Structural Equation Modeling

Petri Nokelainen
Research Centre for Vocational Education
University of Tampere, Finland
petri.nokelainen@uta.fi
http://users.uta.fi/u/petri.nokelainen

Contents

Introduction

- Development of Western science is based on two great achievements: the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the possibility to find out causal relationships by systematic experiment (during the Renaissance).

Albert Einstein
(in Pearl, 2000)

Introduction

- Structural equation modeling (SEM), as a concept, is a combination of statistical techniques such as exploratory factor analysis and multiple regression.
- The purpose of SEM is to examine a set of relationships between one or more Independent Variables (IV) and one or more Dependent Variables (DV).

Introduction

- Both IV's and DV's can be continuous or discrete.
- Independent variables are usually considered either predictor or causal variables because they predict or cause the dependent variables (the response or outcome variables).

Introduction

- Structural equation modeling is also known as ‘causal modeling’ or ‘analysis of covariance structures’.
- Path analysis and confirmatory factor analysis (CFA) are special types of SEM. (Figure 1.)

Introduction

- Genetics S. Wright (1921): *"Prior knowledge of the causal relations is assumed as prerequisite … [in linear structural modeling]".*

  $y = \beta x + \varepsilon$

"In an ideal experiment where we control X to x and any other set Z of variables (not containing X or Y) to z, the value of Y is given by $\beta x + \varepsilon$, where ε is not a function of the settings x and z." (Pearl, 2000)

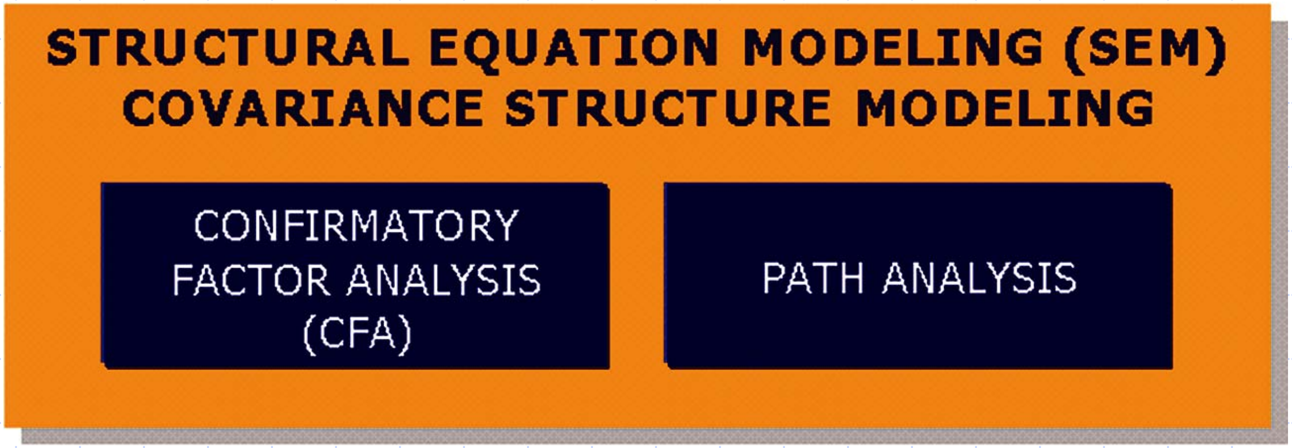
Introduction

- According to Judea Pearl (2000), modern SEM is a far cry from the original causality modeling theme, mainly for the following two reasons:
  - Researchers have tried to build scientific ’credibility’ of SEM by isolating (or removing) references to causality.
  - Causal relationships do not have commonly accepted mathematical notation.

Introduction

- Two main components of SEM are presented in Figure 1.
  - CFA operates with observed and latent variables, path analysis operates only with observed variables.

Figure 1. Components of Structural Equation Modeling



(Nokelainen, 1999.)

Contents

Path Analysis

- Examines how *n* independent (x, IV, Xi, ξ) variables are statistically related to a dependent (y, DV, Eta, η) variable.
- Applies the techniques of regression analysis, aiming at more detailed resolution of the phenomena under investigation.
- Allows
  - *Causal* interpretation of statistical dependencies
  - Examination of how data fits to a theoretical model

Path Analysis

- Once the data is available, conduction of path analysis is straightforward:
  1. Draw a path diagram according to the theory.
  2. Conduct one or more regression analyses.
  3. Compare the regression estimates (B) to the theoretical assumptions or (Beta) other studies.
  4. If needed, modify the model by removing or adding connecting paths between the variables and redo stages 2 and 3.

Path Analysis

- Data assumptions:
  - DV:
    - Continuous, normally distributed (univariate normality assumption)
  - IV:
    - Continuous (no dichotomy or categorical variables)
  - N:
    - About 30 observations for each IV

Path Analysis

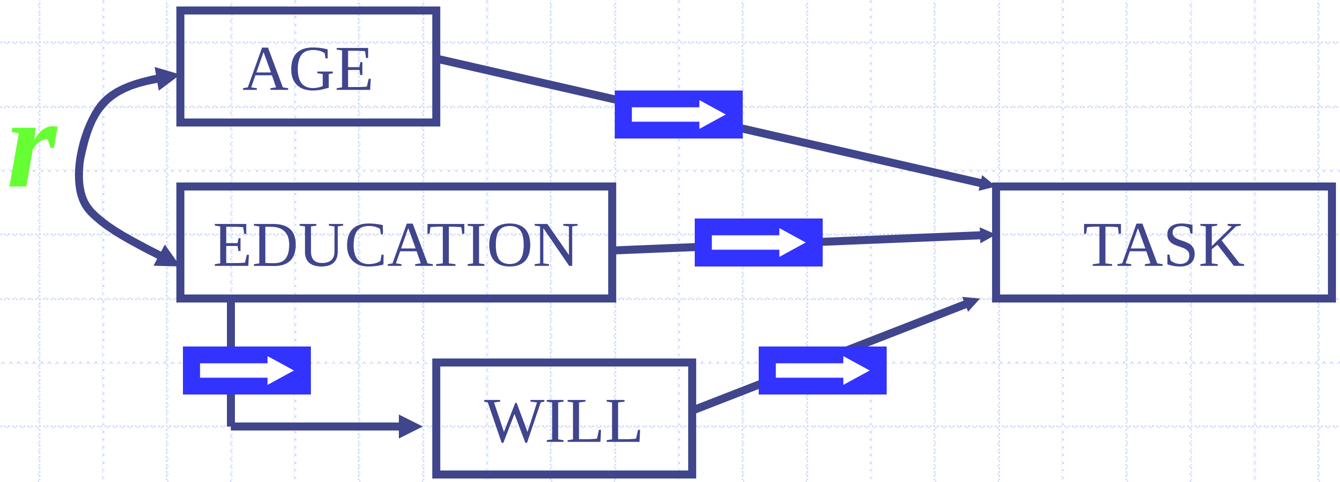
- Theoretical assumptions
  - Causality:
    - X_1 and Y_1 correlate.
    - X_1 precedes Y_1 chronologically.
    - X_1 and Y_1 are still related after controlling other dependencies.
- Statistical assumptions
  - Model needs to be recursive.
  - It is OK to use ordinal data.
  - All variables are measured (and analyzed) without measurement error ($\varepsilon = 0$).

Path Analysis

- As stated earlier, path analysis assumes that the model is recursive.
  - Nature of causal dependency is unidirectional, like a 'one way road' (arc with one head $\longrightarrow$).
  - If there is no a priori information available about the direction of causal dependency, it is assumed to be correlational (arc with two heads $\longleftrightarrow$).

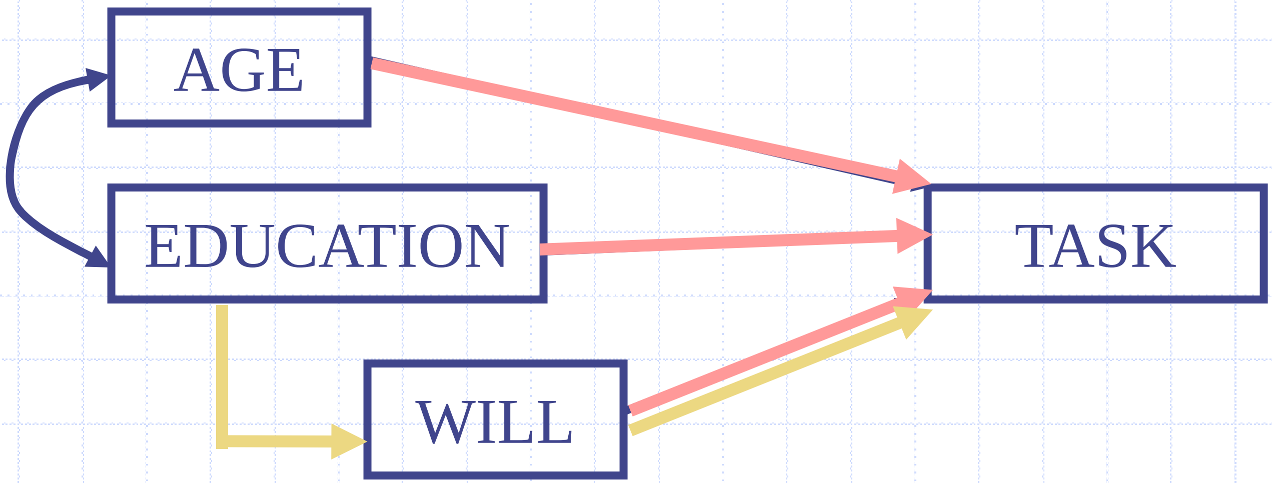
r

Path Analysis



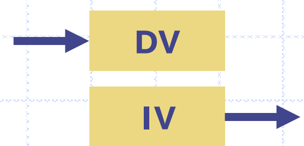
Path Analysis

- Direct and indirect effect



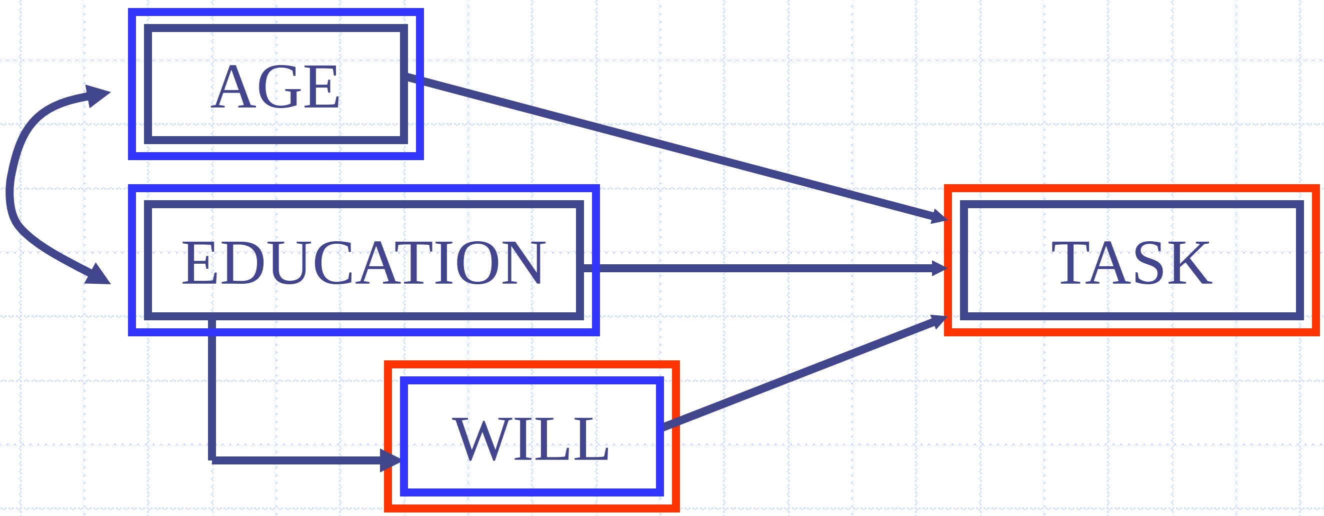
Path Analysis

- There are two types of observed variables:
  - Endogenous (y, DV, Eta η).
  - Exogenous (x, IV, Xi ξ).
- For each endogenous (DV) variable, a regression analysis is performed.



Path Analysis

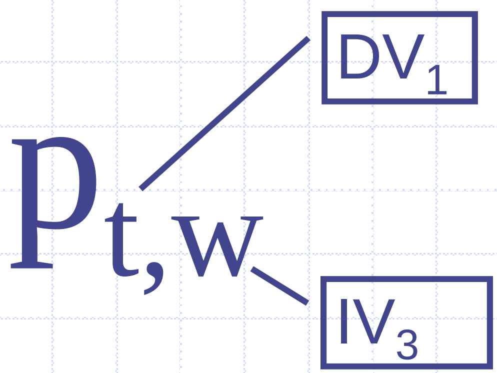
x IV Xi ξ EXOGENIOUS
y DV Eta η ENDOGENIOUS



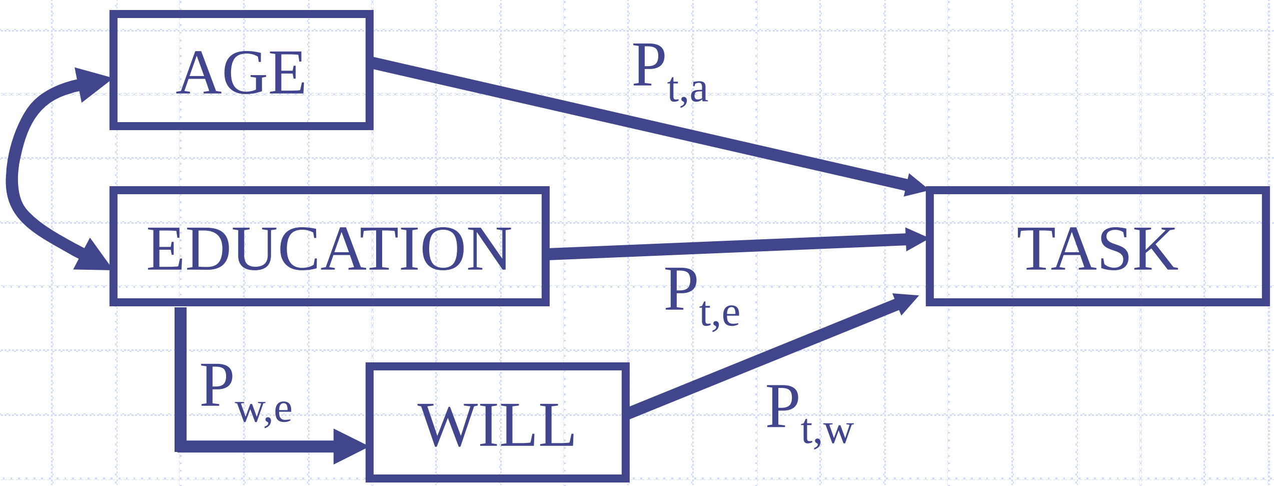
Two regression analyses:
1) AGE + EDUCATION + WILL -> TASK
2) EDUCATION –> WILL

Path Analysis

- Path coefficients are a product of one or more regression analyses.
  - They are indicators of statistical dependency between variables.



