, 1988, 1994; Satorra, 1992) is obtained as the mean-adjusted chi square defined as

$$T_m = 2\, n\, F\,(\hat{\pi}) \,/\, c, \qquad (1)$$

where c is a scaling correction factor,

$$c = tr[\mathbf{U\Gamma}] \,/\, d, \qquad (2)$$

with

$$\mathbf{U} = (\mathbf{W}^{-1} - \mathbf{W}^{-1}\, \mathbf{\Delta}\, (\mathbf{\Delta}'\mathbf{W}^{-1}\mathbf{\Delta})^{-1}\mathbf{\Delta}'\mathbf{W}^{-1}) \qquad (3)$$

and where d is the degrees of freedom of the model.

Model Fit With Non-Normal Continuous Outcomes (Continued)

- Chi-square difference testing with robust (mean-adjusted) chi-square T_{md} (Satorra, 2000, Satorra & Bentler, 1999)

$$T_{md} = (T_0 - T_1)/c_d\,, \qquad (4)$$

$$= (T_{m0}\, c_0 - T_{m1}\, c_1)/c_d\,, \qquad (5)$$

$$c_d = (d_0\, c_0 - d_1\, c_1)/(d_0 - d_1), \qquad (6)$$

where the 0/1 subscript refers to the more/less restrictive model, c refers to a scaling correction factor, and d refers to degrees of freedom.

Common Model Fit Indices

- Root mean square error of approximation (RMSEA) (Browne & Cudeck, 1993; Steiger & Lind, 1980). With continuous outcomes, RMSEA is defined as

$$RMSEA = \sqrt{max[(2\,F_{ML}\,(\hat{\pi})/d - 1/n),0]}\,\sqrt{G} \quad (7)$$

where d is the number of degrees of freedom of the model and G is the number of groups. With categorical outcomes, Mplus replaces d in (7) by $tr[\mathbf{U\Gamma}]$.

- TLI and CFI

$$TLI = (\chi^2_B / d_B - \chi^2_{H_0} / d_{H_0}) / (\chi^2_B / d_B - 1), \quad (8)$$

$$CFI = 1 - max\,(\chi^2_{H_0} - d_{H_0}, 0) / max\,(\chi^2_{H_0} - d_{H_0}, \chi^2_B - d_B, 0), \quad (9)$$

Common Model Fit Indices (Continued)

where d_B and d_{H_0} denote the degrees of freedom of the baseline and H_0 models, respectively. The baseline model has uncorrelated outcomes with unrestricted variances and unrestricted means and / or thresholds.

- SRMR (standardized root mean square residual)

$$SRMR = \sqrt{\sum_j \sum_{k \leq j} r^2_{jk} / e\,.} \quad (10)$$

Here, $e = p\,(p + 1)/2$, where p is the number of outcomes and r_{jk} is a residual in a correlation metric.

A New Model Fit Index

WRMR (weighted root mean square residual) is defined as

$$WRMR = \sqrt{\sum_{r}^{e} \frac{(s_r - \sigma_r)^2}{v_r} \Big/ e}\,, \qquad (20)$$

where s_r is an element of the sample statistics vector, $\hat{\sigma}_r$ is the estimated model counterpart, v_r is an estimate of the asymptotic variance of s_r, and the e is the number of sample statistics. WRMR is suitable for models where sample statistics have widely varying variances, when sample statistics are on different scales such as in models with mean structures, with non-normal continuous outcomes, and with categorical outcomes including models with threshold structures.

Computational Issues Related To CFA

- Scale of observed variables – important to keep them on a similar scale
- Convergence – often related to starting values or the type of model being estimated
  - Program stops because maximum number of iterations has been reached
    - If no negative residual variances, either increase the number of iterations or use the preliminary parameter estimates as starting values
    - If there are large negative residual variances, try better starting values
  - Program stops before the maximum number of iterations has been reached
    - Check if variables are on a similar scale
    - Try new starting values
- Starting values – the most important parameters to give starting values to are residual variances

Mplus MODEL Command For CFA

MODEL command is used to describe the model to be estimated

BY statement is used to define the latent variables or factors

BY is short for "measured by"

Example 1 – standard parameterization

```
MODEL:   f1  BY  y1 y2 y3;
         f2  BY  y4 y5 y6;
```

Defaults

- Factor loading of first variable after BY is fixed to one
- Factor loadings of other variables are estimated
- Residual variances are estimated
- Residual covariances are fixed to zero
- Variances of factors are estimated
- Covariance between the exogenous factors is estimated

Mplus MODEL Command For CFA (Continued)

Example 2 – Alternative parameterization

```
MODEL:  f1 BY y1* y2 y3;
        f2 BY y4* y5 y6;
        f1@1 f2@1;        ! or f1-f2@1;
```

EFA In A CFA Framework

EFA In A CFA Framework (Jöreskog, 1969)

- Purpose
  - To obtain standard errors to determine if factor loadings are statistically significant
  - To obtain modification indices to determine if residual covariances are needed to represent minor factors
- Use the same number of restrictions as an exploratory factor analysis model – m^2
  - Fix factor variances to one for m restrictions
  - Fix factor loadings to zero for the remaining restrictions
    - Find an anchor item for each factor – select an item that has a large loading for the factor and small loadings for other factors
    - Fix the loading of the anchor item to zero for all of the other factors
    - Allow all other factor loadings to be free
- Will get the same model fit as EFA

Exploratory Structural Equation Modeling (ESEM)

- ESEM replaces EFA in a CFA framework
  – SEs for rotated loadings and factor correlations
  – MIs for residual correlations
- ESEM makes EFA possible
  – for multiple time points
  – for multiple groups
  – as a measurement model in SEM
  – using target rotation

EFA In CFA: Selecting Anchor Items

Promax Rotated Loadings – 3 Factor Solution

	Spatial	Memory	Verbal
VISUAL	0.577	0.061	0.035
CUBES	0.602	-0.114	-0.039
PAPER	0.434	0.115	0.033
LOZENGES	**0.765**	**-0.032**	**-0.010**
GENERAL	0.152	-0.029	0.728
PARAGRAP	0.009	0.080	0.777
SENTENCE	**-0.060**	**-0.015**	**0.891**
WORDC	0.149	0.065	0.572
WORDM	0.015	0.037	0.816
WORDR	**-0.023**	**0.611**	**0.010**
NUMBERR	0.116	0.573	-0.114
OBJECT	0.127	0.678	0.043
FIGUREW	0.081	0.351	0.076

Input For Holzinger-Swineford EFA In A CFA Framework Using 13 Variables

```
TITLE:      EFA in a CFA framework using 13 variables from
            Holzinger and Swineford (1939)
DATA:       FILE IS holzall.dat;
            FORMAT IS f3,2f2,f3,2f2/3x,13(1x,f3)/3x,11(1x,f3);
VARIABLE:   NAMES ARE id female grade agey agem school visual
            cubes paper lozenges general paragrap sentence wordc
            wordm addition code counting straight wordr numberr
            figurer object numberf figurew deduct numeric
            problemr series arithmet;

            USEV ARE visual cubes paper lozenges general
            paragrap sentence wordc wordm wordr numberr object
            figurew;

            USEOBS IS school EQ 0;

ANALYSIS:   ESTIMATOR = ML;
```


Input For Holzinger-Swineford EFA In A CFA Framework Using 13 Variables (Continued)

```
MODEL:
         spatial BY visual-figurew*0       ! start all items at 0
                    lozenges*1             ! start anchor item at 1
                    cubes*1                ! start other large items at 1
                    sentence@0 wordr@0;    ! remove 2 indeterminacies

         memory  BY visual-figurew*0       ! start all items at 0
                    wordr*1                ! start anchor item at 1
                    object*1               ! start other large items at 1
                    lozenges@0 sentence@0; ! remove 2 indeterminacies

         verbal  BY visual-figurew*0       ! start all items at 0
                    sentence*1             ! start anchor item at 1
                    wordm*1                ! start other large items at 1
                    lozenges@0 wordr@0;    ! remove 2 indeterminacies

         spatial-verbal@1;                 ! remove 3 indeterminacies

OUTPUT:  STANDARDIZED MODINDICES(3.84) SAMPSTAT FSDETERMINACY;
```

Output Excerpts Holzinger-Swineford EFA In A CFA Framework Using 13 Variables

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                          39.028
             Degrees of Freedom                 42
             P-Value                        0.6022
CFI/TLI
             CFI                             1.000
             TLI                             1.009
RMSEA (Root Mean Square Error Of Approximation)
             Estimate                        0.000
             90 Percent C.I.                 0.000  0.050
             Probability RMSEA <= .05        0.949
SRMR (Standardized Root Mean Square Residual)
             Value                           0.028
```

Factor Determinacies