Path Analysis



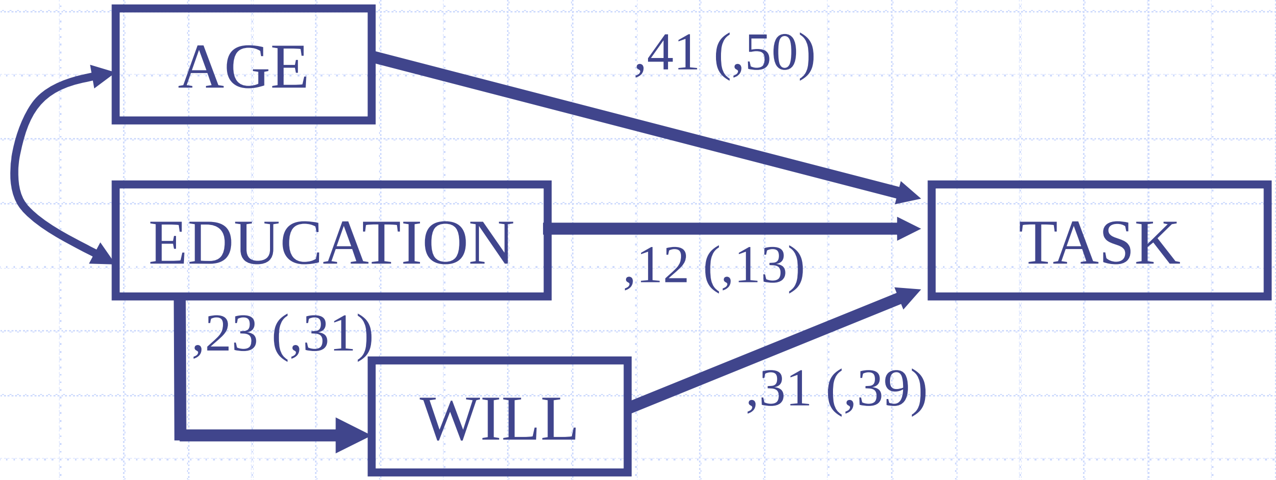
Path Analysis

- Path coefficients are standardized ('Beta') or unstandardized ('B') regression coefficients.
  - Strength of inter-variable dependencies are comparable to other studies when standardized values (z, where $M = 0$ and $SD = 1$) are used.
  - Unstandardized values allow the original measurement scale examination of inter-variable dependencies.

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}} \qquad z = \frac{(x - \bar{x})}{SD}$$

Path Analysis

- Beta (B)

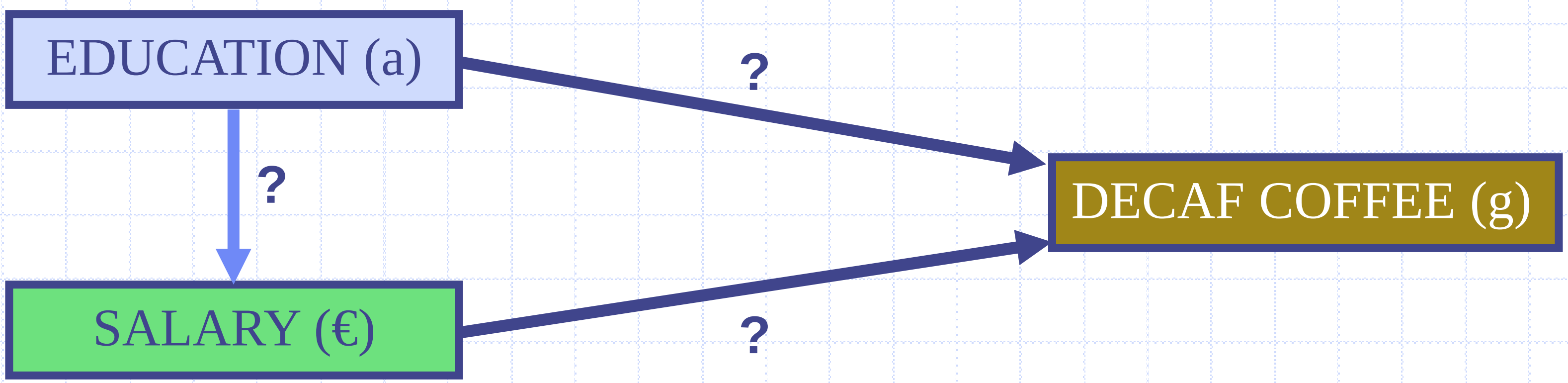


Path Analysis

- Path coefficient ($p_{DV,IV}$) indicates the direct effect of IV to DV.
- If the model contains only one IV and DV variable, the path coefficient equals to correlation coefficient.
  - In those models that have more than two variables (one IV and one DV), the path coefficients equal to partial correlation coefficients.
    - The other path coefficients are controlled while each individual path coefficient is calculated.

Path Analysis

- No need to use LISREL or AMOS
  - Two separate regression analyses in SPSS (Analyze – Regression – Linear)



1. Data (N = 10)

	EDUCATION	SALARY	DECAF_COFFEE
1	20	4000	500
2	19	3500	600
3	12	700	0
4	9	200	0
5	21	5000	1000
6	9	3000	0
7	19	1400	250
8	17	1200	200
9	9	1000	0
10	9	800	0

2. First SPSS regression analysis (SALARY + EDUCATION -> DECAF_COFFEE)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,937[a]	,877	,842	136,383

a. Predictors: (Constant), SALARY, EDUCATION

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-447,534	137,286		-3,260	,014
	EDUCATION	33,224	11,752	,506	2,827	,026
	SALARY	,108	,037	,518	2,894	,023

a. Dependent Variable: DECAF_COFFEE

3. Second SPSS regression analysis (EDUCATION -> SALARY)

Model Summary

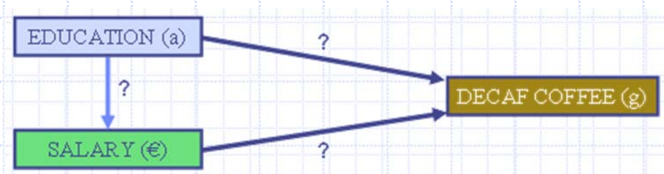
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,673[a]	,453	,385	1295,232

a. Predictors: (Constant), EDUCATION

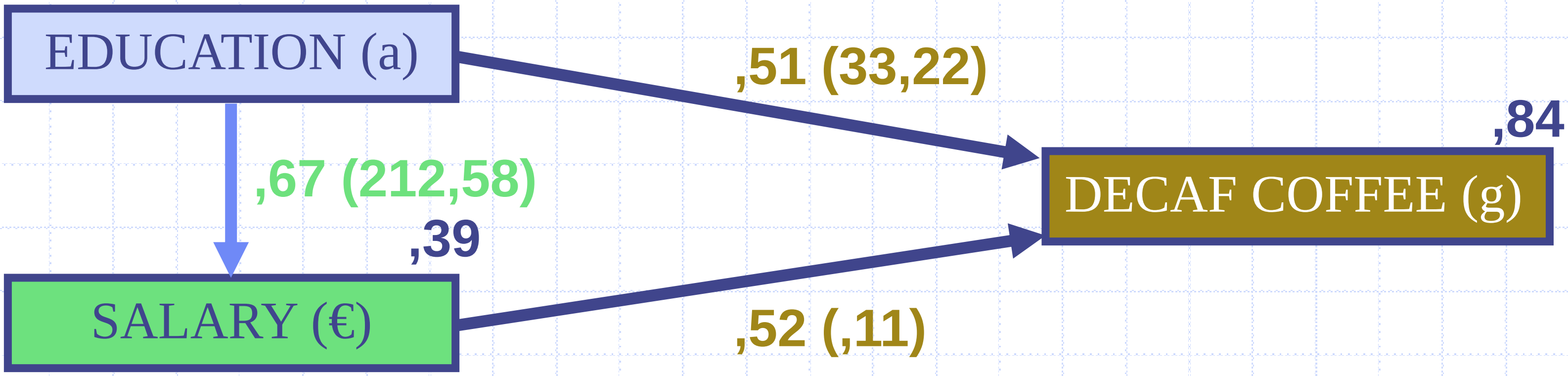
Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-981,169	1256,814		-,781	,457
	EDUCATION	212,581	82,514	,673	2,576	,033

a. Dependent Variable: SALARY

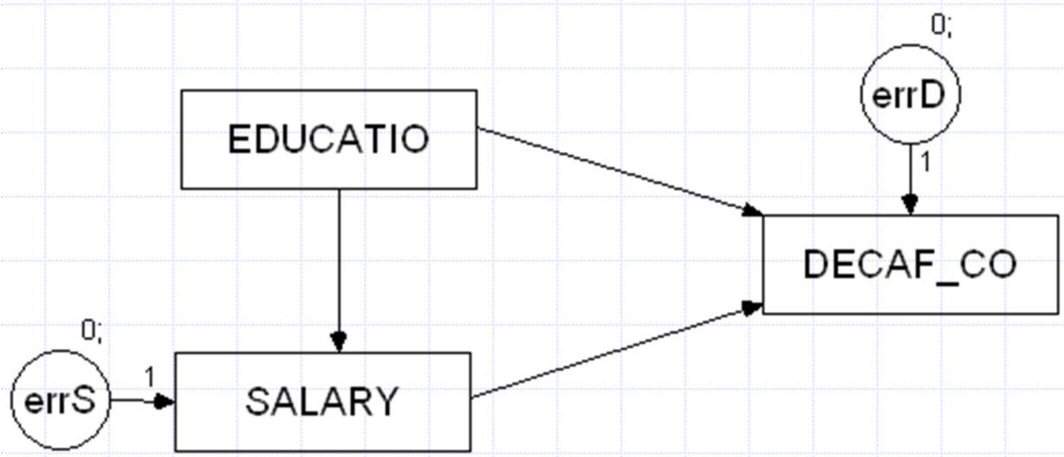


Path Analysis



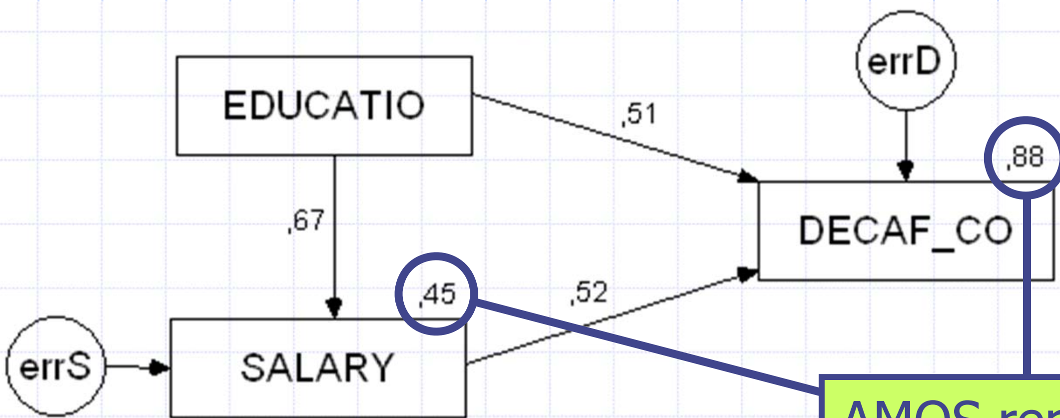
Path Analysis

- Here is the same model in AMOS:



Path Analysis

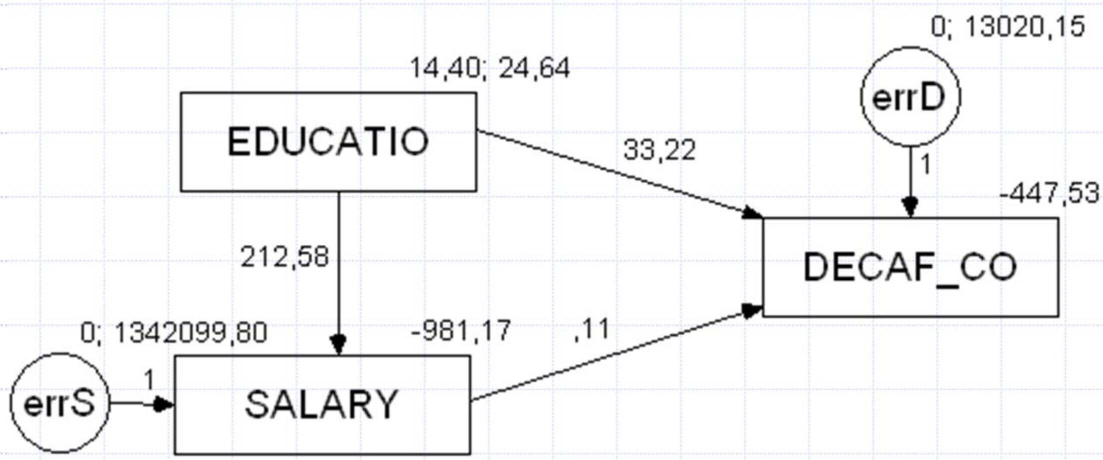
- And the results are naturally the same:
  - Standardized



AMOS reports R square instead of more critical Adjusted R square.

Path Analysis

- And the results are naturally the same:
  - Unstandardized



Contents

Basic Concepts of Factor Analysis

- The fundamental idea underlying the factor analysis is that some but not all variables can be directly observed.
- Those unobserved variables are referred to as either *latent* variables or factors.
- Information about latent variables can be gained by observing their influence on observed variables.
- Factor analysis examines covariation among a set of observed variables trying to *generate a smaller number of latent variables*.

Basic Concepts of Factor Analysis

- ***Exploratory Factor Analysis***
  - In exploratory factor analysis (EFA), observed variables are represented by squares and circles represent latent variables.
  - Causal effect of the latent variable on the observed variable is presented with straight line with arrowhead.

Basic Concepts of Factor Analysis

- ***Exploratory Factor Analysis***
  - The latent factors (ellipses) labeled with ξ's (Xi) are called common factors and the δ's (delta) (usually in circles) are called errors in variables or *residual variables*.
  - Errors in variables have unique effects to one and only one observed variable - unlike the common factors that share their effects in common with more than one of the observed variables.

Basic Concepts of Factor Analysis

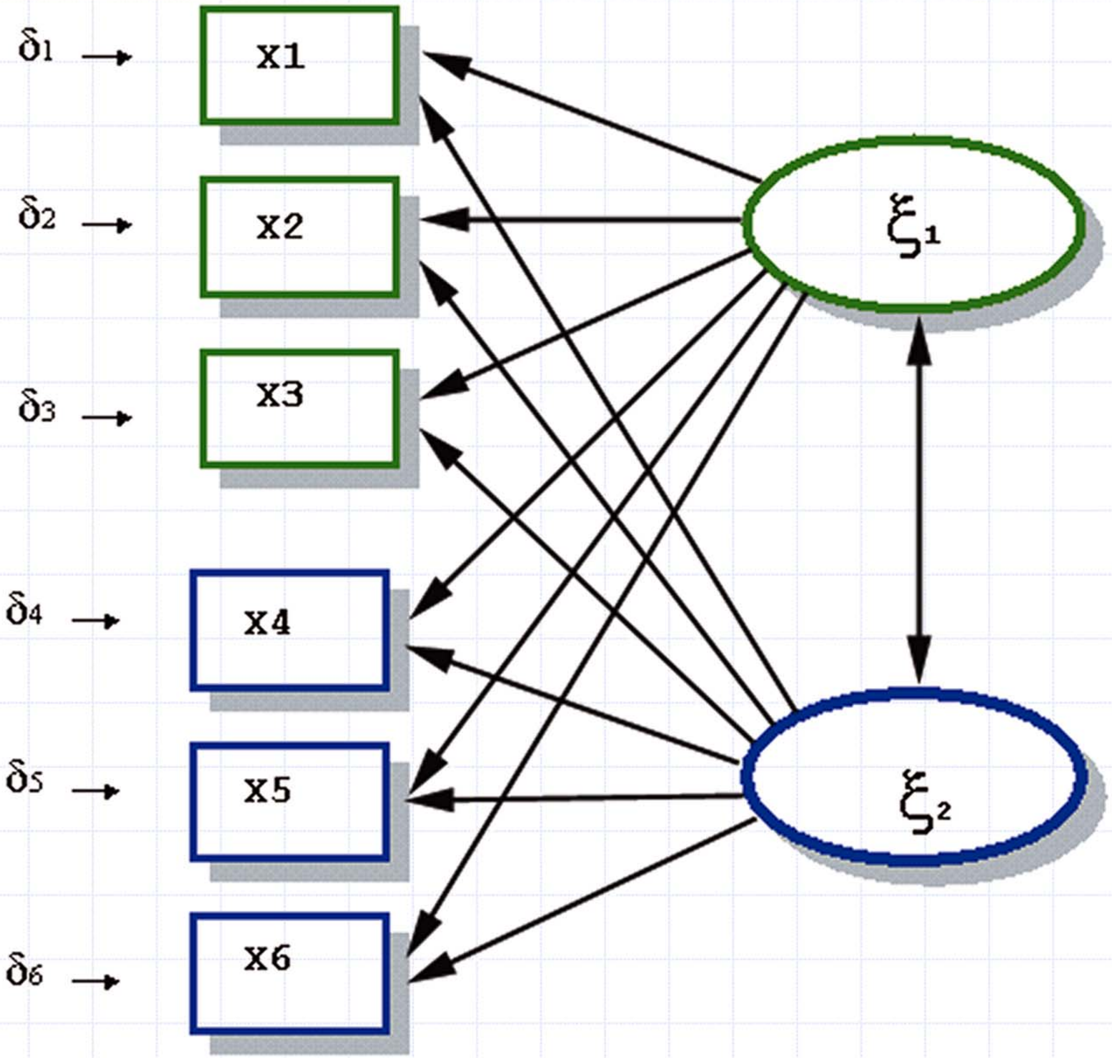


Figure 2.
Exploratory Factor Model
(Nokelainen, 1999.)