```
SPATIAL      0.869
MEMORY       0.841
VERBAL       0.948
```


Output Excerpts Holzinger-Swineford EFA In A CFA Framework Using 13 Variables (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
SPATIAL BY					
VISUAL	**3.933**	**0.811**	**4.848**	**3.933**	**0.571**
CUBES	**2.584**	**0.559**	**4.620**	**2.584**	**0.583**
PAPER	**1.216**	**0.327**	**3.717**	**1.216**	**0.432**
LOZENGES	**6.173**	**0.765**	**8.071**	**6.173**	**0.745**
***GENERAL**	**2.278**	**1.060**	**2.149**	**2.278**	**0.196**
PARAGRAP	0.212	0.307	0.692	0.212	0.063
SENTENCE	0.000	0.000	0.000	0.000	0.000
WORDC	0.994	0.526	1.889	0.994	0.186
WORDM	0.554	0.710	0.780	0.554	0.070
WORDR	0.000	0.000	0.000	0.000	0.000
NUMBERR	0.956	1.019	0.938	0.956	0.127
OBJECT	-0.439	0.663	-0.661	-0.439	-0.096
FIGUREW	0.350	0.441	0.793	0.350	0.098

*Note that theory predicts that GENERAL loads on VERBAL only.

Output Excerpts Holzinger-Swineford EFA In A CFA Framework Using 13 Variables (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
MEMORY BY					
VISUAL	0.580	0.808	0.718	0.580	0.084
CUBES	-0.398	0.558	-0.712	-0.398	-0.090
PAPER	0.374	0.333	1.123	0.374	0.133
LOZENGES	0.000	0.000	0.000	0.000	0.000
GENERAL	-0.100	1.103	-0.091	-0.100	-0.009
PARAGRAP	0.318	0.309	1.030	0.318	0.094
SENTENCE	0.000	0.000	0.000	0.000	0.000
WORDC	0.436	0.540	0.808	0.436	0.082
WORDM	0.425	0.720	0.590	0.425	0.054
WORDR	**6.541**	**1.058**	**6.180**	**6.541**	**0.606**
NUMBERR	**4.291**	**0.977**	**4.392**	**4.291**	**0.571**
OBJECT	**3.040**	**0.646**	**4.704**	**3.040**	**0.668**
FIGUREW	**1.264**	**0.433**	**2.923**	**1.264**	**0.353**

Output Excerpts Holzinger-Swineford EFA In A CFA Framework Using 13 Variables (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
VERBAL BY					
VISUAL	0.265	0.811	0.326	0.265	0.038
CUBES	-0.129	0.546	-0.236	-0.129	-0.029
PAPER	0.096	0.327	0.294	0.096	0.034
LOZENGES	0.000	0.000	0.000	0.00	0.000
GENERAL	**8.130**	**1.058**	**7.682**	**8.130**	**0.700**
PARAGRAP	**2.501**	**0.303**	**8.264**	**2.501**	**0.744**
SENTENCE	**3.954**	**0.322**	**12.263**	**3.954**	**0.853**
WORDC	**2.927**	**0.517**	**5.656**	**2.927**	**0.548**
WORDM	**6.191**	**0.707**	**8.751**	**6.191**	**0.782**
WORDR	0.000	0.000	0.000	0.000	0.000
NUMBERR	-0.870	1.033	-0.842	-0.870	-0.116
OBJECT	0.139	0.653	0.212	0.139	0.030
FIGUREW	0.247	0.433	0.570	0.247	0.069

Output Excerpts Holzinger-Swineford EFA In A CFA Framework Using 13 Variables (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
VERBAL WITH					
SPATIAL	0.467	0.119	3.937	0.467	0.467
MEMORY WITH					
SPATIAL	0.371	0.171	2.173	0.371	0.371
VERBAL	0.459	0.144	3.181	0.459	0.459
Variances					
SPATIAL	1.000	0.000	0.000	1.000	1.000
VERBAL	1.000	0.000	0.000	1.000	1.000
MEMORY	1.000	0.000	0.000	1.000	1.000

Output Excerpts Holzinger-Swineford EFA In A CFA Framework Using 13 Variables (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
Residual Variances					
VISUAL	28.758	4.325	6.649	28.758	0.606
CUBES	13.795	2.049	6.732	13.795	0.703
PAPER	5.801	0.761	7.619	5.801	0.734
LOZENGES	30.640	7.063	4.338	30.640	0.446
GENERAL	47.239	6.824	6.923	47.239	0.350
PARAGRAP	3.637	0.544	6.684	3.637	0.321
SENTENCE	5.831	1.042	5.598	5.831	0.272
WORDC	14.547	1.864	7.803	14.547	0.510
WORDM	18.122	2.878	6.298	18.122	0.289
WORDR	73.589	12.422	5.924	73.589	0.632
NUMBERR	37.595	5.998	6.268	37.595	0.665
OBJECT	11.939	2.377	5.022	11.939	0.576
FIGUREW	10.368	1.319	7.860	10.368	0.807

Output Excerpts Holzinger-Swineford EFA In A CFA Framework Using 13 Variables (Continued)

R-Square

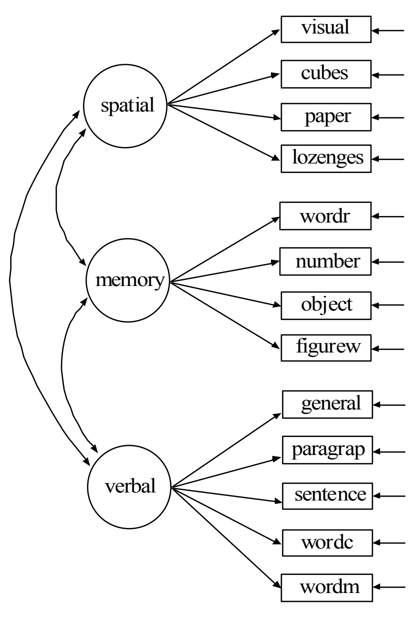
```
Observed          R-Square    ! R-Square =
Variable                      ! 1 – STDYX (residual) = Reliability
                              ! when no covariates are in the model
VISUAL               0.394
CUBES                0.297
PAPER                0.266
LOZENGES             0.554
GENERAL              0.650
PARAGRAP             0.679
SENTENCE             0.728
WORDC                0.490
WORDM                0.711
WORDR                0.368
NUMBERR              0.335
OBJECT               0.424
FIGUREW              0.193
```


Output Excerpts Holzinger-Swineford EFA In A CFA Framework Using 13 Variables (Continued)

Model Modification Indices

```
                        M.I.   E.P.C.  Std E.P.C.   StdYX E.P.C.

WITH Statements

WORDC   WITH SENTENCE   6.586   2.657      2.657     0.107
WORDM   WITH GENERAL    7.121   9.555      9.555     0.104
WORDM   WITH SENTENCE   6.557  -4.238     -4.238    -0.116
```

Simple Structure CFA



Input Excerpts For Holzinger-Swineford Simple Structure CFA Using 13 Variables

```
MODEL:       spatial BY visual-lozenges;
             memory BY wordr-figurew;
             verbal BY general-wordm;

OUTPUT:      STANDARDIZED MODINDICES(3.84) SAMPSTAT FSDETERMINACY;
```


Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                               56.254
             Degrees of Freedom                      62
             P-Value                             0.6817
CFI/TLI
             CFI                                  1.000
             TLI                                  1.012
RMSEA (Root Mean Square Error Of Approximation)
             Estimate                             0.000
             90 Percent C.I.                      0.000  0.041
             Probability RMSEA <= .05             0.983
SRMR (Standardized Root Mean Square Residual)
             Value                                0.046
```

Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables (Continued)

Note: Model fit is better than with the EFA in a CFA framework (p= .6022). This is because the parameters that were fixed to zero were not significant. Thus the gain in degrees of freedom resulted in a higher p-value.

The chi-square difference test between the EFA in a CFA framework and the Simple Structure CFA models is not significant: Chi-square value of 17.23 with 20 degrees of freedom.

Factor Determinacies

SPATIAL	0.867
MEMORY	0.835
VERBAL	0.954

Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
SPATIAL BY					
VISUAL	1.000	.000	.000	4.539	.659
CUBES	.481	.102	4.691	2.182	.492
PAPER	.329	.066	4.975	1.491	.530
LOZENGES	1.303	.219	5.941	5.915	.714
MEMORY BY					
WORDR	1.000	.000	.000	6.527	.605
NUMBERR	.642	.142	4.534	4.191	.557
OBJECT	.435	.091	4.776	2.840	.624
FIGUREW	.247	.063	3.937	1.613	.450
VERBAL BY					
GENERAL	1.000	.000	.000	9.363	.806
PARAGRAP	.295	.027	11.077	2.766	.822
SENTENCE	.413	.037	11.294	3.866	.834
WORDC	.394	.044	8.857	3.688	.691
WORDM	.716	.062	11.513	6.707	.847

Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
VERBAL WITH					
SPATIAL	25.118	5.700	4.407	.591	.591
MEMORY WITH					
SPATIAL	13.323	4.329	3.077	.450	.450
VERBAL	31.883	8.340	3.823	.522	.522
Variances					
SPATIAL	20.597	5.450	3.779	1.000	1.000
VERBAL	87.646	15.363	5.705	1.000	1.000
MEMORY	42.606	13.205	3.226	1.000	1.000

Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables (Continued)

R-Square

VISUAL	0.434
CUBES	0.243
PAPER	0.281
LOZENGES	0.509
GENERAL	0.650
PARAGRAP	0.676
SENTENCE	0.696
WORDC	0.477
WORDM	0.717
WORDR	0.366
NUMBERR	0.311
OBJECT	0.389
FIGUREW	0.203

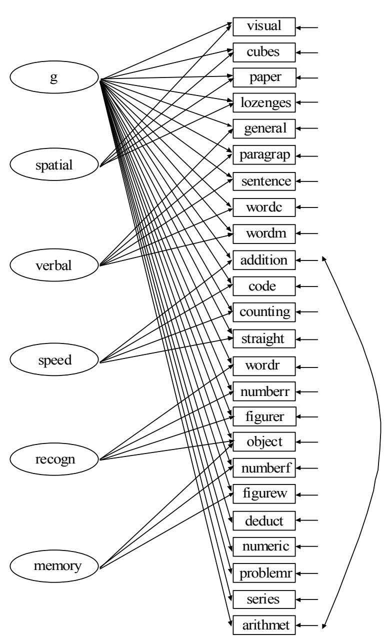
Output Excerpts Holzinger-Swineford Simple Structure CFA Using 13 Variables (Continued)

Model Modification Indices

```
                            M.I.    E.P.C.  Std E.P.C.   StdYX E.P.C.
 WITH Statements

 PARAGRAP WITH GENERAL   4.170    -3.108    -3.108       -0.080
 WORDC    WITH SENTENCE  4.586     2.207     2.207        0.089
 WORDM    WITH GENERAL   4.552     7.582     7.582        0.082
```


Special Factor Analysis Models

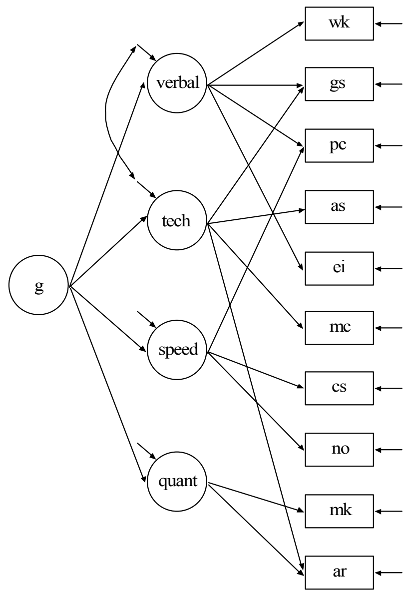
Bi-Factor Model
(Hierarchical Factor Model)