Basic Concepts of Factor Analysis

- ***Exploratory Factor Analysis***
  - The EFA model in Figure 2 reflects the fact that researcher does not specify the structure of the relationships among the variables in the model.
  - When carrying out EFA, researcher must assume that
    - all common factors are correlated,
    - all observed variables are directly affected by all common factors,
    - errors in variables are uncorrelated with one another,
    - all observed variables are affected by a unique factor and
    - all ξ's are uncorrelated with all δ's. (Long, 1983.)

Basic Concepts of Factor Analysis

Confirmatory Factor Analysis

- One of the biggest problems in EFA is its inability to incorporate substantively meaningful constraints.
- That is due to fact that algebraic mathematical solution to solve estimates is not trivial, instead one has to seek for other solutions.
- That problem was partly solved by the development of the confirmatory factor model, which was based on an iterative algorithm (Jöreskog, 1969).

Basic Concepts of Factor Analysis

- ***Confirmatory Factor Analysis***
  - In confirmatory factor analysis (CFA), which is a special case of SEM, the correlations between the factors are an explicit part of the analysis because they are collected in a matrix of factor correlations.
  - With CFA, researcher is able to decide *a priori* whether the factors would correlate or not. (Tacq, 1997.)

Basic Concepts of Factor Analysis

- ***Confirmatory Factor Analysis***
  - Moreover, researcher is able to impose substantively motivated constraints,
    - which common factor pairs that are correlated,
    - which observed variables are affected by which common factors,
    - which observed variables are affected by a unique factor and
    - which pairs of unique factors are correlated. (Long, 1983.)

Basic Concepts of Factor Analysis

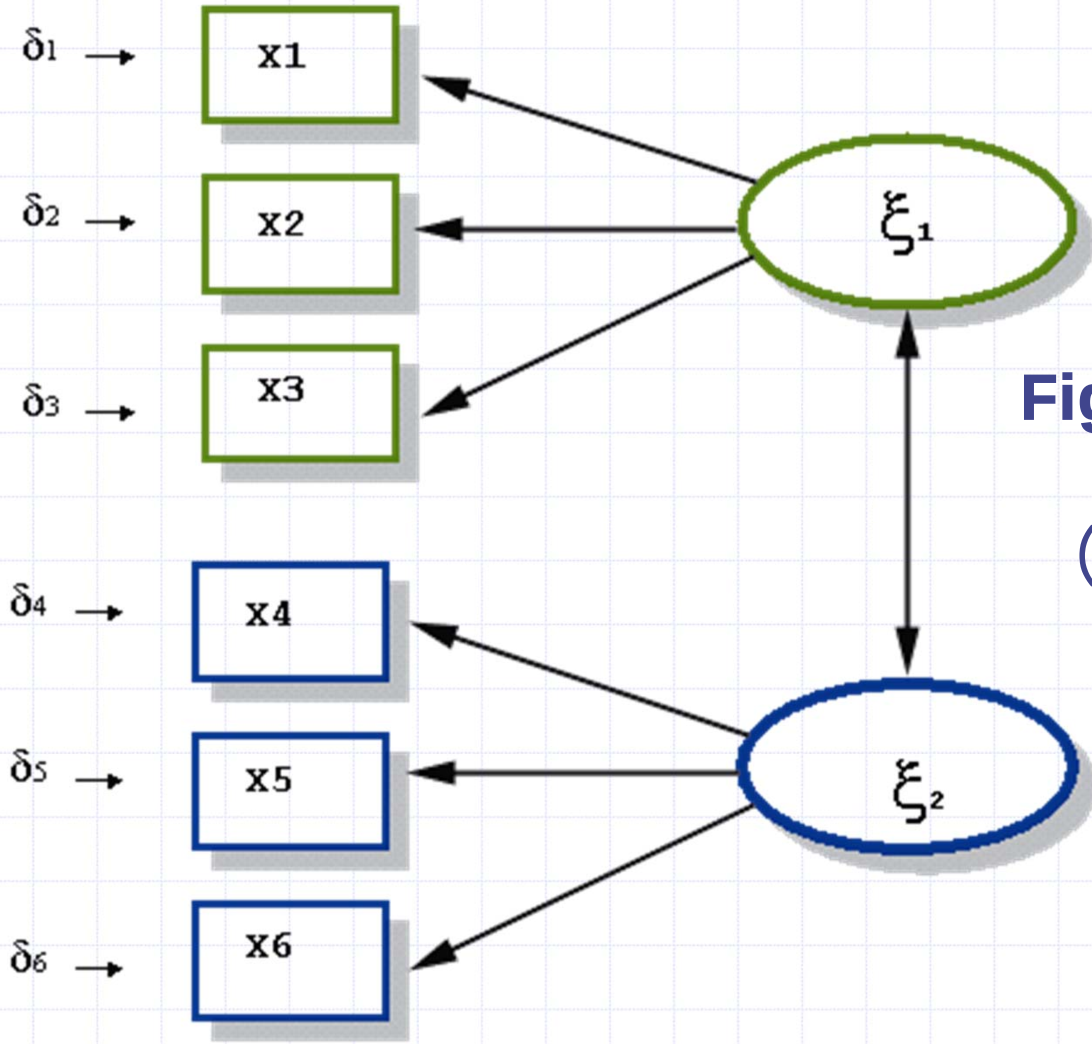


Figure 3. Confirmatory Factor Model (Nokelainen, 1999.)

Contents

Model Constructing

- One of the most well known covariance structure models is called LISREL (**L**ine**I**near **S**tructural **REL**ationships) or Jöreskog-Keesling-Wiley –model.
- LISREL is also a name of the software (Jöreskog et al., 1979), which is later demonstrated in this presentation to analyze a latent variable model.
- The other approach in this study field is Bentler-Weeks -model (Bentler et al., 1980) and EQS –software (Bentler, 1995).

Model Constructing

- The latest software release attempting to implement SEM is graphical and intuitive AMOS (Arbuckle, 1997).
- AMOS has since 2000 taken LISREL's place as a module of a well-known statistical software package SPSS (**S**tatistical **P**ackage for **S**ocial **S**ciences).
- Also other high quality SEM programs exist, such as Mplus (Muthén & Muthén, 2000).
  - MPlus is targeted for professional users, it has only text input mode.

Model Constructing

- In this presentation, I will use both the LISREL 8 –software and AMOS 5 for SEM analysis and PRELIS 2 –software (Jöreskog et al., 1985) for preliminary data analysis.
- All the previously mentioned approaches to SEM use the same pattern for constructing the model:

1. model hypotheses,
2. model specification,
3. model identification and
4. model estimation.

1. Model Hypotheses

- Next, we will perform a CFA model constructing process for a part of a “Commitment to Work and Organization” model.
- This is quite technical approach but unavoidable in order to understand the underlying concepts and a way of statistical thinking.

1. Model Hypotheses

- Next we study briefly basic concepts of factor analysis in order to understand the path which leads to structural equation modeling.
- To demonstrate the process, we study the theoretical model of ‘growth-oriented atmosphere’ (Ruohotie, 1996, 1999) to analyze organizational commitment.
- The data (N = 319), collected from Finnish polytechnic institute for higher education staff in 1998, contains six *continuous* summary variables (Table 1).

By stating ’continuous’, we assume here that mean of *n* Likert scale items with frequency of more than 100 observations produce **a summary item** (component or factor) that **behaves**, according to central limit theorem, **like a continuous variable with normal distribution.**

1. Model Hypotheses

Table 1. Variable Description

	Item	Summary variable	Sample statement
SUPPORTIVE MANAGEMENT	X1	Participative Leadership	It is easy to be touch with the leader of the training programme.
	X2	Elaborative Leadership	This organization improves it's members professional development.
	X3	Encouraging Leadership	My superior appreciates my work.
FUNCTIONAL GROUP	X4	Collaborative Activities	My teacher colleagues give me help when I need it.
	X5	Teacher – Student Connections	Athmosphere on my lectures is pleasant and spontaneous.
	X6	Group Spirit	The whole working community co-operates effectively.

1. Model Hypotheses

- A sample of the data is presented in Table 2.

Table 2. A Sample of the Raw Data Set

	Variables					
	Supportive Management			Functional Group		
Teachers	Participative Leadership (x1)	Elaborative Leadership (x2)	Encouraging Leadership (x3)	Collaborative Activities (x4)	Teacher-student Connections (x5)	Group Spirit (x6)
1.	2.75	3.25	4.00	2.60	3.00	2.00
2.	3.25	3.75	5.00	3.40	4.00	3.00
3.	3.50	3.75	4.00	3.60	4.75	3.00
. . .						
319	5.00	1.00	3.00	3.00	3.00	5.00

1. Model Hypotheses

- The covariance matrix is presented in Table 3.

Table 3. The Covariance Matrix

	Covariance Matrix					
	Supportive Management			Functional Group		
	X1	X2	X3	X4	X5	X6
X1	.734	.343	.438	.220	.104	.275
X2	.343	.668	.467	.234	.037	.307
X3	.438	.467	.938	.306	.165	.391
X4	.220	.234	.306	.459	.182	.308
X5	.104	.037	.165	.182	.387	.114
X6	.275	.307	.391	.308	.114	.552

1. Model Hypotheses

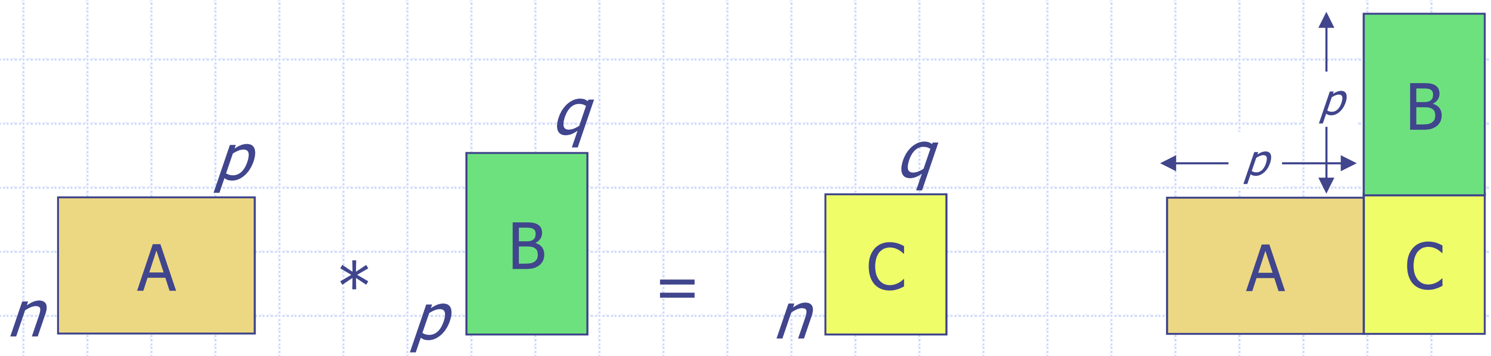
What is covariance matrix?

- Scatter, covariance, and correlation matrix form the basis of a multivariate method.
- The correlation and the covariance matrix are also often used for a first inspection of relationships among the variables of a multivariate data set.
- All of these matrices are calculated using the matrix multiplication (A · B).
- The only difference between them is how the data is scaled before the matrix multiplication is executed:
  - scatter: no scaling
  - covariance: mean of each variable is subtracted before multiplication
  - correlation: each variable is standardized (mean subtracted, then divided by standard deviation)

1. Model Hypotheses

What is matrix multiplication?

- Let (a_{rs}), (b_{rs}), and (c_{rs}) be three matrices of order $m_x n$ $n_x p$ and $p_x q$ respectively. Each element c_{rs} of the matrix *C*, the result of the *matrix product* **A•B,** is then calculated by the inner product of the *s* th row of *A* with the *r* th column of *B*.

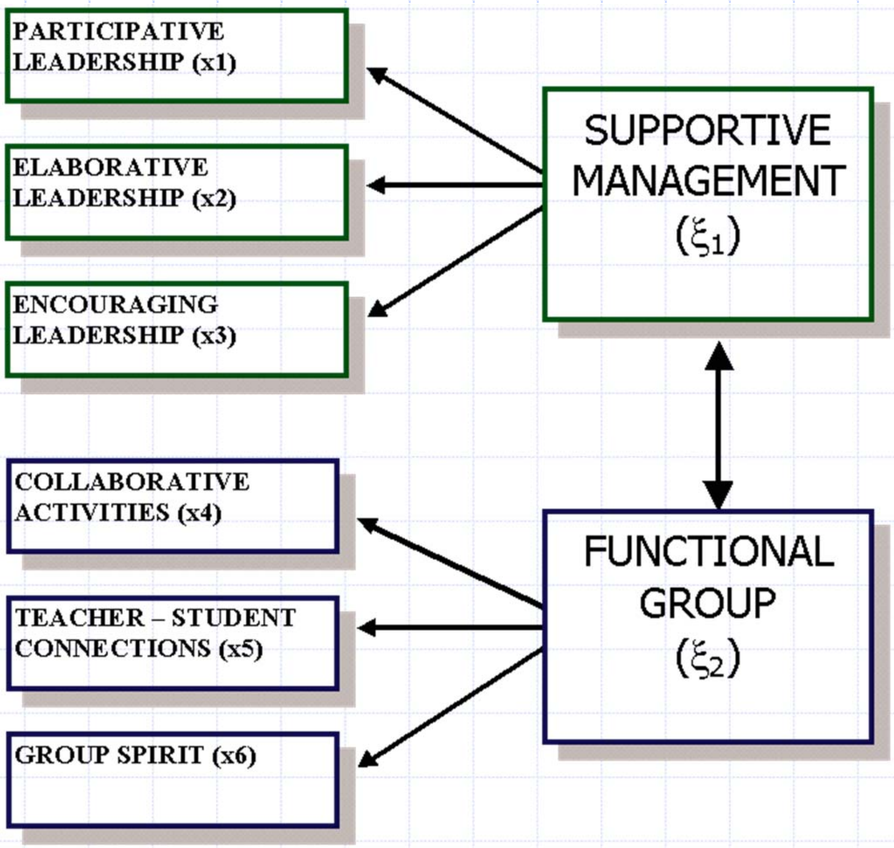


1. Model Hypotheses

- The basic components of the confirmatory factor model are illustrated in Figure 4.
- Hypothesized model is sometimes called *a structural model*.

1. Model Hypotheses

Figure 4. Hypothesized Model



1. Model Hypotheses

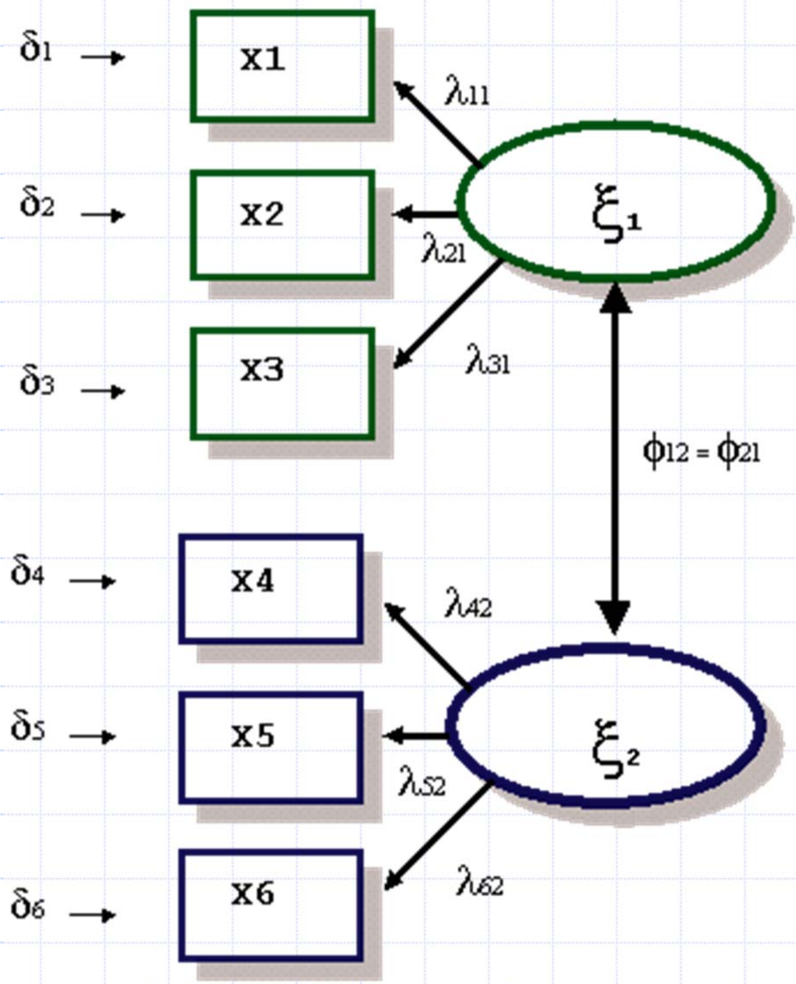
- Two main hypotheses of interest are:
  - Does a two-factor model fit the data?
  - Is there a significant covariance between the supportive and functional factors?

2. Model Specification

- Because of confirmatory nature of SEM, we continue our model constructing with the model specification to the stage, which is referred as *measurement model* (Figure 5).

2. Model Specification

Figure 5. Measurement Model



2. Model Specification

- One can specify a model with different methods, e.g., Bentler-Weeks or LISREL.
  - In Bentler-Weeks method every variable in the model is either an IV or a DV.
  - The parameters to be estimated are
    - the regression coefficients and
    - the variances and the covariances of the independent variables in the model. (Bentler, 1995.)

2. Model Specification

- Specification of the confirmatory factor model requires making formal and explicit statements about
  - the number of common factors,
  - the number of observed variables,
  - the variances and covariances among the common factors,
  - the relationships among observed variables and latent factors,
  - the relationships among residual variables and
  - the variances and covariances among the residual variables. (Jöreskog et al., 1989.)

2. Model Specification

- We start model specification by describing factor equations in a two-factor model: a Supportive Management factor (x1 – x3) and a Functional Group factor (x4 – x6), see Figure 5.
  - Note that the observed variables do not have direct links to all latent factors.

2. Model Specification

- The relationships for this part of the measurement model can now be specified in a set of *factor equations* in a scalar form:

$$x_1 = \lambda_{11}\xi_1 + \delta_1 \qquad x_2 = \lambda_{21}\xi_1 + \delta_2$$
$$x_3 = \lambda_{31}\xi_1 + \delta_3 \qquad x_4 = \lambda_{42}\xi_2 + \delta_4$$
$$x_5 = \lambda_{52}\xi_2 + \delta_5 \qquad x_6 = \lambda_{62}\xi_2 + \delta_6 \qquad (1)$$

  - δ_i is the residual variable (error) which is the unique factor affecting x_i. λ_{ij} is the loading of the observed variables x_i on the common factor ξ_j .

2. Model Specification

- Note that factor equations are similar to a familiar regression equation:

$$Y = X\beta + \varepsilon \qquad (2)$$

2. Model Specification

- Most of the calculations are performed as matrix computations because SEM is based on covariance matrices.
  - To translate equation (1) into a more matrix friendly form, we write:

$$x_1 = \lambda_{11}\xi_1 + 0\xi_2 + \delta_1 \quad (3a)$$

$$x_2 = \lambda_{21}\xi_1 + 0\xi_2 + \delta_2 \quad (3b)$$

$$x_3 = \lambda_{31}\xi_1 + 0\xi_2 + \delta_3 \quad (3c)$$

$$x_4 = 0\xi_1 + \lambda_{42}\xi_2 + \delta_4 \quad (3d)$$

$$x_5 = 0\xi_1 + \lambda_{52}\xi_2 + \delta_5 \quad (3e)$$

$$x_6 = 0\xi_2 + \lambda_{62}\xi_2 + \delta_6 \quad (3f)$$

2. Model Specification

- Mathematically, the relationship between the observed variables and the factors is expressed as matrix equation

  $$x = \Lambda_x \xi + \delta \qquad (4)$$

  and the matrix form for the measurement model is now written in a matrix form:

2. Model Specification

$$
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix} \quad (5)
$$

x_1 is defined as a linear combination of the latent variables ξ_1 ξ_2 and δ_1.

The coefficient for x_1 is λ_{11} indicating that a unit change in a latent variable ξ_1 results in an average change in x_1 of λ_{11} units.

The coefficient for ξ_2 is fixed to zero.

Each observed variable x_i has also residual factor δ_i which is the error of measurement in the x_i's on the assumption that the factors do not fully account for the indicators.

2. Model Specification

- The covariances between factors in Figure 5 are represented with arrows connecting ξ_1 and ξ_2.
  - This covariance is labeled $\phi_{12} = \phi_{21}$ in Φ.

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \qquad (6)$$

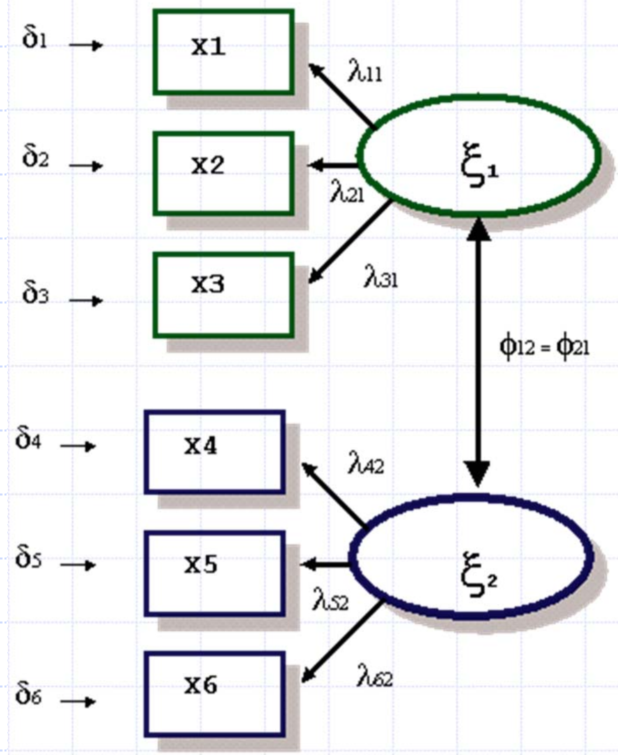
2. Model Specification

- The diagonal elements of Φ are the variances of the common factors.
  - Variances and covariances among the error variances are contained in Θ.

2. Model Specification

- In this model (see Figure 5), error variances are assumed to be uncorrelated:

$$\Theta = \begin{bmatrix} \theta_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{66} \end{bmatrix} \quad (7)$$



2. Model Specification

- Because the factor equation (4) cannot be directly estimated, the covariance structure of the model is examined.
- Matrix Σ contains the structure of covariances among the observed variables after multiplying equation (4) by its transpose

  $$\Sigma = E(xx') \qquad (8)$$

  and taking expectations

  $$\Sigma = E[(\Lambda\xi+\delta)(\Lambda\xi+\delta)'] \qquad (9)$$

2. Model Specification

- Next we apply the matrix algebral information that the transpose of a sum matrices is equal to the sum of the transpose of the matrices, and the transpose of a product of matrices is the product of the transposes in reverse order (see Backhouse et al., 1989):

$$\Sigma = E[(\Lambda\xi+\delta)\ (\xi'\Lambda'+\delta')] \qquad (10)$$

2. Model Specification

- Applying the distributive property for matrices we get

  $\Sigma = E[\Lambda\xi\xi'\Lambda' + \Lambda\xi\delta' + \delta\xi'\Lambda' + \delta\delta']$ (11)

- Next we take expectations

  $\Sigma = E[\Lambda\xi\xi'\Lambda'] + E[\Lambda\xi\delta'] + E[\delta\xi'\Lambda'] + E[\delta\delta']$ (12)

2. Model Specification

- Since the values of the parameters in matrix Λ are constant, we can write

$$\Sigma = \Lambda E[\xi\xi'] \Lambda' + \Lambda E[\xi\delta'] + E[\delta\xi'] \Lambda' + E[\delta\delta'] \qquad (13)$$

2. Model Specification

- Since $E[\xi\xi'] = \Phi$, $E[\delta\delta'] = \Theta$, and δ and ξ are uncorrelated, previous equation can be simplified to *covariance equation*:

  $$\Sigma = \Lambda\Phi\Lambda' + \Theta \qquad (14)$$

- The left side of the equation contains the number of unique elements $q(q+1)/2$ in matrix Σ.
- The right side contains $qs + s(s+1)/2 + q(q+1)/2$ unknown parameters from the matrices Λ, Φ, and Θ.
- Unknown parameters have been tied to the population variances and covariances among the observed variables which can be directly estimated with sample data.

3. Model Identification

- Identification is a theoretical property of a model, which depends neither on data or estimation.
  - When our model is identified we obtain unique estimates of the parameters.
- “Attempts to estimate models that are not identified result in arbitrary estimates of the parameters.” (Long, 1983, p. 35.)

3. Model Identification

- A model is identified if it is possible to solve the covariance equation $\Sigma = \Lambda\Phi\Lambda' + \Theta$ for the parameters in Λ, Φ and Θ.
  - Estimation assumes that model is identified.
- There are three conditions for identification:
  - *necessary conditions*, which are essential but not sufficient,
  - *sufficient conditions*, which if met imply that model is identified but if not met do not imply opposite (unidentified),
  - *necessary and sufficient conditions*.

3. Model Identification

- Necessary condition is simple to test since it relates the number of independent covariance equations to the number of independent parameters.
- Covariance equation (14) contains $q(q+1)/2$ independent equations and $qs + s(s+1)/2 + q(q+1)/2$ possible independent parameters in Λ, Φ and Θ.
  - Number of independent, unconstrained parameters of the model must be less than or equal to $q(q+1)/2$.

3. Model Identification

- We have six observed variables and, thus, 6(6+1)/2 = 21 distinct variances and covariances in Σ.
  - There are 15 independent parameters:

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 0 \\ 0 & 4 \\ 0 & 5 \\ 0 & 6 \end{bmatrix} \qquad \Phi = \begin{bmatrix} 7 & 8 \\ 8 & 9 \end{bmatrix} \qquad \Theta = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 11 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 13 & 0 & 0 \\ 0 & 0 & 0 & 0 & 14 & 0 \\ 0 & 0 & 0 & 0 & 0 & 15 \end{bmatrix}$$

3. Model Identification

- Since the number of independent parameters is smaller than the independent covariance equations (15<21), the necessary condition for identification is satisfied.

3. Model Identification

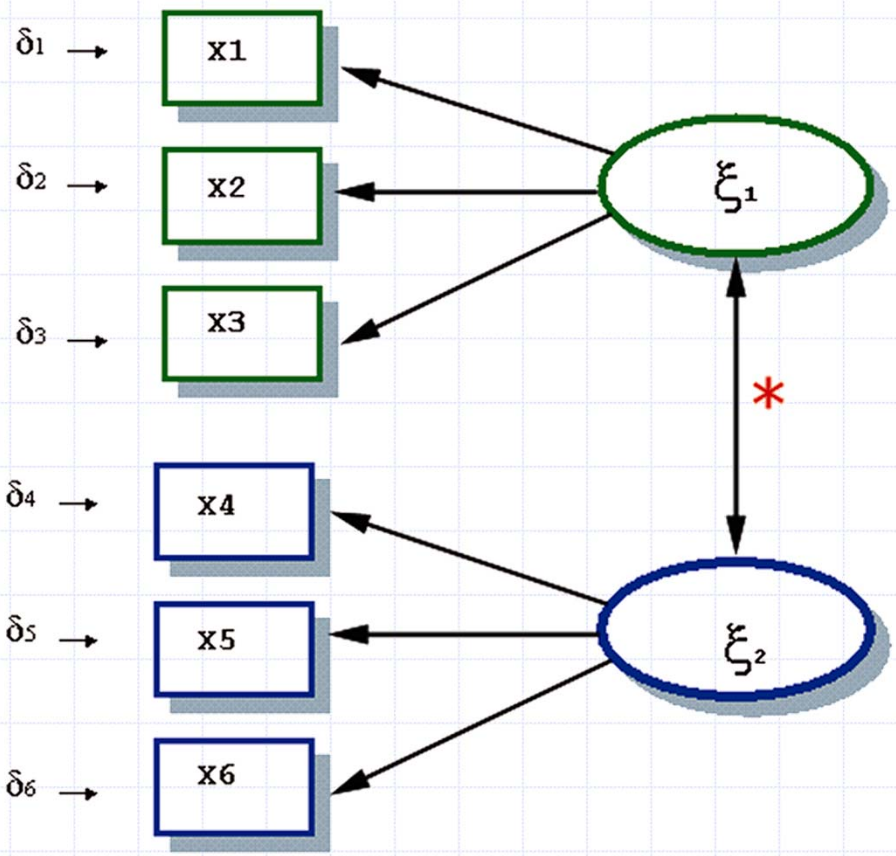
- The most effective way to demonstrate that a model is identified is to show that each of the parameters can be solved in terms of the population variances and covariances of the observed variables.
  - Solving covariance equations is time-consuming and there are other 'recipe-like' solutions.

3. Model Identification

- We gain constantly an identified model if
  - each observed variable in the model measures only one latent factor and
  - factor scale is fixed (Figure 6) or one observed variable per factor is fixed (Figure 7). (Jöreskog et al., 1979, pp. 196-197; 1984.)