Input Excerpts Holzinger-Swineford General-Specific (Bi-Factor) Factor Model

```
MODEL:      g BY visual-arithmet;
            spatial BY visual-lozenges;
            verbal BY general-wordm;
            speed BY addition-straight;
            recogn BY wordr-object;
            memory BY numberf object figurew;

!           uncorrelated factors because of the general factor:

            g WITH spatial-memory @0;
            spatial WITH verbal-memory @0;
            verbal WITH speed-memory @0;
            speed WITH recogn-memory @0;
            recogn WITH memory @0;

!           correlated residual ("doublet factor"):

            addition WITH arithmet;

OUTPUT:     STANDARDIZED MODINDICES(3.84) SAMPSTAT FSDTERMINACY;
```


Second-Order Factor Model



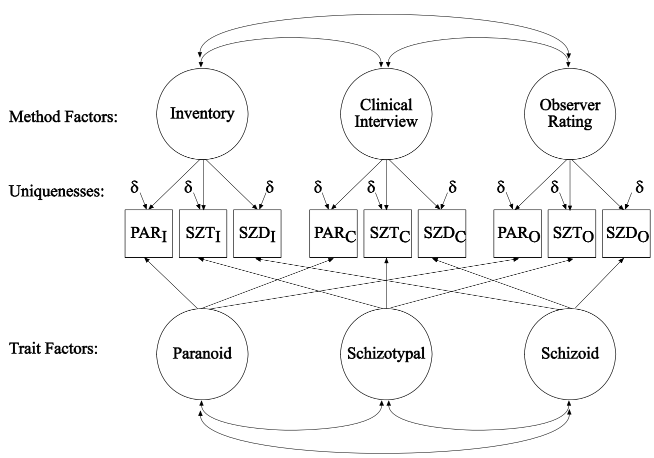
Input For Second-Order Factor Analysis Model

```
TITLE:      Second-order factor analysis model
DATA:       FILE IS asvab.dat;
            !  Armed services vocational aptitude battery
            NOBSERVATIONS = 20422;
            TYPE=COVARIANCE;
VARIABLE:   NAMES ARE ar wk pc mk gs no cs as mc ei;
            USEV = wk gs pc as ei mc cs no mk ar;

            !WK   Word Knowledge
            !GS   General Science
            !PC   Paragraph Comprehension
            !AS   Auto and Shop Information
            !EI   Electronics information
            !MC   Mechanical Comprehension
            !CS   Coding Speed
            !NO   Numerical Operations
            !MK   Mathematical Knowledge
            !AR   Arithmetic Reasoning

ANALYSIS:   ESTIMATOR = ML;
```

Input For Second-Order Factor Analysis Model (Continued)

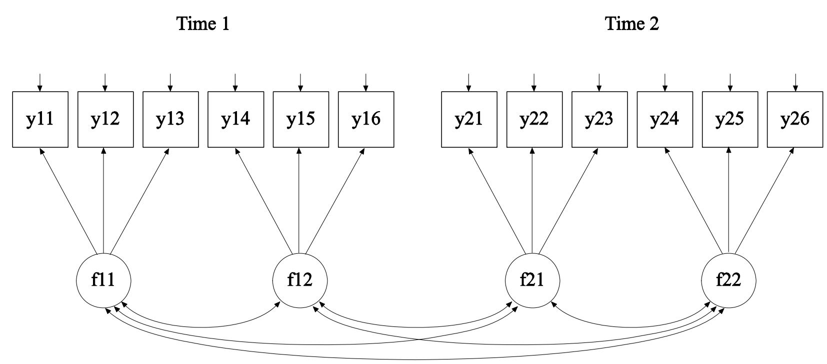
```
MODEL:      verbal BY wk gs pc ei;
            tech BY gs mc ar;
            speed BY pc cs no;
            quant BY mk ar;
            g BY verbal tech speed quant;
            tech WITH verbal;

OUTPUT:     SAMPSTAT MOD(0) STAND TECH1 RESIDUAL;
```


Multi-Trait, Multi-Method (MTMM) Model



Source: Brown (2006)

Longitudinal Factor Analysis Model



Further Readings On CFA

Bollen, K.A. (1989). Structural equations with latent variables. New York: John Wiley.

Brown, T.A. (2006). Confirmatory factor analysis for applied researchers. New York: The Guilford Press.

Joreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34, 183-202.

Lawley, D.N. & Maxwell, A.E. (1971). Factor analysis as a statistical method. London: Butterworths.

Long, S. (1983). Confirmatory factor analysis. Sage University Paper series on Quantitative Applications in the Social Sciences, No 33. Beverly Hills, CA: Sage.

Mulaik, S. (1972). The foundations of factor analysis. McGraw-Hill.

Further Readings On CFA

Mulaik, S.A. (2009). Linear causal modeling with structural equations. Boca Raton, FL: Chapman & Hall.

Personality and Individual Differences, 42 (5), May 1997: Special issue on SEM model testing.

Yung Y.F, Thissen, D. & McLeod, L.D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. Psychometrika, 64, 113-128.

Measurement Invariance And Population Heterogeneity

Models To Study Measurement Invariance And Population Heterogeneity

To further study a set of factors or latent variables established by an EFA/CFA, questions can be asked about the invariance of the measures and the heterogeneity of populations.

Measurement Invariance – Does the factor model hold in other populations or at other time points?

- Same number of factors
- Zero loadings in the same positions
- Equality of factor loadings
- Equality of intercepts
  - Test difficulty

Population Heterogeneity – Are the factor means, variances, and covariances the same for different populations?

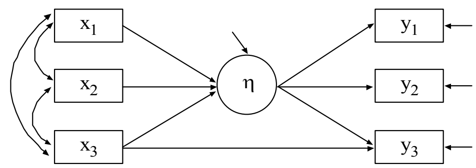
Models To Study Measurement Invariance And Population Heterogeneity (Continued)

Models To Study Measurement Invariance and Population Heterogeneity

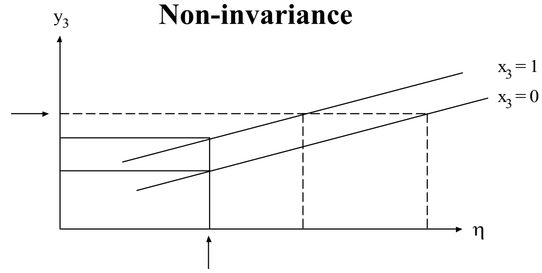
- CFA with covariates
  - Parsimonious
  - Small sample advantage
  - Advantageous with many groups
- Multiple group analysis
  - More parameters to represent non-invariance
    - Factor loadings and observed residual variances/covariances in addition to intercepts
    - Factor variances and covariances in addition to means
  - Interactions

CFA With Covariates

Non-invariance



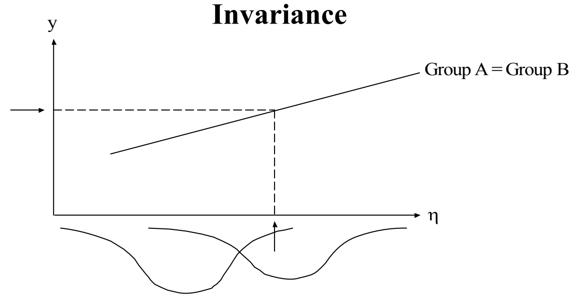
Non-invariance



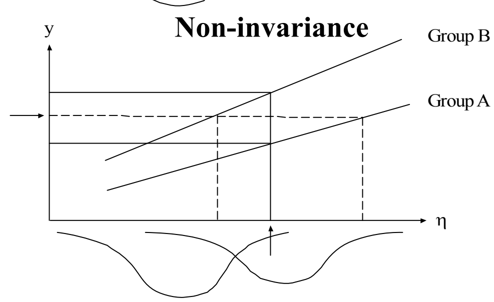
Conditional on η, y is different for the two groups

Multiple Group Analysis

Invariance



Non-invariance



CFA With Covariates (MIMIC)

CFA With Covariates (MIMIC)

Used to study the effects of covariates or background variables on the factors and outcome variables to understand measurement invariance and heterogeneity

- Measurement non-invariance – direct relationships between the covariates and factor indicators that are not mediated by the factors – if they are significant, this indicates measurement non-invariance due to differential item functioning (DIF)

- Population Heterogeneity – relationships between the covariates and the factors – if they are significant, this indicates that the factor means are different for different levels of the covariates.

CFA With Covariates (MIMIC) (Continued)

Model Assumptions

- Same factor loadings and observed residual variances / covariances for all levels of the covariates
- Same factor variances and covariances for all levels of the covariates

Model identification, estimation, testing, and modification are the same as for CFA.

Steps In CFA With Covariates

- Establish a CFA or EFA/CFA model
- Add covariates – check that factor structure does not change and study modification indices for possible direct effects
- Add direct effects suggested by modification indices – check that factor structure does not change
- Interpret the model
  - Factors
  - Effects of covariates on factors
  - Direct effects of covariates on factor indicators

NELS Data

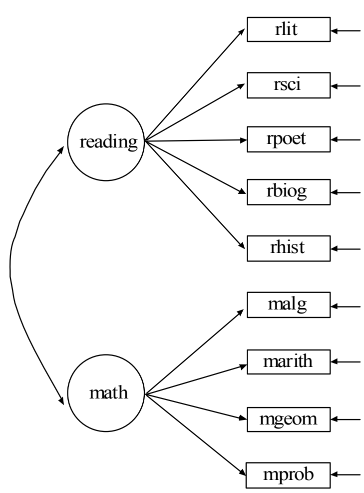
The NELS data consist of 16 testlets developed to measure the achievement areas of reading, math, science, and other school subjects. The sample consists of 4,154 eighth graders from urban, public schools.

Data for the analysis include five reading testlets and four math testlets. The entire sample is used.

Variables

rlit – reading literature
rsci – reading science
rpoet – reading poetry
rbiog – reading biography
rhist – reading history

malg – math algebra
marith – math arithmetic
mgeom – math geometry
mprob – math probability



Input For NELS CFA

```
TITLE:     CFA using NELS data
DATA:      FILE IS ft21.dat;
VARIABLE:  NAMES ARE ses rlit rsci rpoet rbiog rhist malg
           marith mgeom mprob searth schem slife smeth hgeog
           hcit hhist gender schoolid minorc;

           USEVARIABLES ARE rlit-mprob;

MODEL:     reading BY rlit-rhist;
           math BY malg-mprob;

OUTPUT:    STANDARDIZED MODINDICES;
```


Output Excerpts NELS CFA

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                          128.872
             Degrees of Freedom                  26
             P-Value                         0.0000
CFI/TLI
             CFI                              0.993
             TLI                              0.990
RMSEA (Root Mean Square Error Of Approximation)
             Estimate                         0.031
             90 Percent C.I.                  0.026  0.036
             Probability RMSEA <= .05         1.000
SRMR (Standardized Root Mean Square Residual)
             Value                            0.016
```

Output Excerpts NELS CFA (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
READING BY					
RLIT	1.000	.000	.000	.845	.657
RSCI	1.383	.038	36.451	1.168	.672
RPOET	1.130	.030	37.558	.955	.698
RBIOG	1.300	.034	37.791	1.098	.703
RHIST	1.287	.037	34.436	1.087	.627
MATH BY					
MALG	1.000	.000	.000	1.018	.868
MARITH	1.026	.015	69.297	1.045	.890
MGEOM	.655	.020	32.637	.667	.494
MPROB	1.066	.028	38.300	1.086	.565
MATH WITH					
READING	.723	.024	30.067	.840	.840

Output Excerpts NELS CFA (Continued)

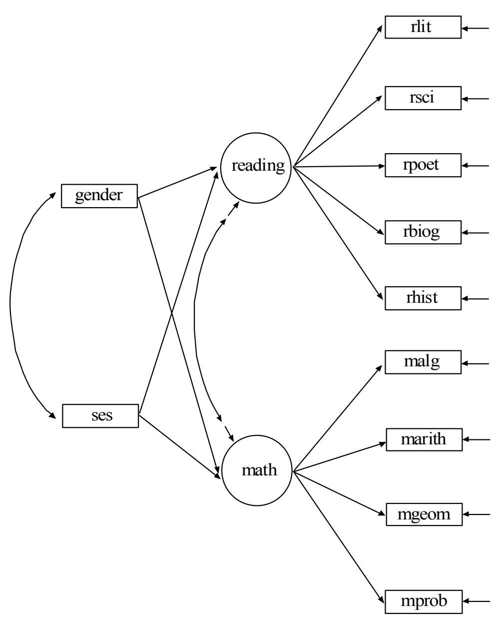
Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
Residual Variances					
RLIT	.939	.024	39.516	.939	.568
RSCI	1.657	.042	39.000	1.657	.548
RPOET	.962	.025	37.986	.962	.513
RBIOG	1.234	.033	37.745	1.234	.506
RHIST	1.822	.045	40.416	1.822	.606
MALG	.339	.012	27.759	.339	.246
MARITH	.285	.012	24.067	.285	.207
MGEOM	1.379	.031	43.922	1.379	.756
MPROB	2.518	.058	43.165	2.518	.681
Variances					
READING	.714	.032	22.231	1.000	1.000
MATH	1.037	.031	33.659	1.000	1.000

Output Excerpts NELS CFA (Continued)

R-Square

RLIT	.432
RSCI	.452
RPOET	.487
RBIOG	.494
RHIST	.394
MALG	.754
MARITH	.793
MGEOM	.244
MPROB	.319



Input For NELS CFA With Covariates

```
TITLE:     CFA with covariates using NELS data

DATA:      FILE IS ft21.dat;

VARIABLE:  NAMES ARE ses rlit rsci rpoet rbiog rhist malg
           marith mgeom mprob searth schem slife smeth hgeog
           hcit hhist gender schoolid minorc;

           USEVARIABLES ARE rlit-mprob ses gender;

MODEL:     reading BY rlit-rhist;
           math BY malg-mprob;

           reading math ON ses gender;   ! female = 0, male = 1

OUTPUT:    STANDARDIZED MODINDICES (3.84);
```


Output Excerpts NELS CFA With Covariates

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                              202.935
             Degrees of Freedom                      40
             P-Value                             0.0000
CFI/TLI
             CFI                                  0.990
             TLI                                  0.986
RMSEA (Root Mean Square Error Of Approximation)
             Estimate                             0.031
             90 Percent C.I.                      0.027  0.036
             Probability RMSEA <= .05             1.000
SRMR (Standardized Root Mean Square Residual)
             Value                                0.018
```

Output Excerpts NELS CFA With Covariates

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
READING BY					
RLIT	1.000	.000	.000	.846	.658
RSCI	1.370	.038	36.437	1.159	.667
RPOET	1.133	.030	37.907	.959	.700
RBIOG	1.296	.034	37.998	1.097	.702
RHIST	1.291	.037	34.758	1.092	.630
MATH BY					
MALG	1.000	.000	.000	1.015	.866
MARITH	1.031	.015	70.136	1.047	.892
MGEOM	.659	.020	32.794	.669	.495
MPROB	1.071	.028	38.435	1.088	.566

Output Excerpts NELS CFA With Covariates (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
READING ON					
SES	**.344**	**.014**	**24.858**	**.407**	**.438**
GENDER	**-.186**	**.027**	**-6.901**	**-.220**	**-.110**
MATH ON					
SES	**.418**	**.015**	**28.790**	**.412**	**.444**
GENDER	**.044**	**.030**	**1.457**	**.044**	**022**
MATH WITH					
READING	.558	.019	29.142	.649	.649

Output Excerpts NELS CFA With Covariates (Continued)

Residual Variances

RLIT	.937	.024	39.695	.937	.567
RSCI	1.679	.043	39.407	1.679	.555
RPOET	.955	.025	38.136	.955	.510
RBIOG	1.237	.033	38.046	1.237	.507
RHIST	1.812	.045	40.521	1.812	.603
MALG	.345	.012	28.752	.345	.251
MARITH	.281	.012	24.388	.281	.204
MGEOM	1.377	.031	43.946	1.377	.754
MPROB	2.513	.058	43.207	2.513	.680
READING	.572	.026	21.920	.799	.799
MATH	.826	.025	32.943	.801	.801

Output Excerpts NELS CFA With Covariates (Continued)

R-Square

RLIT	.433
RSCI	.445
RPOET	.490
RBIOG	.493
RHIST	.397
MALG	.749
MARITH	.796
MGEOM	.246
MPROB	.320

Latent Variable	R-Square
READING	.201
MATH	.199

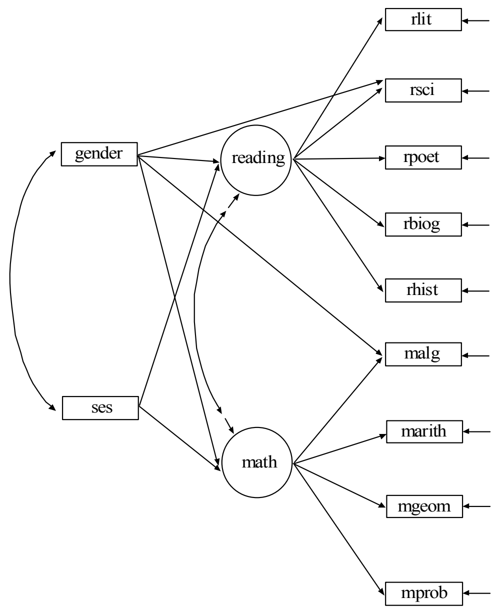
Input For Modification Indices For Direct Effects NELS CFA With Covariates

```
TITLE:      Modification indices for direct effects
            CFA with covariates using NELS data

DATA:       FILE IS ft21.dat;

VARIABLE:   NAMES ARE ses rlit rsci rpoet rbiog rhist malg
            marith mgeom mprob search schem slife smeth hgeog
            hcit hhist gender schoolid minorc;

            USEVARIABLES ARE rlit-mprob ses gender;

MODEL:      reading BY rlit-rhist;
            math BY malg-mprob;

            reading math ON ses gender;   !female = 0, male = 1

            rlit-mprob ON ses-gender@0;

OUTPUT:     STANDARDIZED MODINDICES(3.84);
```


Output Excerpts Modification Indices For Direct Effects NELS CFA With Covariates

Modification Indices

		M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
RSCI	**ON GENDER**	**31.730**	**0.253**	**0.253**	**0.073**
RPOET	ON GENDER	12.715	-0.124	-0.124	-0.045
RHIST	ON SES	6.579	0.062	0.062	0.038
MALG	**ON GENDER**	**26.616**	**-0.120**	**-0.120**	**-0.051**
MARITH	ON GENDER	10.083	0.075	0.075	0.032
MGEON	ON SES	4.201	0.040	0.040	0.032
MPROB	ON GENDER	7.922	0.143	0.143	0.037



Summary Of Analysis Results For NELS CFA With Covariates And Direct Effects

Model	Chi-square (d.f.)	Difference (d.f. diff.)
No direct effects	202.935 (40)	
rsci ON gender	171.006 (39)	31.929* (1)
rsci ON gender and malg ON gender	144.728 (38)	26.728* (1)

Input For NELS CFA With Covariates And Two Direct Effects

```
TITLE:      CFA with covariates and two direct effects using
            NELS data

DATA:       FILE IS ft21.dat;

VARIABLE:   NAMES ARE ses rlit rsci rpoet rbiog rhist malg
            marith mgeom mprob search schem slife smeth hgeog
            hcit hhist gender schoolid minorc;

            USEVARIABLES ARE rlit-mprob ses gender;

MODEL:      reading BY rlit-rhist;
            math BY malg-mprob;

            reading math ON  ses gender;   !female = 0, male = 1

            rsci ON gender;
            malg ON gender;

OUTPUT:     STANDARDIZED MODINDICES(3.84);
```