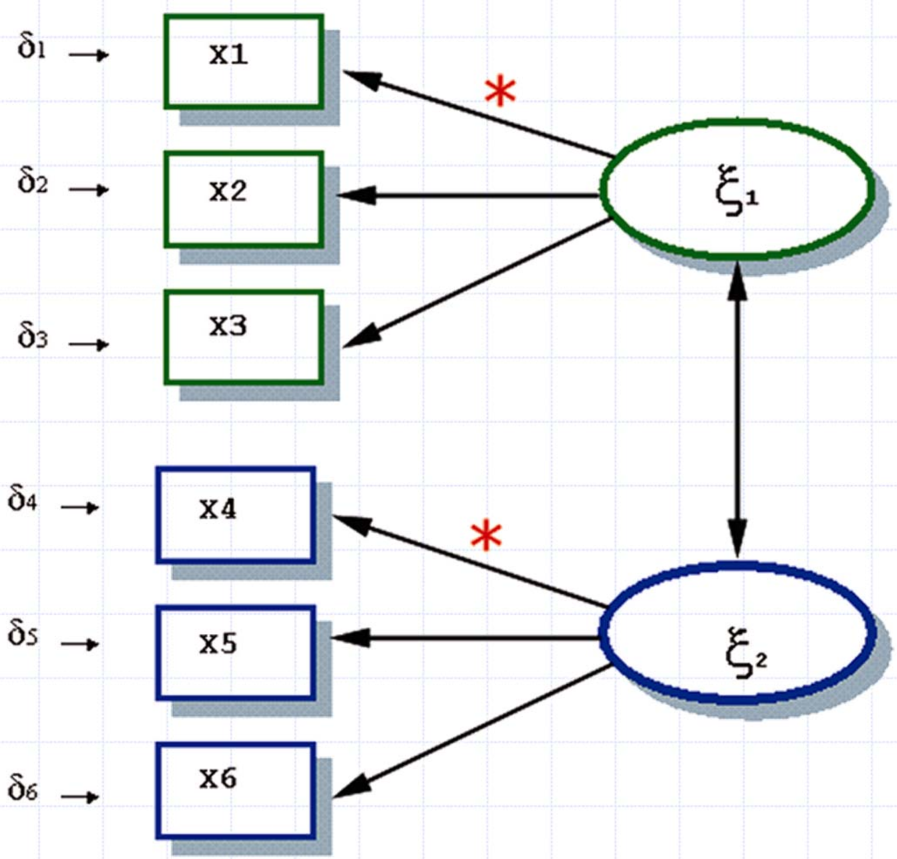
3. Model Identification

Figure 6. Factor Scale Fixed



3. Model Identification

Figure 7. One Observed Variable per Factor is Fixed



4. Model Estimation

- When identification is approved, estimation can proceed.
- If the observed variables are normal and linear and there are more than 100 observations (319 in our example), Maximum Likelihood estimation is applicable.

4. Model Estimation

Figure 8. LISREL 8 Input File

```
NONE - HAMKK part 01 - LISREL - pn1999

OBSERVED VARIABLES
x1 - x6

COVARIANCE MATRIX FROM FILE ha_01_x6.cov

SAMPLE SIZE 319

LATENT VARIABLES
SUP FUN

RELATIONSHIPS
COM = STI
x1 = 1*SUP
x2 x3 = SUP
x4 = 1*FUN
x5 x6 = FUN

LISREL OUTPUT ALL

PATH DIAGRAM

END OF PROBLEM

□
```

4. Model Estimation

Figure 4. Hypothesized Model

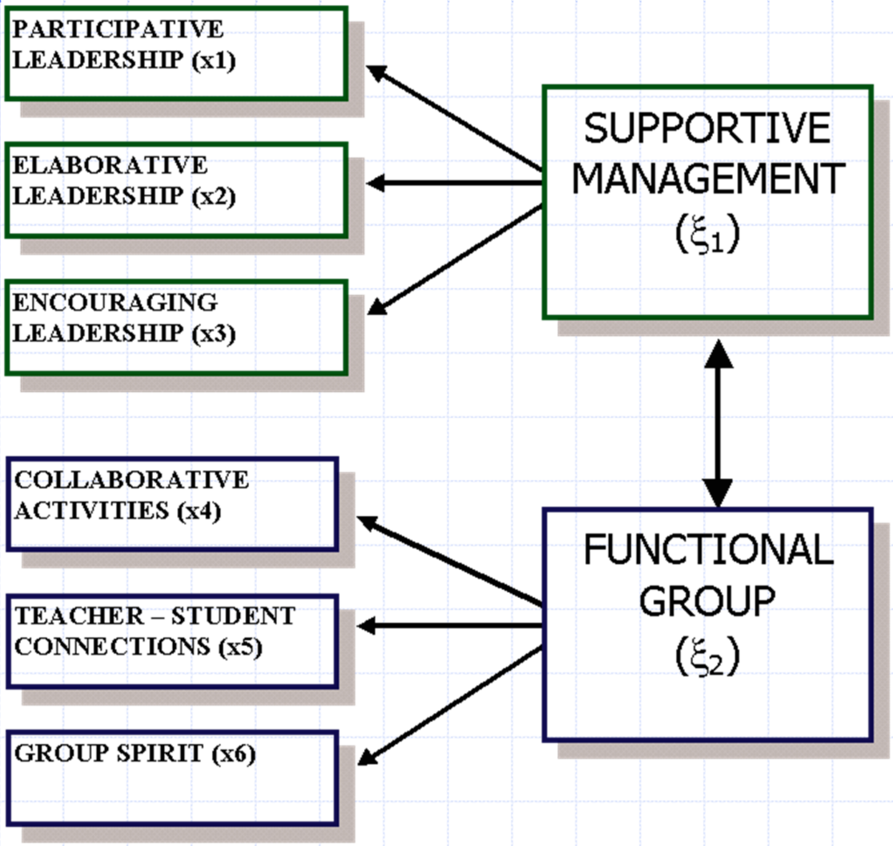
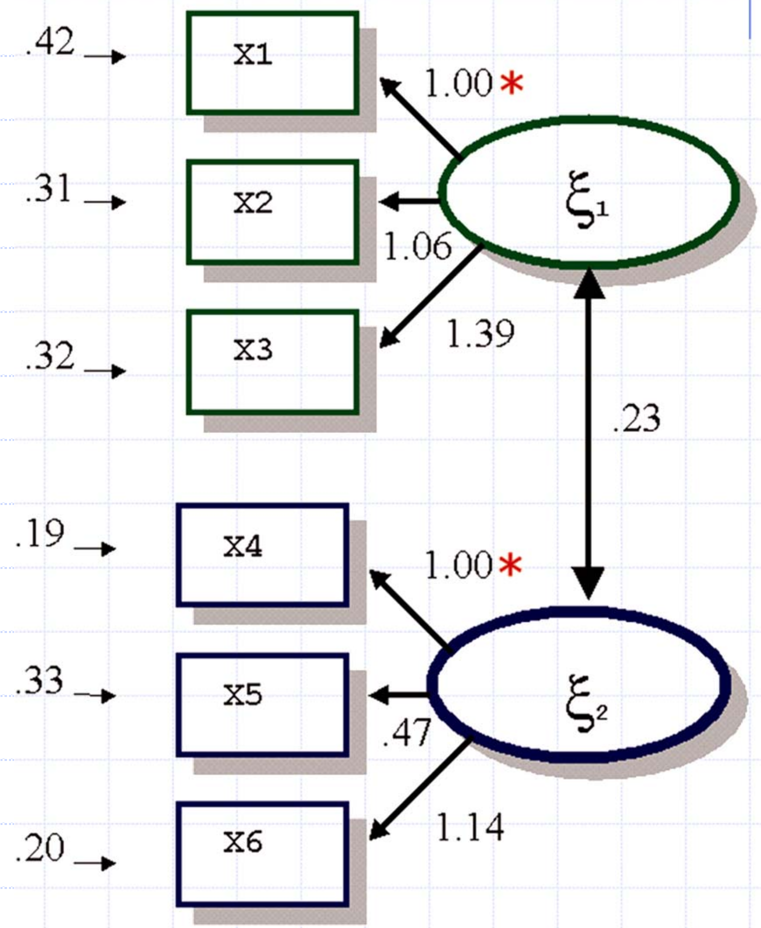


Figure 9. Parameter Estimates



Contents

An Example of SEM: Commitment to Work and Organization

Background

- In 1998 RCVE undertook a growth-oriented atmosphere study in a Finnish polytechnic institute for higher education (later referred as 'organization').
- The organization is a training and development centre in the field of vocational education.

An Example of SEM: Commitment to Work and Organization

Background

- In addition to teacher education, this organization promotes vocational education in Finland through developing vocational institutions and by offering their personnel a variety of training programmes which are tailored to their individual needs.
- The objective of the study was to obtain information regarding the current attitudes of teachers of the organization to their commitment to working environment (e.g., O'Neill et al., 1998).

An Example of SEM: Commitment to Work and Organization

Table 6. Dimensions of the Commitment to Work and Organization Model

DV (Eta$_1$ η_1)	**Commitment to Work and Organization**	**COM**	Commitment to work and organization	CO
IV$_1$ (Xi$_1$ ξ_1)	**Supportive Management**	**SUP**	Participative Leadership	PAR
			Elaborative Leadership	ELA
			Encouraging Leadership	ENC
IV$_2$ (Xi$_2$ ξ_2)	**Functional Group**	**FUN**	Collaborative Activities	COL
			Teacher – Student Connections	CON
			Group Spirit	SPI
IV$_3$ (Xi$_3$ ξ_3)	**Stimulating Job**	**STI**	Inciting Values	INC
			Job Value	VAL
			Influence on Job	INF

An Example of SEM: Commitment to Work and Organization

- **Sample**
  - A drop-off and mail-back methodology was used with a paper and pencil test.
  - Total of 319 questionnaires out of 500 (63.8%) was returned.
  - The sample contained 145 male (46%) and 147 female (46%) participants (*n* = 27, 8% missing data).
  - Participants most common age category was 40-49 years (*n* = 120, 37%).
  - Participants were asked to report their opinions on a 'Likert scale' from 1 (totally disagree) to 5 (totally agree).
  - All the statements were in positive wording.

An Example of SEM: Commitment to Work and Organization

- Model hypotheses
  - The following hypotheses were formulated:
    - ***Hypothesis 1.*** Supportive management (SUP), functional group (FUN) and stimulating job (STI) will be positively associated with commitment towards work and organization (COM).
    - ***Hypothesis 2.*** Significant covariance exists between the supportive (SUP), functional (FUN) and stimulating (STI) factors.

An Example of SEM: Commitment to Work and Organization

- Model Specification
  - The hypothesized model includes both
    - the *structural model* presenting the theoretical relationships among a set of latent variables, and
    - the *measurement model* presenting the latent variables as a linear combinations of the observed indicator variables.
  - The structural model (Figure 13) and measurement model (Figure 14) are built on the basis of the two hypotheses:

An Example of SEM: Commitment to Work and Organization

Figure 13. Structural Model

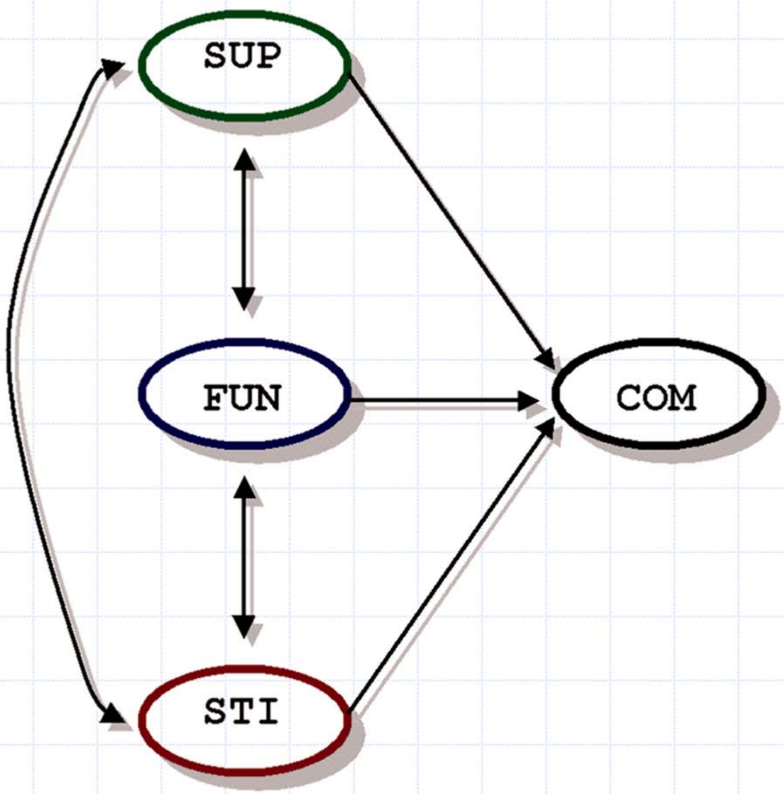
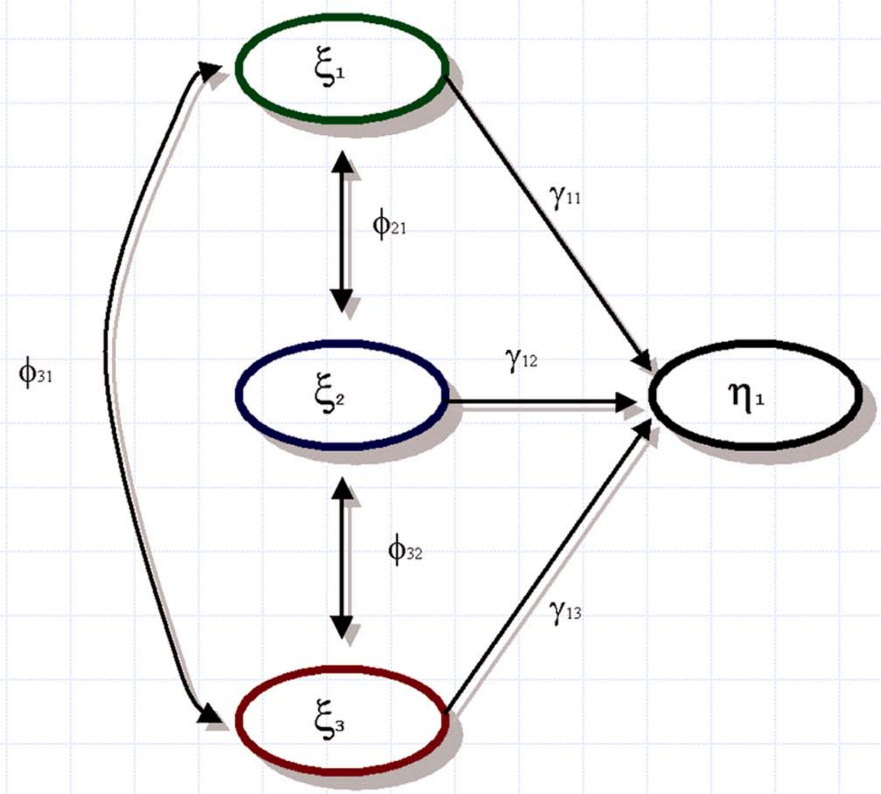