Output Excerpts NELS CFA With Covariates And Two Direct Effects

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                              144.278
             Degrees of Freedom                      38
             P-Value                             0.0000
CFI/TLI
             CFI                                  0.993
             TLI                                  0.991
RMSEA (Root Mean Square Error Of Approximation)
             Estimate                             0.026
             90 Percent C.I.                      0.022  0.031
             Probability RMSEA <= .05             1.000
SRMR (Standardized Root Mean Square Residual)
             Value                                0.014
```

Output Excerpts NELS CFA With Covariates And Two Direct Effects (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
READING BY					
RLIT	1.000	.000	.000	.846	.658
RSCI	1.389	.038	36.609	1.175	.676
RPOET	1.133	.030	37.958	.959	.701
RBIOG	1.294	.034	37.991	1.095	.701
RHIST	1.290	.037	34.760	1.091	.630
MATH BY					
MALG	1.000	.000	.000	1.019	.869
MARITH	1.027	.015	70.300	1.047	.892
MGEOM	.657	.020	32.833	.670	.496
MPROB	1.068	.028	38.524	1.089	.566

Output Excerpts NELS CFA With Covariates And Two Direct Effects (Continued)

Model Results

	Estimates	S.E.	Est./S.E.	Std	StdYX
READING ON					
SES	.343	.014	24.854	.406	.437
GENDER	-.222	.028	-7.983	-.262	-.131
MATH ON					
SES	.419	.015	28.807	.411	.444
GENDER	**.092**	**.032**	**2.873**	**.090**	**.045**
RSCI ON					
GENDER	**.254**	**.045**	**5.649**	**.254**	**.073**
MALG ON					
GENDER	**-.121**	**.023**	**-5.171**	**-.121**	**-.051**

Interpretation Of Direct Effects

Rsci On Gender

- Indirect effect of gender on rsci
  - Reading factor has a negative relationship with gender – males have a lower mean than females on the reading factor
  - Rsci has a positive loading on the reading factor
  - Conclusion: Males are expected to have a lower mean on rsci

- Direct effect of gender on rsci
  - Direct effect is positive – for a given reading factor value, males do better than expected on rsci
  - Conclusion – rsci is not invariant. Males may have had more exposure to science reading.

Interpretation Of Direct Effects (Continued)

Malg On Gender

- Indirect effect of gender on malg
  - Math factor has a positive relationship with gender – males have a higher mean than females in math
  - Malg has a positive loading on the math factor
  - Conclusion: Males are expected to have a higher mean on malg
- Direct effect of gender on malg
  - Direct effect is negative – for a given math factor value, males do worse than expected on malg
  - Conclusion: malg is not invariant

Multiple Group Analysis

Multiple Group Analysis

Used to study group differences in measurement and structural parameters by simultaneous analysis of several groups of individuals

Advantages Of Multiple Group Analysis Versus Factor Analysis With Covariates

- More parameters to represent non-invariance
  - Factor loadings and observed residual variances/covariances in addition to intercepts
  - Factor variances and covariances in addition to means
- Interactions

Multiple Group Analysis (Continued)

Disadvantages Of Multiple Group Analysis Versus Factor Analysis With Covariates

- Less parsimonious model
- Requires sufficiently large sample size for each group
- Difficult to carry out with many groups

Model Specification

- Comparison of factor variances and covariances meaningful only when factor loadings are invariant
- Comparison of factor means meaningful only when factor loadings and measurement intercepts are invariant
- Partial invariance possible

Model identification, estimation, testing, and modification are the same as for CFA.

Steps In Multiple Group Analysis

- Fit the model separately in each group
- Fit the model in all groups allowing all parameters to be free
- Fit the model in all groups holding factor loadings equal to test the invariance of the factor loadings
- Fit the model in all groups holding factor loadings and intercepts equal to test the invariance of the intercepts
- Add covariates
- Modify the model

Mplus Input For Multiple Group Analysis

- General rules
  - MODEL command is used to describe the overall analysis model for all groups
  - Group-specific MODEL commands are used to specify differences between the overall analysis model and the model for that group
  - Equalities specified in the MODEL command apply across groups
  - Equalities specified in the group-specific MODEL commands apply to only the specific group

Mplus Input For Multiple Group Analysis (Continued)

- Defaults
  - Factor loadings are held equal across the groups
  - All other free parameters are not held equal across groups
  - When means are included in the model
    - Intercepts of observed variables are held equal across group
    - Factor means are fixed at zero in the first group and are free to be estimated in the other groups

Mplus Input For Multiple Group Analysis (Continued)

- Example 1 – factor loading invariance across groups

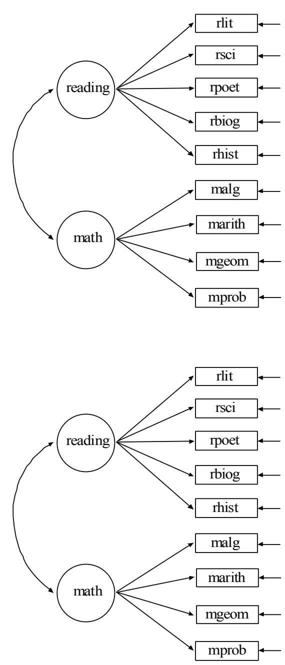
```
MODEL:      f1 BY y1 y2 y3;
            f2 BY y4 y5 y6;
```

- Example 2 – factor loading non-invariance for 2 groups

```
MODEL:      f1 BY y1 y2 y3;
            f2 BY y4 y5 y6;
MODEL g2:   f1 BY y2 y3;
            f2 BY y5 y6;
```


Males

Females



Inputs For NELS Single Group Analyses Without Measurement Invariance

Single Group Analyses

```
TITLE:      Single group CFA for males using NELS data

DATA:       FILE IS ft21.dat;

VARIABLE:   NAMES ARE ses rlit rsci rpoet rbiog rhist malg
            marith mgeom mprob searth schem slife smeth hgeog
            hcit gender schoolid minorc;

            USEVARIABLES ARE rlit-mprob;

            USEOBSERVATIONS ARE (gender EQ 1);  !  change 1 to
                                                !  0 for females
MODEL:      reading BY rlit-rhist;
            math BY malg-mprob;
```


Input For NELS Multiple Group Analysis Without Measurement Invariance

```
TITLE:        Multiple group CFA for males and females using NELS
              data with no measurement invariance

DATA:         FILE IS ft21.dat;

VARIABLE:     NAMES ARE ses rlit rsci rpoet rbiog rhist malg
              marith mgeom mprob searth schem slife smeth hgeog
              hcit gender schoolid minorc;

              GROUPING IS gender (0=female 1=male);

              USEVARIABLES ARE rlit-mprob;

MODEL:        reading BY rlit-rhist;
              math BY malg-mprob;
              [reading@0 math@0];

MODEL male:   reading BY rsci-rhist;
              math BY marith-mprob;
              [rlit-mprob];
```

Summary Of Analysis Results For NELS Single And Multiple Group Analyses Without Measurement Invariance

	Chi-square	RMSEA
Males (n=2048)	72.555 (26) .0000	.030
Females (n=2106)	86.274 (26) .0000	.033
Together (n=4154)	158.829 (52) .0000	.031

Input For NELS Multiple Group Analyses With Measurement Invariance

Invariance Of Factor Loadings

```
TITLE:        Multiple group CFA for males and females using NELS
              data with measurement invariance of factor loadings

DATA:         FILE IS ft21.dat;

VARIABLE:     NAMES ARE ses rlit rsci rpoet rbiog rhist malg
              marith mgeom mprob searth schem slife smeth hgeog
              hcit gender schoolid minorc;
              GROUPING IS gender (0=female 1=male);
              USEVARIABLES ARE rlit-mprob;
MODEL:        Reading BY rlit-rhist;
              Math BY malg-mprob;
              [reading@0  math@0];

MODEL MALE:   [rlit - mprob];

OUTPUT:       STANDARDIZED MODINDICES(3.84);
```

Input For NELS Multiple Group Analyses With Measurement Invariance (Continued)

Invariance Of Factor Loadings And Intercepts

```
TITLE:      Multiple group CFA for males and females using NELS
            data with measurement invariance of factor loadings
            and intercepts

DATA:       FILE IS ft21.dat;

VARIABLE:   NAMES ARE ses rlit rsci rpoet rbiog rhist malg
            marith mgeom mprob searth schem slife smeth hgeog
            hcit gender schoolid minorc;
            GROUPING IS gender (0=female 1=male);
            USEVARIABLES ARE rlit-mprob;

MODEL:      reading BY rlit-rhist;
            math BY malg-mprob;

OUTPUT:     STANDARDIZED MODINDICES(3.84);
```


Summary Of Analysis Results For NELS Single And Multiple Group Analyses With Measurement Invariance

Model	Chi-square	Difference
Measurement non-invariance	158.829 (52)	
Factor loading invariance	170.386 (59)	11.557 (7)
Factor loading and intercept invariance	238.847 (66)	68.461* (7)

Summary Of Analysis Results For NELS Single And Multiple Group Analyses With Measurement Invariance (Continued)

Explanation of Chi-Square Differences

Factor loading invariance	(7)	7 factor loadings instead of 14
Factor loading and intercept invariance	(7)	9 intercepts and 2 factor means instead of 18 intercepts

Summary Of Analysis Results For NELS Single And Multiple Group Analyses With Measurement Invariance (Continued)

Modification Indices (Excerpts)

Group MALE

	M.I.	E.P.C.	Std E.P.C.	StdYX E.P.C.
Means/Intercepts/Thresholds				
[RSCI]	**31.794**	**.154**	**.154**	**.089**
[RPOET]	12.856	-.081	-.081	-.058
[MALG]	**26.574**	**-.085**	**-.085**	**-.071**
[MARITH]	10.084	.056	.056	.047
[MPROB]	7.903	.075	.075	.039

Input Excerpts For NELS Multiple Group Analysis With Partial Measurement Invariance

```
MODEL:       reading BY rlit-rhist;
             math BY malg-mprob;

MODEL male:  [rsci malg];

OUTPUT:      STANDARDIZED  MODINDICES (3.84);
```


Summary Of Analysis Results For NELS Multiple Group Analysis With Partial Measurement Invariance

Model	Chi-square	Difference
Factor loading invariance	170.386 (59)	
Factor loading and partial intercept invariance	180.110 (64)	9.724 (5)

Input Excerpts For NELS Multiple Group Analysis With Partial Measurement Invariance And Invariant Residual Variances

```
MODEL:        reading    BY  rlit-rhist;
              math       BY  malg-mprob;
              rlit-mprob (1-9);

MODEL male:   [rsci malg];

OUTPUT:       STANDARDIZED  MODINDICES (3.84);
```


Summary Of Analysis Results For NELS Multiple Group Analysis With Partial Measurement Invariance And Invariant Residual Variances

Model	Chi-square	Difference
Partial invariance	180.110 (64)	
Item residual invariance	197.513 (73)	17.403 (9)*

Input Excerpts For NELS Multiple Group Analysis With Partial Measurement Invariance And Invariant Factor Variances And Covariance: A Test Of Population Heterogeneity

```
MODEL:        reading BY rlit-rhist;
              math BY malg-mprob;

              reading (1);
              math    (2);
              reading WITH math (3);

MODEL male:   [rsci malg];

OUTPUT:       STANDARDIZED MODINDICES (3.84);
```


Summary Of Analysis Results For NELS Multiple Group Analysis With Partial Measurement Invariance And Invariant Factor Variances And Covariance: A Test Of Population Heterogeneity

Model	Chi-square	Difference
Partial invariance	180.110 (64)	
Invariant factor variances and covariance	183.442 (67)	3.312 (3)

Input Excerpts For NELS Multiple Group Analysis With Partial Measurement Invariance And Invariant Factor Variances, Covariance, And Means: A Test Of Population Heterogeneity

```
MODEL:        reading BY rlit-rhist;
              math BY malg-mprob;
              reading (1);
              math    (2);
              reading WITH math (3);

MODEL male :  [rsci malg reading@0 math@0];

OUTPUT:       STANDARDIZED MODINDICES (3.84);
```


Summary Of Analysis Results For NELS Multiple Group Analysis With Partial Measurement Invariance And Invariant Factor Variances, Covariance, And Means: A Test Of Population Heterogeneity

Model	Chi-square	Difference
Partial invariance	180.110 (64)	
Invariant factor variances and covariance	183.422 (67)	3.312 (3)
Invariant factor variances, covariance, and means	340.498 (69)	157.076 (2)*

Technical Aspects Of Multiple-Group Factor Analysis Modeling

$$\boldsymbol{y}_{ig} = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g\, \boldsymbol{\eta}_{ig} + \boldsymbol{\varepsilon}_{ig}\,, \qquad (21)$$

$$E(\boldsymbol{y}_g) = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g\, \boldsymbol{\alpha}_g\,, \qquad (22)$$

$$V(\boldsymbol{y}_g) = \boldsymbol{\Lambda}_g\, \boldsymbol{\Psi}_g\, \boldsymbol{\Lambda}_g' + \boldsymbol{\Theta}_g\,. \qquad (23)$$

ML estimation with G *independently* observed groups:

$$F_{ML}\,(\boldsymbol{\pi}) = 1/2 \sum_{g=1}^{G} \{n_g[ln \mid \boldsymbol{\Sigma}_g \mid + trace\,(\boldsymbol{\Sigma}_g^{-1}\boldsymbol{T}_g) - ln \mid \boldsymbol{S}_g \mid - p]\}/n, \qquad (24)$$

where n_g is the sample size in group g, $n = \Sigma_g^G n_g$, and

$$\boldsymbol{T}_g = \boldsymbol{S}_g + (\bar{\boldsymbol{y}}_g - \boldsymbol{\mu}_g)(\bar{\boldsymbol{y}}_g - \boldsymbol{\mu}_g)' \qquad (25)$$

(e.g. Jöreskog & Sörbom, 1979; Browne & Arminger, 1995).

225

Technical Aspects Of Multiple-Group Factor Analysis Modeling (Continued)

Two main cases:

- No mean structure
  - Assume $\boldsymbol{\Lambda}$ invariance
  - Study ($\boldsymbol{\Theta}_g$ and) $\boldsymbol{\Psi}_g$ differences
  - ($\boldsymbol{\nu}_g$ free, $\boldsymbol{\alpha} = \boldsymbol{0}$, so that $\widehat{\boldsymbol{\mu}}_g = \bar{\boldsymbol{y}}_g$)

- Mean structure
  - Assume $\boldsymbol{\nu}$ and $\boldsymbol{\Lambda}$ invariance
  - Study ($\boldsymbol{\Theta}_g$ and) $\boldsymbol{\alpha}_g$ and $\boldsymbol{\Psi}_g$ differences ($\boldsymbol{\alpha}_1 = \boldsymbol{0}$)

226

Further Readings On MIMIC And Multiple-Group Analysis

Joreskog, K.G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

Meredith, W. (1964). Notes on factorial invariance. Psychometrika, 29, 177-185.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika, 58, 525-543.

Muthen, B. (1989a). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585. (#24)

Sorbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 27, 229-239.

Structural Equation Modeling (SEM)

Structural Equation Modeling (SEM)

Used to study relationships among multiple outcomes often involving latent variables

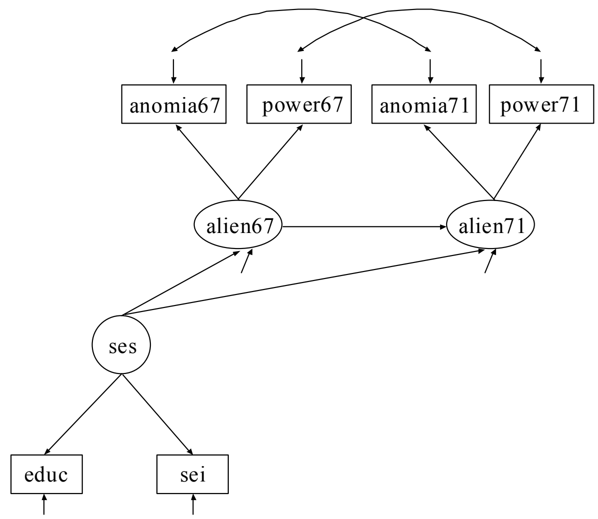
- Estimate and test direct and indirect effects in a system of regression equations for latent variables without the influence of measurement error
- Estimate and test theories about the absence of relationships among latent variables

Model identification, estimation, testing, and modification are the same as for CFA.

Steps In SEM

- Establish a CFA model when latent variables are involved
- Establish a model of the relationships among the observed or latent variables
- Modify the model

Classic Wheaton Et Al. SEM



Input For Classic Wheaton Et Al. SEM

```
TITLE:      Classic structural equation model with multiple
            indicators used in a study of the stability of
            alienation.

DATA:       FILE IS wheacov.dat
            TYPE IS COVARIANCE;
            NOBS ARE 932;

VARIABLE:   NAMES ARE anomia67 power67 anomia71 power71 educ
            sei;

MODEL:      ses       BY educ sei;
            alien67   BY anomia67 power67;
            alien71   BY anomia71 power71;

            alien71   ON alien67 ses;
            alien67   ON ses;

            anomia67  WITH anomia71;
            power67   WITH power71;

OUTPUT:     SAMPSTAT STANDARDIZED MODINDICES (0);
```

Output Excerpts
Classic Wheaton Et Al. SEM

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                              4.771
             Degrees of Freedom                     4
             P-Value                            .3111

RMSEA (Root Mean Square Error Of Approximation)
             Estimate                            .014
             90 Percent C.I.                     .000  .053
             Probability RMSEA <= .05            .928
```


Output Excerpts
Classic Wheaton Et Al. SEM (Continued)

Model Results

```
                  Estimates      S.E.  Est./S.E.     Std     StdYX
 SES        BY
    EDUC              1.000      .000      .000    2.607      .841
    SEI               5.221      .422    12.367   13.612      .642

 ALIEN67    BY
    ANOMIA67          1.000      .000      .000    2.663      .775
    POWER67            .979      .062    15.896    2.606      .852

 ALIEN71    BY
    ANOMIA71          1.000      .000      .000    2.850      .805
    POWER71            .922      .059    15.500    2.627      .832
```

Output Excerpts
Classic Wheaton Et Al. SEM (Continued)

ALIEN71 ON					
ALIEN67	.607	.051	11.895	.567	.567
SES	-.227	.052	-4.337	-.208	-.208
ALIEN67 ON					
SES	-.575	.056	-10.197	-.563	-.563
ANOMIA67 WITH					
ANOMIA71	1.622	.314	5.173	1.622	.133
POWER67 WITH					
POWER71	.340	.261	1.302	.340	.035

Output Excerpts
Classic Wheaton Et Al. SEM (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
Residual Variances					
ANOMIA67	4.730	.453	10.438	4.730	.400
POWER67	2.564	.403	6.362	2.564	.274
ANOMIA71	4.397	.515	8.537	4.397	.351
POWER71	3.072	.434	7.077	3.072	.308
EDUC	2.804	.507	5.532	2.804	.292
SEI	264.532	18.125	14.595	264.532	.588
ALIEN67	4.842	.467	10.359	.683	.683
ALIEN71	4.084	.404	10.104	.503	.503
Variances					
SES	6.796	.649	10.476	1.000	1.000

Output Excerpts Classic Wheaton Et Al. SEM (Continued)

R-Square

Observed Variable	R-Square
ANOMIA67	.600
POWER67	.726
ANOMIA71	.649
POWER71	.692
EDUC	.708
SEI	.412

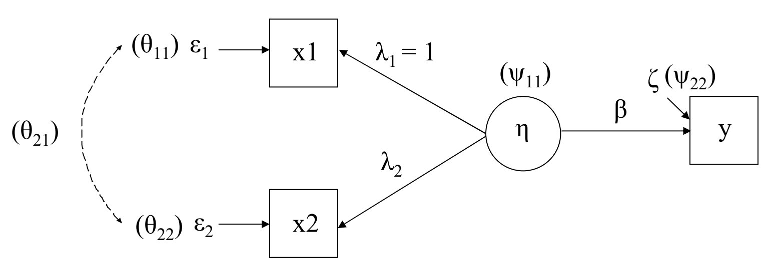
Latent Variable	R-Square
ALIEN67	.317
ALIEN71	.497

Modeling Issues In SEM

- Model building strategies
  - Bottom up
  - Measurement versus structural parts
- Number of indicators
  - Identifiability
  - Robustness to misspecification
- Believability
  - Measures
  - Direction of arrows
  - Other models
- Quality of estimates
  - Parameters, s.e.'s, power
  - Monte Carlo study within the substantive study

Model Identification

Model Identification Issues: A (Simple?) SEM With Measurement Errors In The *x*'s



Model Identification Issues (Continued)

A non-identified parameter gives a non-invertible information matrix (no s.e.s.; indeterminacy involving parameter #...).

A fixed or constrained parameter with a derivative (MI) different from zero would be identified if freed and would improve F.