Figure 14. Measurement Model

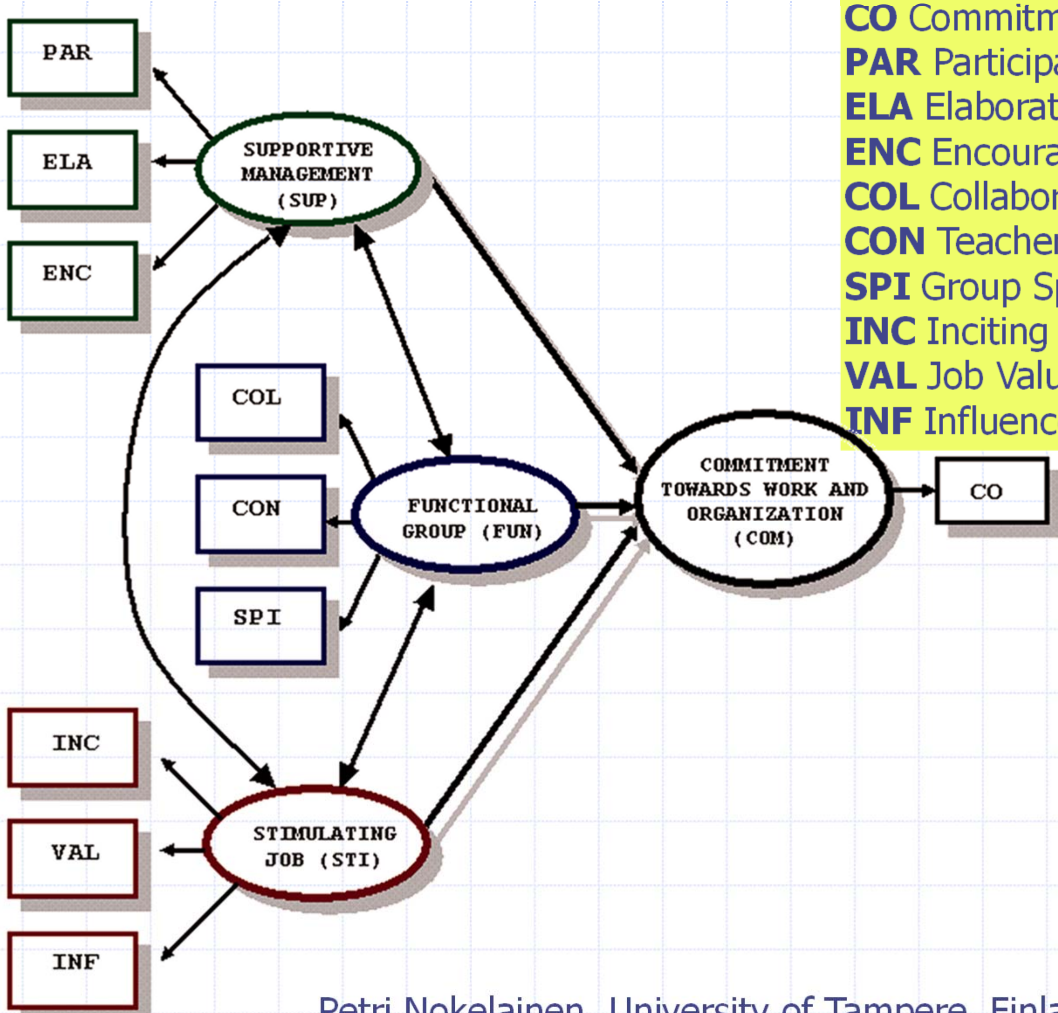


An Example of SEM: Commitment to Work and Organization

- The hypothesized model is presented in Figure 15:
  - A Commitment towards work and organization (COM η_1) with
    - CO (Y_1),
  - A Supportive Management (SUP ξ_1) with
    - PAR (X_1) ELA (X_2) and ENC (X_3),
  - A Functional group (FUN ξ_2) with
    - COL (X_4) CON (X_5) and SPI ($X6_)$, and
  - A Stimulating work (STI ξ_3) with
    - INC (X_7) VAL (X_8) and INF (X_9).

An Example of SEM: Commitment to Work and Organization

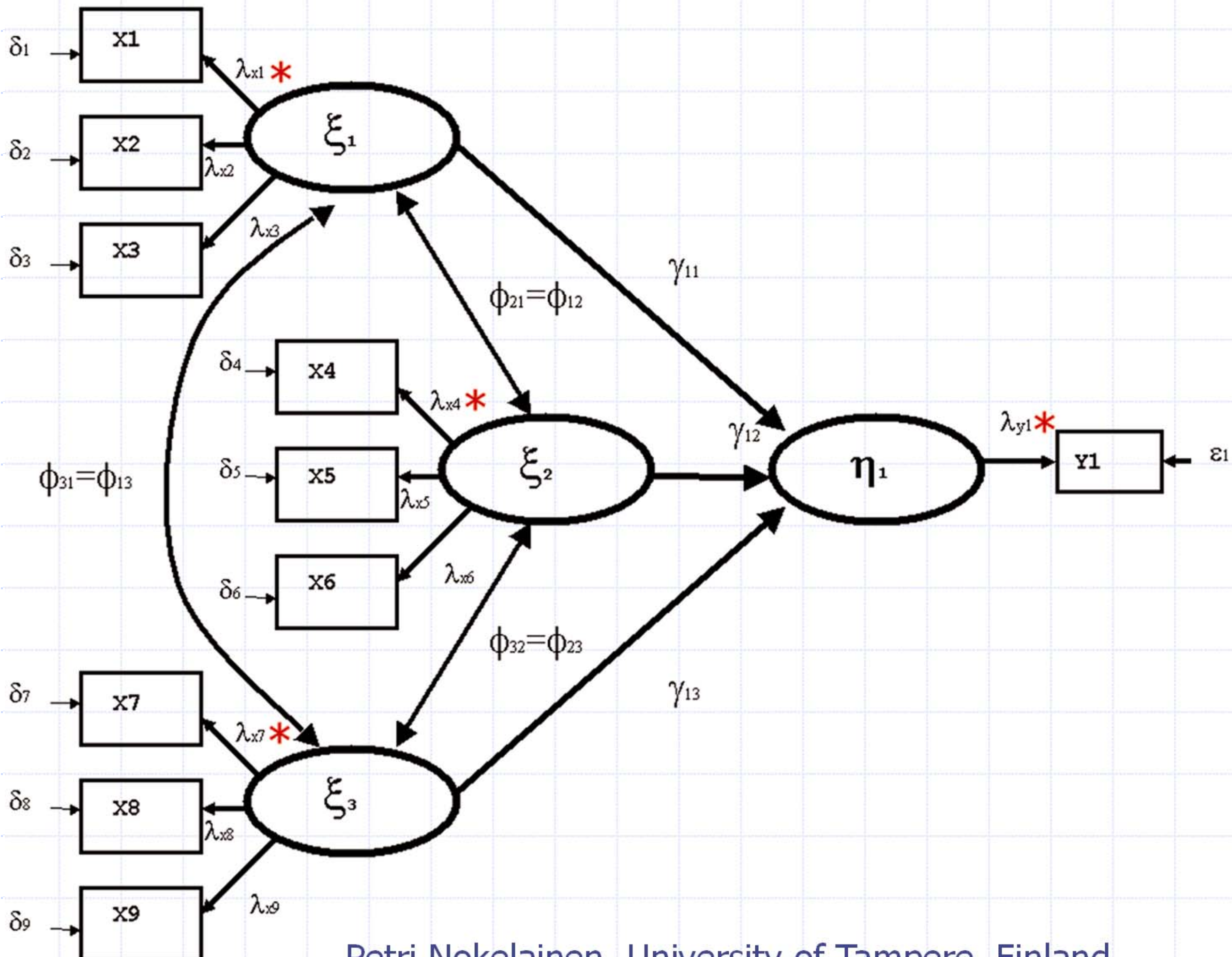
Figure 15. Hypothesized Structural Model



CO Commitment to work and organization
PAR Participative Leadership
ELA Elaborative Leadership
ENC Encouraging Leadership
COL Collaborative Activities
CON Teacher - Student Connections
SPI Group Spirit
INC Inciting Values
VAL Job Value
INF Influence on Job

An Example of SEM: Commitment to Work and Organization

Figure 16. Hypothesized Measurement Model



An Example of SEM: Commitment to Work and Organization

- Model Identification
  - First we examine necessary condition (quantitative approach) for identification by comparing the number of data points to the number of parameters to be estimated.
  - With 10 observed variables there are 10(10+1)/2 = 55 data points.
  - The hypothesized model in Figure 16 indicates that 25 parameters are to be estimated.
  - The model is over-identified with df 30 (55 - 25).
- The necessary and sufficient condition for identification is filled when each observed variable measures one and only one latent variable and one observed variable per latent factor is fixed (Jöreskog, 1979, pp. 191-197).
  - Fixed variables are indicated with red asterisks in Figure 16.

An Example of SEM: Commitment to Work and Organization

- Preliminary Analysis of the Data
  - *Sample size* should be at least 100 units, preferably more than 200.
  - This demand is due to the fact that parameter estimates (ML) and chi-square tests of fit are sensitive to sample size.
  - One should notice that with smaller sample sizes the generalized least-squares method (GLS) is still applicable.
  - Our data has 319 observations, so we may continue with standard settings.

An Example of SEM: Commitment to Work and Organization

- Preliminary Analysis of the Data
  - *Missing data* is another problem, but fortunately with several solutions since researcher may
    - delete cases or variables,
    - estimate missing data,
    - use a missing data correlation matrix, or
    - treat missing data as data. (Tabachnick et al., 1996, pp. 62-65.)
  - We applied list wise deletion since the sample size was adequate for statistical operations (N = 325 was reduced to N = 319 observations).

An Example of SEM: Commitment to Work and Organization

- Preliminary Analysis of the Data
  - *Outliers* are cases with out-of-range values due to
    - incorrect data entry (researcher's mistake or misunderstanding)
    - false answer (respondent's mistake or misunderstanding),
    - failure to specify missing value codes in a statistical software (researcher's mistake).
  - One can detect the most obvious *univariate* outliers by observing min./max. values of summary statistics (Table 7).

An Example of SEM: Commitment to Work and Organization

Table 7. Univariate Summary Statistics for Continuous Variables

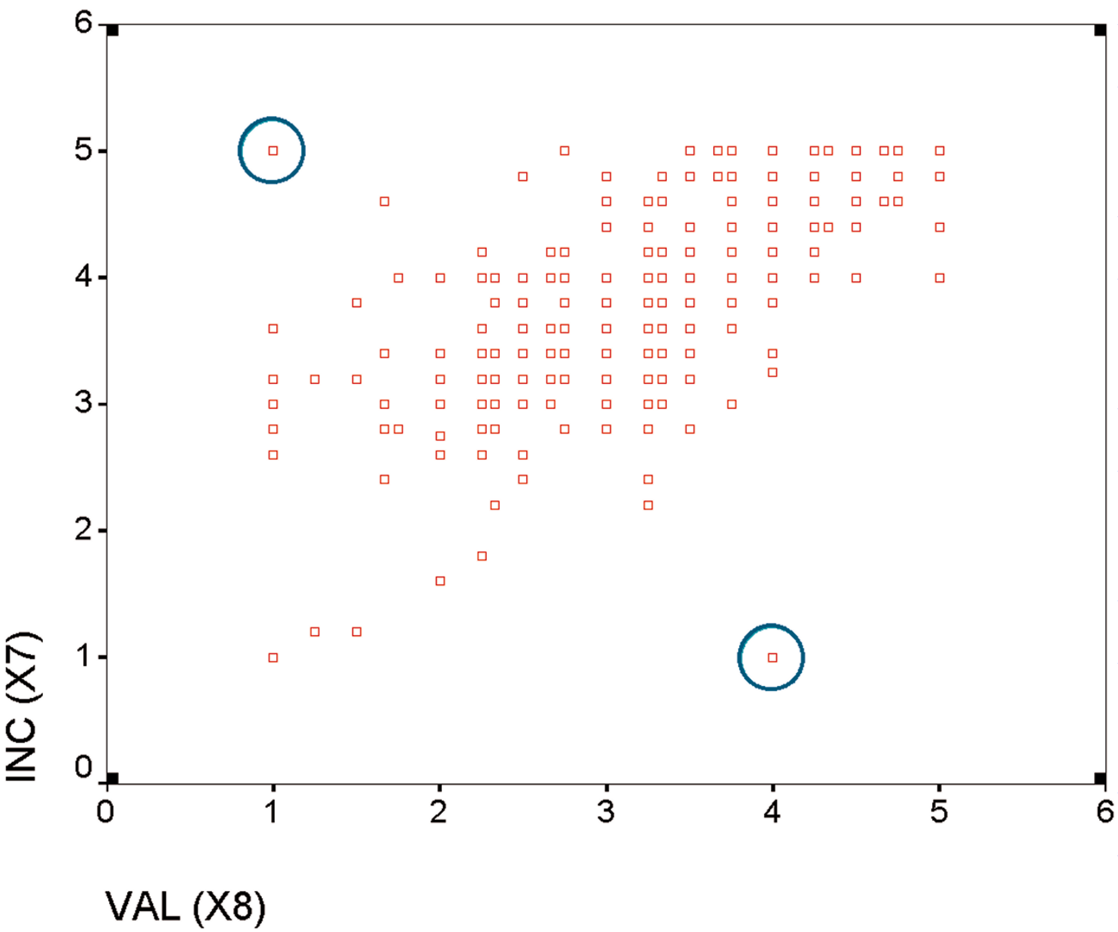
VARIABLE		MEAN	ST. DEV.	SKEWNESS	KURTOSIS	MIN.	FREQ.	MAX.	FREQ.
CO	Y1	3.693	.770	-.353	-.263	1.250	1	5.000	18
PAR	X1	3.485	.856	-.499	.188	1.000	5	5.000	14
ELA	X2	3.184	.817	-.213	.092	1.000	5	5.000	9
ENC	X3	3.284	.968	-.224	-.299	1.000	8	5.000	24
COL	X4	3.426	.678	.064	.021	1.000	1	5.000	4
CON	X5	3.487	.622	.312	.121	1.000	1	5.000	9
SPI	X6	3.280	.743	-.259	.546	1.000	5	5.000	7
INC	X7	3.827	.775	-.651	.725	1.000	2	5.000	27
VAL	X8	3.223	.866	-.247	-.105	1.000	7	5.000	9
INF	X9	3.633	.848	-.521	.285	1.000	4	5.000	24

An Example of SEM: Commitment to Work and Organization

- Preliminary Analysis of the Data
  - A more exact (but tedious!) way to identify possible *bivariate outliers* is to produce scatter plots.
    - Figure 17 is produced with SPSS (Graphs – Interactive – Dot).

An Example of SEM: Commitment to Work and Organization

Figure 17. Bivariate Scatterplot



An Example of SEM: Commitment to Work and Organization

- Preliminary Analysis of the Data
  - *Multivariate normality* is the assumption that each variable and all linear combinations of the variables are normally distributed.
  - When previously described assumption is met, the residuals are also normally distributed and independent.
  - This is important when carrying out SEM analysis.
  - Histograms provide a good graphical look into data (Table 8) to seek for skewness.

An Example of SEM: Commitment to Work and Organization

Table 8. Histograms for Continuous Variables

FRQ	PER	LOW. CLASS LIMIT		FRQ	PER	LOW. CLASS LIMIT	
Y1				X5			
2	.6	1.250	▪	1	.3	1.000	
7	2.2	1.625	▪▪▪▪	0	.0	1.400	
3	.9	2.000	▪	1	.3	1.800	
26	8.2	2.375	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	8	2.5	2.200	▪▪
38	11.9	2.750	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	132	41.4	2.600	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪
43	13.5	3.125	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	21	6.6	3.000	▪▪▪▪▪▪
65	20.4	3.500	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	55	17.2	3.400	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪
64	20.1	3.875	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	52	16.3	3.800	▪▪▪▪▪▪▪▪▪▪▪▪▪▪
37	11.6	4.250	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	36	11.3	4.200	▪▪▪▪▪▪▪▪▪▪
34	10.7	4.625	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	13	4.1	4.600	▪▪▪
X1				X6			
7	2.2	1.000	▪▪	5	1.6	1.000	▪
7	2.2	1.400	▪▪	5	1.6	1.400	▪
11	3.4	1.800	▪▪▪▪	12	3.8	1.800	▪▪▪▪
17	5.3	2.200	▪▪▪▪▪▪▪	29	9.1	2.200	▪▪▪▪▪▪▪▪▪▪
9	2.8	2.600	▪▪▪	101	31.7	2.600	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪
91	28.5	3.000	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	25	7.8	3.000	▪▪▪▪▪▪▪▪▪
50	15.7	3.400	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	72	22.6	3.400	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪
54	16.9	3.800	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	35	11.0	3.800	▪▪▪▪▪▪▪▪▪▪▪▪▪
44	13.8	4.200	▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪	26	8.2	4.200	▪▪▪▪▪▪▪▪▪
29	9.1	4.600	▪▪▪▪▪▪▪▪▪▪▪▪	9	2.8	4.600	▪▪▪

An Example of SEM: Commitment to Work and Organization

- Preliminary Analysis of the Data
  - By examining the Table 9 we notice that distribution of variables X_1 and X_9 is negatively skewed.
  - Furthermore, observing skewness values (Table 9) we see that bias is statistically significant (X_1= -2.610, p = .005; X_9= -2.657, p=.004 and X_7= -2.900, p = .002).

An Example of SEM: Commitment to Work and Organization

Table 9. Test of Univariate Normality for Continuous Variables

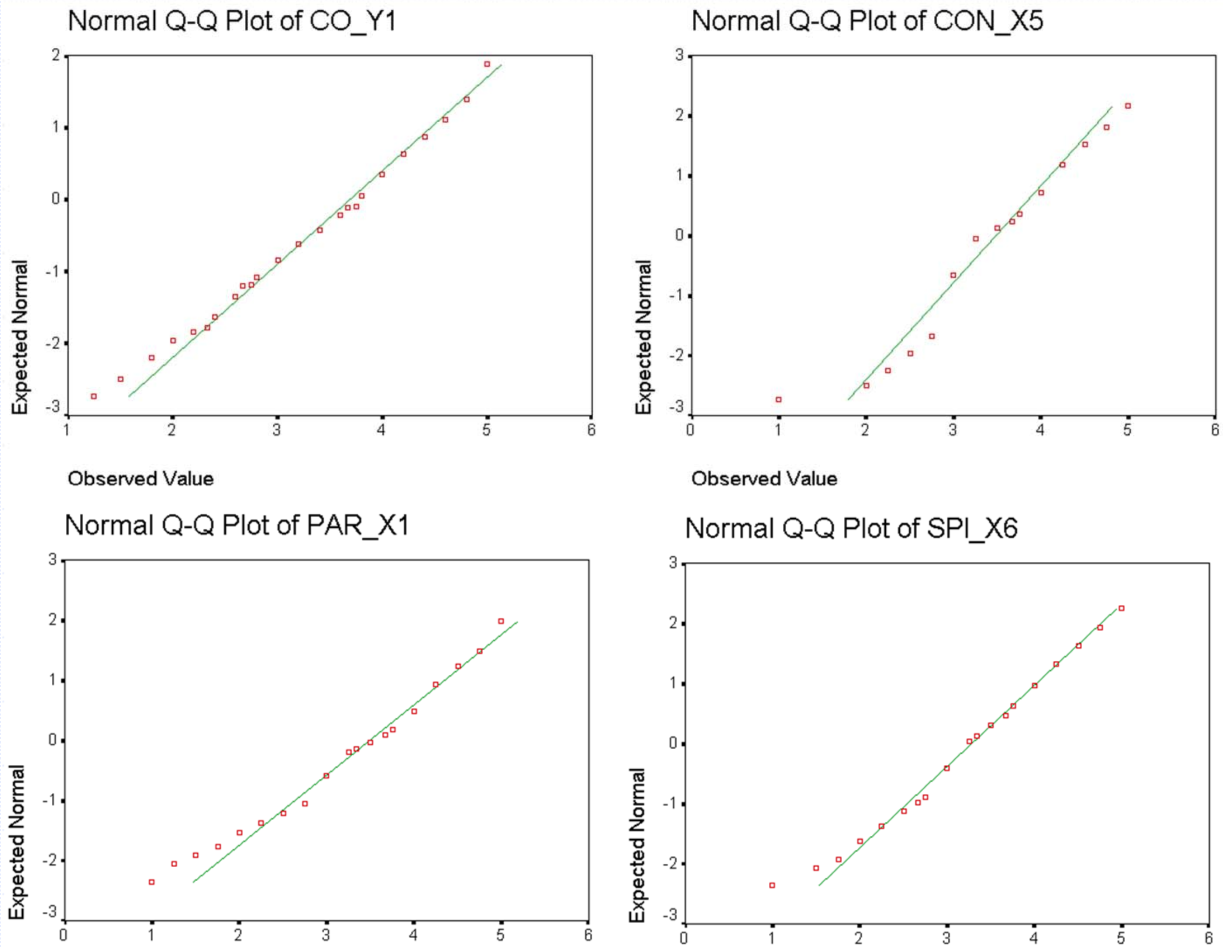
		SKEWNESS		KURTOSIS		SKEWNESS AND KURTOSIS	
VARIABLE		Z-SCORE	P-VALUE	Z-SCORE	P-VALUE	CHI-SQUARE	P-VALUE
CO	Y1	-2.237	.013	-.926	.177	5.860	.053
PAR	X1	-2.610	.005	.840	.200	7.517	.023
ELA	X2	-1.712	.043	.525	.300	3.205	.201
ENC	X3	-1.761	.039	-1.105	.135	4.322	.115
COL	X4	.698	.242	.273	.393	.562	.755
CON	X5	2.103	.018	.624	.266	4.814	.090
SPI	X6	-1.909	.028	1.833	.033	7.004	.030
INC	X7	-2.900	.002	2.244	.012	13.444	.001
VAL	X8	-1.862	.031	-.215	.415	3.513	.173
INF	X9	-2.657	.004	1.138	.128	8.352	.015

An Example of SEM: Commitment to Work and Organization

- Preliminary Analysis of the Data
  - In large samples (>200), significance level (alpha) is not as important as its actual size and the visual appearance of the distribution (Table 10).
  - Perhaps the most essential thing in this case is that now we *know* the bias and instead of excluding those variables immediately we can monitor them more accurately.
  - Table 10 is produced with SPSS (Analyze – Descriptive Statistics – Q-Q Plots).

An Example of SEM: Commitment to Work and Organization

Table 10. Expected Normal Probability Plot



An Example of SEM: Commitment to Work and Organization

- The final phase of the preliminary analysis is to examine the *covariance* (or *correlation*) *matrix* (Table 11).
  - SPSS: Analyze – Correlate – Bivariate (Options: Cross-product deviances and covariances).

Table 11. Covariance Matrix

.593									
.205	.734								
.320	.343	.668							
.324	.438	.467	.938						
.222	.220	.234	.306	.459					
.098	.104	.037	.165	.182	.387				
.262	.275	.307	.391	.308	.114	.552			
.350	.170	.262	.381	.239	.213	.259	.601		
.358	.264	.327	.505	.333	.230	.361	.421	.750	
.337	.275	.338	.443	.298	.182	.345	.434	.496	.719

An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - The model is estimated here by using LISREL 8 to demonstrate textual programming, in the computer exercises, we use AMOS 5 to demonstrate graphical programming.
    - Naturally, both programs lead to similar results.

An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - The LISREL input file is presented in Table 12.
    - SIMPLIS language was applied on the LISREL 8 engine to program the problem.

Table 12. LISREL 8 Input

```
              L I S R E L  8.03

                      BY

         KARL G JORESKOG AND DAG SORBOM


(c) 1999, NONE -project and hopeno@uta.fi

 OBSERVED VARIABLES
 y1 x1 - x9
 COVARIANCE MATRIX FROM FILE hamkk.cov
 SAMPLE SIZE 319
 LATENT VARIABLES
 COM SUP FUN STI
 RELATIONSHIPS
 COM = SUP FUN STI
 x1 = 1*SUP
 x2 x3 = SUP
 x4 = 1*FUN
 x5 x6 = FUN
 x7 = 1*STI
 x8 x9 = STI
 y1 = 1*COM
 LET THE ERROR VARIANCE FOR Y1 BE .31
 LISREL OUTPUT ALL
 PATH DIAGRAM
 END OF PROBLEM
```

An Example of SEM: Commitment to Work and Organization

Model Estimation

- Table 13 lists each matrix specified in the model numbering free parameters ($N = 25$).
- Since the free parameters are numbered successively, we can calculate the degrees of freedom: $10(10+1)/2 = 55$ variances and covariances, and 25 free parameters, resulting in $55 - 25 = 30$ degrees of freedom.
- The model estimates (Maximum Likelihood) are represented in Table 14.

Table 13. Parameter Specifications

LAMBDA X

	SUP	FUN	STI
x1	0	0	0
x2	1	0	0
x3	2	0	0
x4	0	0	0
x5	0	3	0
x6	0	4	0
x7	0	0	0
x8	0	0	5
x9	0	0	6

GAMMA

	SUP	FUN	STI
COM	7	8	9

PHI

	SUP	FUN	STI
SUP	10		
FUN	11	12	
STI	13	14	15

PSI

COM
16

THETA EPS

y1
0

THETA DELTA

x1	x2	x3	x4	x5	x6
17	18	19	20	21	22

THETA DELTA

x7	x8	x9
23	24	25

An Example of SEM: Commitment to Work and Organization

- Model Estimation

Table 14. Model Estimates

LAMBDA Y

	COM
y1	1.00

LAMBDA X

	SUP	FUN	STI
x1	1.00	.00	.00
x2	1.12	.00	.00
x3	1.49	.00	.00
x4	.00	1.00	.00
x5	.00	.54	.00
x6	.00	1.10	.00
x7	.00	.00	1.00
x8	.00	.00	1.20
x9	.00	.00	1.16

GAMMA

	SUP	FUN	STI
COM	.22	-.15	.82

COVARIANCE MATRIX OF ETA AND KSI

	COM	SUP	FUN	STI
COM	.28			
SUP	.24	.29		
FUN	.23	.22	.28	
STI	.31	.25	.27	.36

PHI

	SUP	FUN	STI
SUP	.29		
FUN	.22	.28	
STI	.25	.27	.36

PSI

COM
.01

An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - Table 15 contains measures of fit of the model.
    - The chi-square (χ^2) tests the hypothesis that the factor model is adequate for the data.
      - Non-significant χ^2 is desired which is true in this case (p >.05) as it implies that the model and the data are *not* statistically significantly different.
      - Goodness of Fit Index (GFI) is good for the model with the value of .92 (should be >.90).
        - However, the *adjusted* GFI goes below the .90 level indicating the model is not perfect.
      - The value of Root Mean Square Residual (RMSR) should be as small as possible, the value of .03 indicates good-fitting model.

An Example of SEM: Commitment to Work and Organization

- Model Estimation

Table 15. Goodness of Fit Statistics

```
CHI-SQUARE WITH 30 DEGREES OF FREEDOM = 43.31 (P = 0.055)
GOODNESS OF FIT INDEX = 0.92
ADJUSTED GOODNESS OF FIT INDEX = 0.854
ROOT MEAN SQUARE RESIDUAL = 0.0299
CROSS-VALIDATION INDEX = 87.653
```

An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - Standardized residuals are residuals divided by their standard errors (Jöreskog, 1989, p. 103).
  - All residuals have moderate values (min. -2.81, max. 2.28), which means that the model estimates adequately relationships between variables.
  - QPLOT of standardized residuals is presented in Table 16 where a *x* represents a single point, and an * multiple points.

An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - The plot provides visual way of examining residuals; steeper plot (than diagonal line) means good fit and shallower means opposite.
  - If residuals are normally distributed the x's are around the diagonal.
  - Non-linearities are indicators of specification errors in the model or of unnormal distributions.
    - We can see from the Table 16 that plotted points follow the diagonal and there are neither outliers nor non-linearity.

An Example of SEM: Commitment to Work and Organization

- Model Estimation

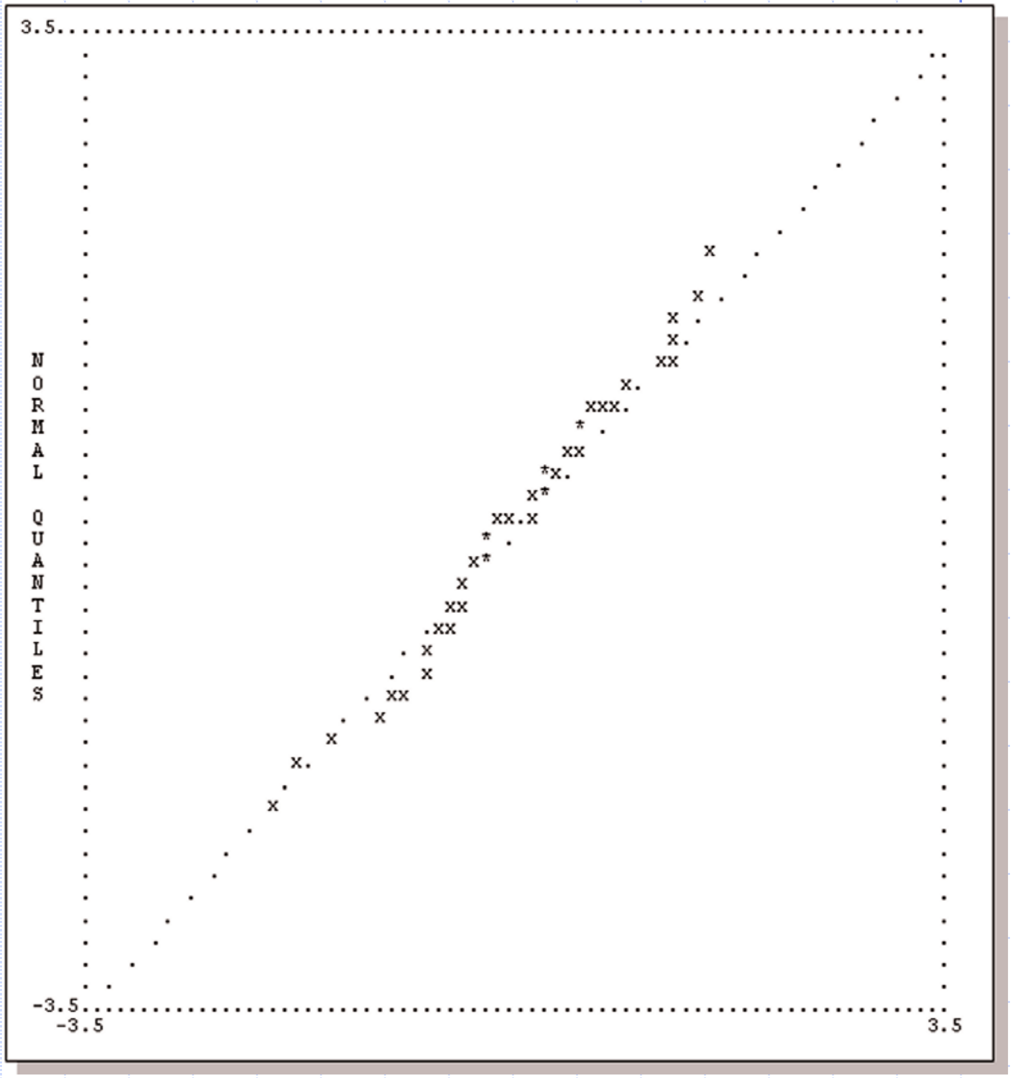


Table 16. QPLOT of Standardized Residuals

An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - The standard errors show how accurately the values of the free parameters have been estimated (Jöreskog, 1989, p. 105) in the model.
  - Standard errors should be small, as seen in Table 17 (min. .05, max. .35).