Example (alcohol consumption, dietary fat intake, blood pressure):

Two indicators of a single latent variable that predicts a later observed outcome (6 parameters; just identified model):

$$x_{ij} = \lambda_j \eta_i + \varepsilon_{ij} \ (j = 1,2), \tag{28}$$

$$y_i = \beta \eta_i + \zeta_i. \tag{29}$$

Model Identification Issues (Continued)

Show identification by solving for the parameters in terms of the $\mathbf{\Sigma}$ elements (fixing $\lambda_1 = 1$):

$V(x_1) = \sigma_{11} = \psi_{11} + \theta_{11}$, (33) $V(x_2) = \sigma_{22} = \lambda_2^2 \psi_{11} + \theta_{22}$, (34)

$Cov(x_2, x_1) = \sigma_{21} = \lambda_2 \psi_{11}$, (35) $Cov(y, x_1) = \sigma_{31} = \beta \psi_{11}$, (36)

$Cov(y, x_2) = \sigma_{32} = \lambda_2 \beta \psi_{11}$, (37) $V(y) = \sigma_{33} = \beta^2 \psi_{11} + \psi_{22}$. (38)

Solving for β:

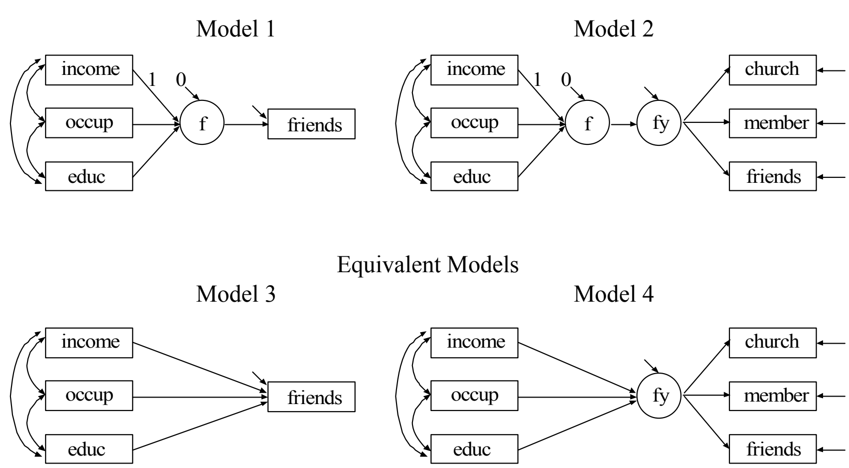
$$\frac{Cov(y, x_2)}{Cov(x_2, x_1)} = \frac{\lambda_2 \beta \psi_{11}}{\lambda_2 \psi_{11}} = \beta$$

With correlated error θ_{21}:

$$\frac{Cov(y, x_2)}{Cov(x_2, x_1)} = \frac{\lambda_2 \beta \psi_{11}}{\lambda_2 \psi_{11} + \theta_{21}} \neq \beta$$

Formative Indicators

Formative Indicators



Hodge-Treiman Social Status Indicators

Social participation related to social status (n = 530 women)

Social participation measures:
- Church membership
- Memberships
- Friends seen

Social status measures:
- Income
- Occupation
- Education

Source: Hodge-Treiman (1968), American Sociological Review

Input For Social Status Formative Indicators, Model 1

```
TITLE:       Hodge-Treiman social status modeling

DATA:        FILE = htmimicn1.dat;
             TYPE = COVARIANCE;
             NOBS = 530;

VARIABLE:    NAMES = church member friends income occup educ;
             USEV = friends-educ;

MODEL:       f BY;  ! defining the formative factor
             f ON income@1 occup educ;
             f@0;
             friends ON f;

OUTPUT:      TECH1 STANDARDIZED;
```

Output Excerpts Social Status Formative Indicators, Model 1

Tests Of Model Fit

```
Chi-Square Test of Model Fit
             Value                              0.000
             Degrees of Freedom                     0
             P-Value                           0.0000
```

Model Results

```
                  Estimates      S.E.  Est./S.E.     Std     StdYX
 F          ON
   INCOME           1.000      0.000      0.000    0.427     0.427
   OCCUP            0.380      0.481      0.790    0.162     0.162
   EDUC             1.640      0.877      1.870    0.700     0.699

 FRIENDS    ON
   F                0.109      0.045      2.410    0.255     0.256

 Residual Variances
   FRIENDS          0.933      0.057     16.279    0.933     0.935
   F                0.000      0.000      0.000    0.000     0.000
```


Input Excerpts Social Status Formative Indicators, Model 2

```
VARIABLE:    NAMES ARE church members friends income occup educ;
             USEV = church-educ;

MODEL:       fy BY church-friends;
             f BY; ! defining the formative factor
             f ON income@1 occup educ;
             f@0;
             fy ON f;
```

Output Excerpts Social Status Formative Indicators, Model 2

Tests Of Model Fit

Chi-Square Test of Model Fit

Value	12.582
Degrees of Freedom	6
P-Value	0.0502

Model Results

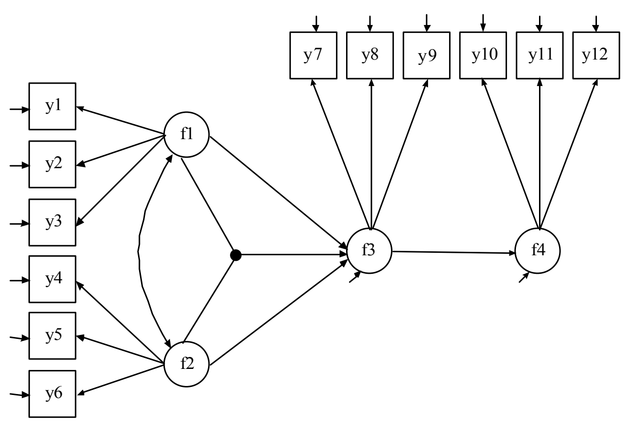
	Estimates	S.E.	Est./S.E.	Std	StdYX
FY BY					
CHURCH	1.000	0.000	0.000	0.466	0.466
MEMBER	1.579	0.235	6.732	0.735	0.736
FRIENDS	0.862	0.143	6.046	0.402	0.402
FY ON					
F	0.108	0.028	3.825	0.508	0.508
F ON					
INCOME	1.000	0.000	0.000	0.457	0.457
OCCUP	0.418	0.276	1.515	0.191	0.191
EDUC	1.438	0.453	3.173	0.658	0.657

Output Excerpts Social Status Formative Indicators, Model 2 (Continued)

	Estimates	S.E.	Est./S.E.	Std	StdYX
Residual Variances					
CHURCH	0.781	0.057	13.620	0.781	0.783
MEMBER	0.457	0.075	6.092	0.457	0.458
FRIENDS	0.837	0.058	14.528	0.837	0.838
FY	0.161	0.037	4.361	0.742	0.742
F	0.000	0.000	0.000	0.000	0.000

Latent Variable Interactions

Structural Equation Model With Interaction Between Latent Variables



Klein & Moosbrugger (2000)
Marsh et al. (2004)

Monte Carlo Simulations

Input Monte Carlo Simulation Study For A CFA With Covariates

```
TITLE:       This is an example of a Monte Carlo simulation study
             for a CFA with covariates (MIMIC) with continuous
             factor indicators and patterns of missing data

MONTECARLO:  NAMES ARE y1-y4 x1 x2;
             NOBSERVATIONS = 500;
             NREPS = 500;
             SEED = 4533;
             CUTPOINTS = x2(1);
             PATMISS = y1(.1) y2(.2) y3(.3) y4(1) |
                       y1(1) y2(.1) y3(.2) y4(.3);
             PATPROBS = .4 | .6;

MODEL POPULATION:
             [x1-x2@0];
             x1-x2@1;
             f BY y1@1 y2-y4*1;
             f*.5;
             y1-y4*.5;
             f ON x1*1 x2*.3;
```

Input Monte Carlo Simulation Study For A CFA With Covariates (Continued)

```
MODEL:        f BY y1@1 y2-y4*1;
              f*.5;
              y1-y4*.5;
              f ON x1*1 x2*.3;
OUTPUT:       TECH9;
```


Output Excerpts Monte Carlo Simulation Study For A CFA With Covariates

Tests Of Model Fit

```
Number of Free Parameters                          14

Chi-Square Test of Model Fit

              Degrees of Freedom                    8

              Mean                              8.297
              Std Dev                           4.122
              Number of successful computations   500
```

Output Excerpts Monte Carlo Simulation Study For A CFA With Covariates (Continued)

Proportions		Percentiles	
Expected	Observed	Expected	Observed
0.990	0.996	1.646	2.008
0.980	0.990	2.032	2.597
0.950	0.940	2.733	2.592
0.900	0.896	3.490	3.441
0.800	0.814	4.594	4.711
0.700	0.706	5.527	5.605
0.500	0.542	7.344	7.663
0.300	0.326	9.524	9.993
0.200	0.238	11.030	11.726
0.100	0.120	13.362	14.313
0.050	0.052	15.507	15.575
0.020	0.016	18.168	17.986
0.010	0.006	20.090	19.268

Output Excerpts Monte Carlo Simulation Study For A CFA With Covariates (Continued)

Model Results

	ESTIMATES			S.E.	M. S. E.	95%	%Sig
	Population	Average	Std. Dev.	Average		Cover	Coeff
F BY							
Y1	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Y2	1.000	1.0083	0.0878	0.0847	0.0078	0.932	1.000
Y3	1.000	1.0035	0.0859	0.0801	0.0074	0.938	1.000
Y4	1.000	1.0032	0.0637	0.0654	0.0041	0.954	1.000
F ON							
X1	1.000	0.9990	0.0630	0.0593	0.0040	0.936	1.000
X2	0.300	0.3029	0.1083	0.1056	0.0117	0.954	0.806

MODEL CONSTRAINT

The MODEL CONSTRAINT Command

```
MODEL:
    f1 BY y1
    y2-y3 (lam2-lam3);
    f2 BY y4
    y5-y6 (lam5-lam6);
    f1-f2 (vf1-vf2);
    y1-y6 (ve1-ve6);

MODEL CONSTRAINT:
    NEW(rel2 rel5 stan3 stan6);
    rel2 = lam2**2*vf1/(lam2**2*vf1 + ve2);
    rel5 = lam5**2*vf2/(lam5**2*vf2 + ve5);
    rel5 = rel2;
    stan3 = lam3*sqrt(vf1)/sqrt(lam3**2*vf1 + ve3);
    stan6 = lam6*sqrt(vf2)/sqrt(lam6**2*vf2 + ve6);
```

The MODEL CONSTRAINT Command (Continued)

- New parameters
- 0 = parameter function
- Inequalities
- Constraints involving observed variables

MODEL TEST

- Wald chi-square test of restrictions on parameters
- Restrictions not imposed by the model (unlike MODEL CONSTRAINT)
- Can use labels from the MODEL command and the MODEL CONSTRAINT command

Example: Testing equality of loadings

```
MODEL:
f BY y1-y3* (p1-p3);
f@1;
MODEL TEST:
p2 = p1;
p3 = p1;
```

Technical Aspects Of Structural Equation Modeling

General model formulation for G groups

$$\boldsymbol{y}_{ig} = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g \boldsymbol{\eta}_{ig} + \mathbf{K}_g \mathbf{x}_{ig} + \boldsymbol{\varepsilon}_{ig}, \tag{26}$$

$$\boldsymbol{\eta}_{ig} = \boldsymbol{\alpha}_g + \mathbf{B}_g \boldsymbol{\eta}_{ig} + \boldsymbol{\Gamma}_g \mathbf{x}_{ig} + \boldsymbol{\zeta}_{ig}, \tag{27}$$

The covariance matrices $\boldsymbol{\Theta}_g = V(\varepsilon_{ig})$ and $\boldsymbol{\Psi}_g = V(\zeta_{ig})$ are also allowed to vary across the G groups.

Further Readings On SEM

Bollen, K.A. (1989). Structural equations with latent variables. New York: John Wiley.

Browne, M.W. & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C.C. Clogg & M.E. Sobel (Eds.), Handbook of statistical modeling for the social and behavioral sciences (pp. 311-359). New York: Plenum Press.

Joreskog, K.G., & Sorbom, D. (1979). Advances in factor analysis and structural equation models. Cambridge, MA: Abt Books.

Muthen, B. & Muthen, L. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling, 4, 599-620.

Skrondal, A. & Rabe-Hesketh, S. (2004). Generalized latent variable modeling. Multilevel, longitudinal, and structural equation models. London: Chapman Hall.

References

Regression Analysis

Agresti, A. & Finlay, B. (1997). Statistical methods for the social sciences. Third edition. New Jersey: Prentice Hall.

Amemiya, T. (1985). Advanced econometrics. Cambridge, Mass.: Harvard University Press.

Hamilton, L.C. (1992). Regression with graphics. Belmont, CA: Wadsworth.

Johnston, J. (1984). Econometric methods. Third edition. New York: McGraw-Hill.

Lewis-Beck, M.S. (1980). Applied regression: An introduction. Newbury Park, CA: Sage Publications.

Moore, D.S. & McCabe, G.P. (1999). Introduction to the practice of statistics. Third edition. New York: W.H. Freeman and Company.

Pedhazur, E.J. (1997). Multiple regression in behavioral research. Third Edition. New York: Harcourt Brace College Publishers.

References (Continued)

Path Analysis

MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G. & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. Psychological Methods, 7, 83-104.

MacKinnon, D.P., Lockwood, C.M. & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. Multivariate Behavioral Research, 39, 99-128.

Shrout, P.E. & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. Psychological Methods, 7, 422-445.

EFA

Bartholomew, D.J. (1987). Latent variable models and factor analysis. New York: Oxford University Press.

Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research, 36, 111-150.

References (Continued)

Cudeck, R. & O'Dell, L.L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. Psychological Bulletin, 115, 475-487.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4, 272-299.

Gorsuch, R.L. (1983). Factor Analysis. 2nd edition. Hillsdale, N.J.: Lawrence Erlbaum.

Harman, H.H. (1976). Modern factor analysis. 3rd edition. Chicago: The University of Chicago Press.

Holzinger, K.J. & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. Supplementary Educational Monographs. Chicago, Ill.: The University of Chicago.

Kim, J.O. & Mueller, C.W. (1978). An introduction to factor analysis: what it is and how to do it. Sage University Paper series on Quantitative Applications in the Social Sciences, No 07-013. Beverly Hills, CA: Sage.

Jöreskog, K.G. (1977). Factor analysis by least-squares and maximum-likelihood methods. In Statistical methods for digital computers, K. Enslein, A. Ralston, and H.S. Wilf (Eds.). New York: John Wiley & Sonds, pp. 125-153.

References (Continued)

Jöreskog, K.G. (1979). Author's addendum. In Advances in factor analysis and structural equation models, J. Magidson (Ed.). Cambridge, Massachusetts: Abt Books, pp. 40-43.

Kim, J.O. & Mueller, C.W. (1978). An introduction to factor analysis: what it is and how to do it. Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-013. Beverly Hills, CA: Sage.

Mulaik, S. (1972). The foundations of factor analysis. McGraw-Hill.

Schmid, J. & Leiman, J.M. (1957). The development of hierarchical factor solutions. Psychometrika, 22, 53-61.

Spearman, C. (1927). The abilities of man. New York: Macmillan.

Thurstone, L.L. (1947). Multiple factor analysis. Chicago: University of Chicago Press.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC: American Psychological Association.

Tucker, L.R. (1971). Relations of factor score estimates to their use. Psychometrika, 36, 427-436.

References (Continued)

CFA

Bollen, K.A. (1989). Structural equations with latent variables. New York: John Wiley.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

Lawley, D.N. & Maxwell, A.E. (1971). Factor analysis as a statistical method. London: Butterworths.

Long, S. (1983). Confirmatory factor analysis. Sage University Paper series on Qualitative Applications in the Social Sciences, No. 3. Beverly Hills, CA: Sage.

Meredith, W. (1964). Notes on factorial invariance. Psychometrika, 29, 177-185.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika, 58, 525-543.

Millsap, R.E. (2001). When trivial constraints are not trivial: the choice of uniqueness constraints in confirmatory factor analysis. Structural Equation Modeling, 8, 1-17.

References (Continued)

Mulaik, S. (1972). The foundations of factor analysis. McGraw-Hill.

Muthén, B. (1989b). Factor structure in groups selected on observed scores. British Journal of Mathematical and Statistical Psychology, 42, 81-90.

Muthén, B. (1989c). Multiple-group structural modeling with non-normal continuous variables. British Journal of Mathematical and Statistical Psychology, 42, 55-62.

Muthén, B. & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. British Journal of Mathematical and Statistical Psychology, 38, 171-189.

Muthén, B. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. British Journal of Mathematical and Statistical Psychology, 45, 19-30.

Personality and Individual Differences, 42 (5), May 1997: Special issue on SEM model testing.

Mulaik, S.A. (2009). Linear causal modeling with structural equations. Boca Raton, FL: Chapman & Hall.

References (Continued)

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 27, 229-239.

Yung Y.F, Thissen, D. & McLeod, L.D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. Psychometrika, 64, 113-128.

MIMIC and Multiple Group Analysis

Hauser, R.M. & Goldberger, A.S. (1971). The treatment of unobservable variables in path analysis. In H. Costner (Ed.), Sociological Methodology (pp. 81-117). American Sociological Association: Washington, D.C.

Joreskog, K.G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

References (Continued)

Jöreskog, K.G. & Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. Journal of the American Statistical Association, 70, 631-639.

Meredith, W. (1964). Notes on factorial invariance. Psychometrika, 29, 177-185.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika, 58, 525-543.

Muthén, B. (1989a). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 27, 229-239.

SEM

Amemiya, T. (1985). Advanced econometrics. Cambridge, Mass.: Harvard University Press.

Beauducel, A. & Wittmann, W. (2005) Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure. Structural Equation Modeling, 12, 1, 41-75.

References (Continued)

Bollen, K.A. (1989). Structural equations with latent variables. New York: John Wiley.

Browne, M.W. & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C.C. Clogg & M.E. Sobel (Eds.), Handbook of statistical modeling for the social and behavioral sciences (pp. 311-359). New York: Plenum Press.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & K. Long (Eds.), Testing structural equation models (pp. 136-162). Newbury Park: Sage.

Fan, X. & Sivo, S.A. (2005) Sensitivity of fit indices to misspecified structural or measurement model components: rationale of two-index strategy revisited. Structural Equation Modeling, 12, 3, 343-367.

Hodge, R.W., Treiman, D.J. (1968). Social participation and social status. American Sociological Review, 33, 722-740.

Hu, L. & Bentler, P.M. (1998). Fit indices in covariance structure analysis: Sensitivity to underparameterized model misspecification. Psychological Methods, 3, 424-453.

Hu, L. & Bentler, P.M. (1999). Cutoff criterion for fit indices in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1-55.

References (Continued)

Jöreskog, K.G. (1973). A general method for estimating as linear structural equation system. In Structural equation models in the social sciences, A.S. Goldberger and O.D. Duncan, Eds.). New York: Seminar Press, pp. 85-12.

Jöreskog, K.G., & Sörbom, D. (1979). Advances in factor analysis and structural equation models. Cambridge, MA: Abt Books.

Kaplan, D. (2000). Structural equation modeling. Foundations and extensions. Thousand Oakes, CA: Sage Publications.

Klein, A. & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. Psychometrika, 65, 457-474.

MacCallum, R.C. & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. Annual Review of Psyhcology, 51, 201-226.

MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G. & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects Psychological Methods, 7, 83-104.

Marsh, H.W., Kit-Tai Hau & Z. Wen (2004) In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Structural Equation Modeling, 11, 3, 320-341.

References (Continued)

Marsh, H.W., Wen, X, & Hau, K.T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. Psychological Methods, 9, 275-300.

Muthén, .B & Muthén, L. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling, 4, 599-620.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In Heijmans, R.D.H., Pollock, D.S.G. & Satorra, A. (eds.), Innovations in Multivariate Statistical Analysis. A Festschrift for Heinz Neudecker (pp. 233-247). London: Kluwer Academic Publishers.

Satorra, A. & Bentler, P.M. (1999). A scaled difference chi-square test statistic for moment structure analysis. Technical report, University of California, Los Angeles.

Shrout, P.E. & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. Psychological Methods, 7, 422-445.

Skrondal, A. & Rabe-Hesketh, S. (2004). Generalized latent variable modeling. Multilevel, longitudinal, and structural equation models. London: Chapman Hall.

Sorbom, D. (1989). Model modifications. Psychometrika, 54, 371-384.

References (Continued)

Steiger, J.H. & Lind, J.M. (1980). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Wheaton, B., Muthén, B., Alwin, D. & Summers, G. (1977). Assessing reliability and stability in panel models. In D.R. Heise (Ed.), Sociological Methodology 1977 (pp. 84-136). San Francisco: Jossey-Bass.

Yu, C.Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Doctoral dissertation, University of California, Los Angeles.

General

Lord, F.M. & Novick, M.R. (1968). Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley Publishing Co.

Muthén, L.K. & Muthén, B.O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural Equation Modeling, 4, 599-620.

References (continued)

http://www.gsu.edu/~mkteer/bookfaq.html

http://gsm.uci.edu/~joelwest/SEM/SEMBooks.html

http://www2.chass.ncsu.edu/garson/pa765/structur.htm is a fairly complete (15) pages general overview of SEM.

Join SEMNET: http://bama.ua.edu/archives/semnet.html