An Example of SEM: Commitment to Work and Organization

- Model Estimation

Table 17. Standard Errors

LAMBDA X

	SUP	FUN	STI
x1	.00	.00	.00
x2	.28	.00	.00
x3	.35	.00	.00
x4	.00	.00	.00
x5	.00	.18	.00
x6	.00	.21	.00
x7	.00	.00	.00
x8	.00	.00	.20
x9	.00	.00	.19

GAMMA

	SUP	FUN	STI
COM	.40	.56	.47

PHI

	SUP	FUN	STI
SUP	.13		
FUN	.08	.09	
STI	.09	.08	.12

PSI

COM
.07

THETA DELTA

x1	x2	x3	x4	x5	x6
.10	.08	.10	.05	.06	.06

THETA DELTA

x7	x8	x9
.06	.07	.06

An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - A T-value is produced for each free parameter in the model by dividing its parameter estimate by its standard error.
    - T-values between -1.96 and 1.96 are not statistically significant.
  - Table 18 proves our second hypothesis about significant covariances between latent Xi - variables (IV's in the model) since T-values indicate that the covariances are significantly different from zero.

An Example of SEM: Commitment to Work and Organization

- Model Estimation

Table 18. T-values

LAMBDA X

	SUP	FUN	STI
x1	.00	.00	.00
x2	4.03	.00	.00
x3	4.29	.00	.00
x4	.00	.00	.00
x5	.00	3.00	.00
x6	.00	5.12	.00
x7	.00	.00	.00
x8	.00	.00	6.02
x9	.00	.00	5.93

GAMMA

	SUP	FUN	STI
COM	.54	-.26	1.74

PHI

	SUP	FUN	STI
SUP	2.26		
FUN	2.90	2.97	
STI	2.95	3.41	3.10

PSI

COM
.17

THETA DELTA

x1	x2	x3	x4	x5	x6
4.35	3.82	2.89	3.44	4.70	3.45

THETA DELTA

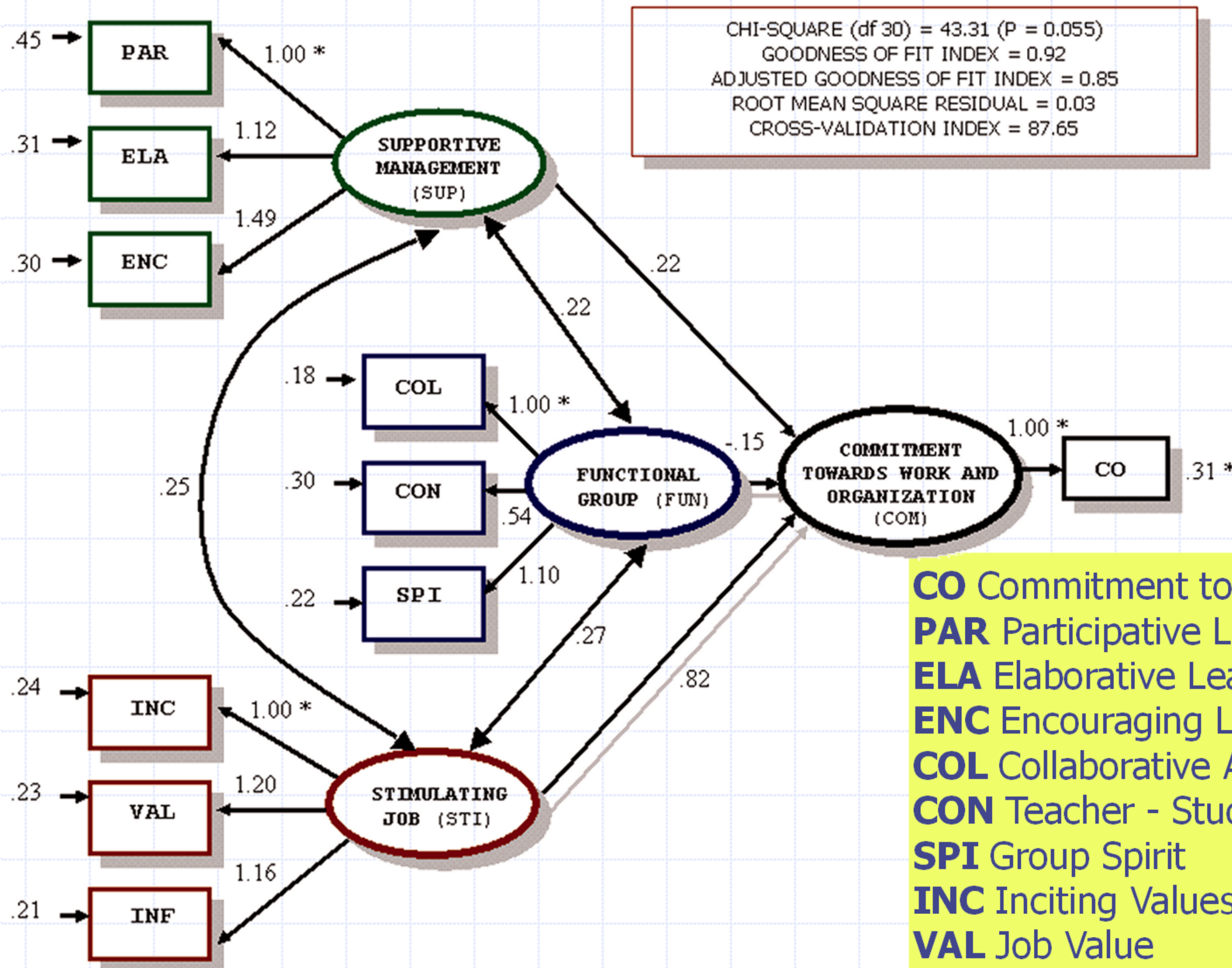
x7	x8	x9
4.03	3.56	3.68

An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - Figure 18 represents estimated "Commitment to Work and Organization" model.
    - Unstandardized coefficients are reported here.
    - Stimulating job increases commitment to work (.82) more than superior's encouragement (.22) or community spirit (-.15).

An Example of SEM: Commitment to Work and Organization

Figure 18. Commitment to Work and Organization Model



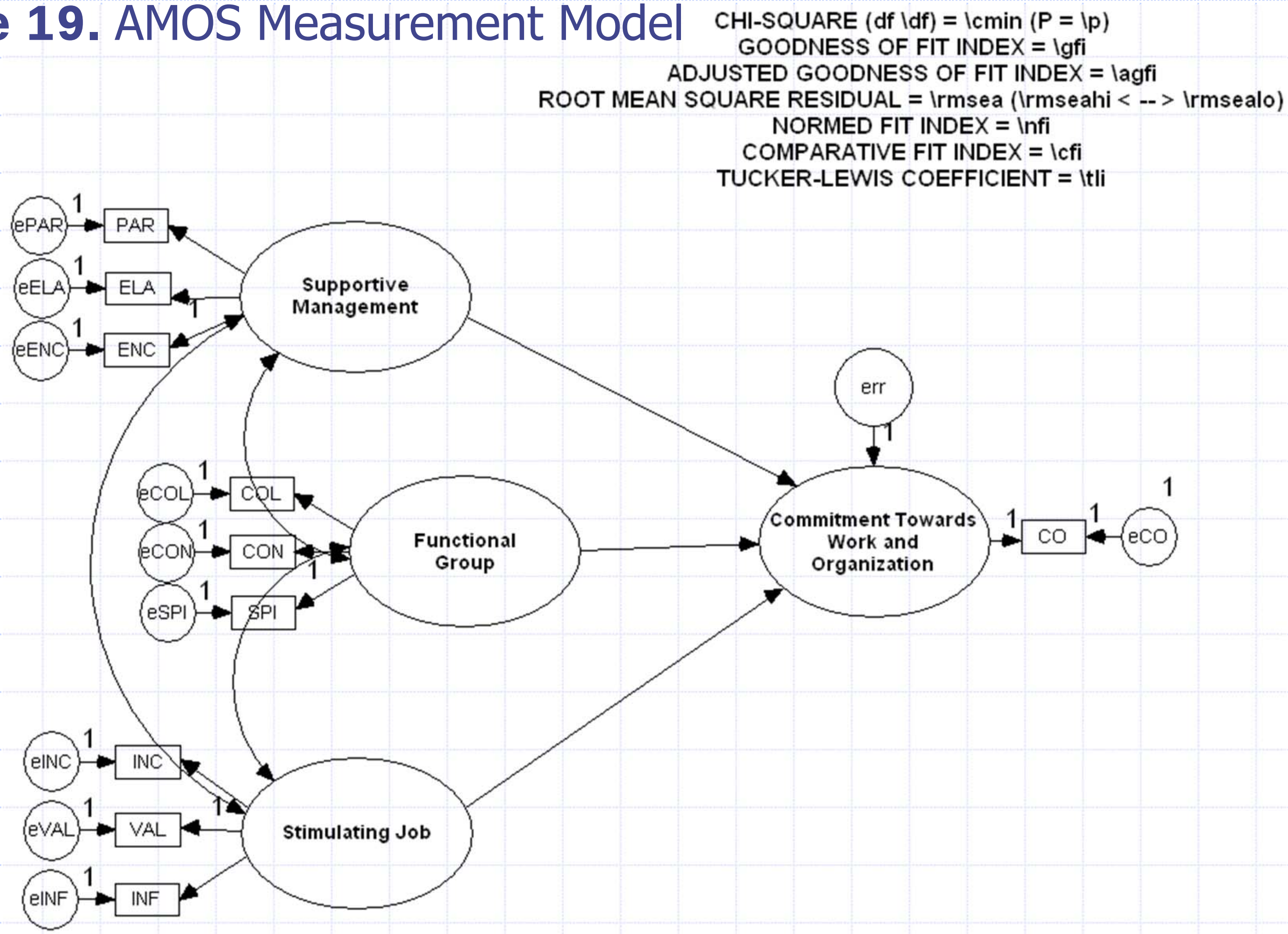
CO Commitment to work and organization
PAR Participative Leadership
ELA Elaborative Leadership
ENC Encouraging Leadership
COL Collaborative Activities
CON Teacher - Student Connections
SPI Group Spirit
INC Inciting Values
VAL Job Value
INF Influence on Job

An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - Figures 19 and 20 represent the same model before and after AMOS 5 analysis.
    - AMOS uses SPSS data matrix as an input file.

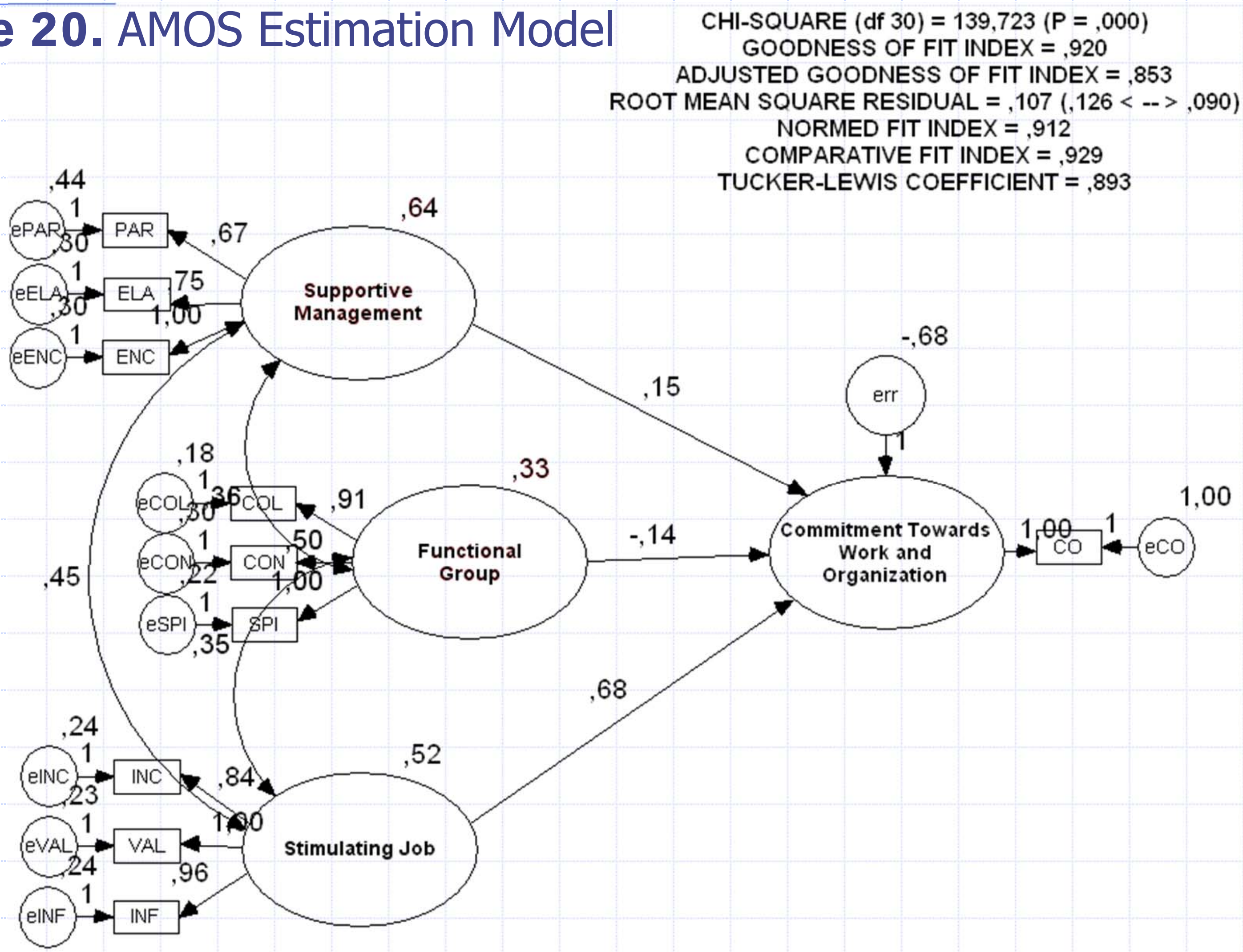
An Example of SEM: Commitment to Work and Organization

Figure 19. AMOS Measurement Model



An Example of SEM: Commitment to Work and Organization

Figure 20. AMOS Estimation Model



An Example of SEM: Commitment to Work and Organization

- Model Estimation
  - Naturally, both LISREL and AMOS produce similar results:
    - Unstandardized coefficients are reported here.
    - Stimulating job increases in both models commitment to work (.82/.68) more than superior's encouragement (.22/.15) or community spirit (-.15/-.14).

Contents

Conclusions

- SEM has proven to be a very versatile statistical toolbox for educational researchers when used to confirm theoretical structures.
- Perhaps the greatest strength of SEM is the requirement of a prior knowledge of the phenomena under examination.
  - In practice, this means that the researcher is testing a theory which is based on an exact and explicit plan or design.
  - One may also notice that relationships among factors examined are free of measurement error because it has been estimated and removed, leaving only common variance.
  - Very complex and multidimensional structures can be measured with SEM; in that case SEM is the only *linear* analysis method that allows complete and simultaneous tests of all relationships.

Conclusions

- Disadvantages of SEM are also simple to point out.
  - Researcher must be very careful with the study design when using SEM for *exploratory* work.
  - As mentioned earlier, the use of the term 'causal modeling' referring to SEM is misleading because there is nothing causal, in the sense of inferring causality, about the use of SEM.
  - SEM's ability to analyze more complex relationships produces more complex models: Statistical language has turned into jargon due to vast supply of analytic software (LISREL, EQS, AMOS).
  - When analyzing scientific reports methodologically based on SEM, usually a LISREL model, one notices that they lack far too often decent identification inspection which is a prerequisite to parameter estimation.

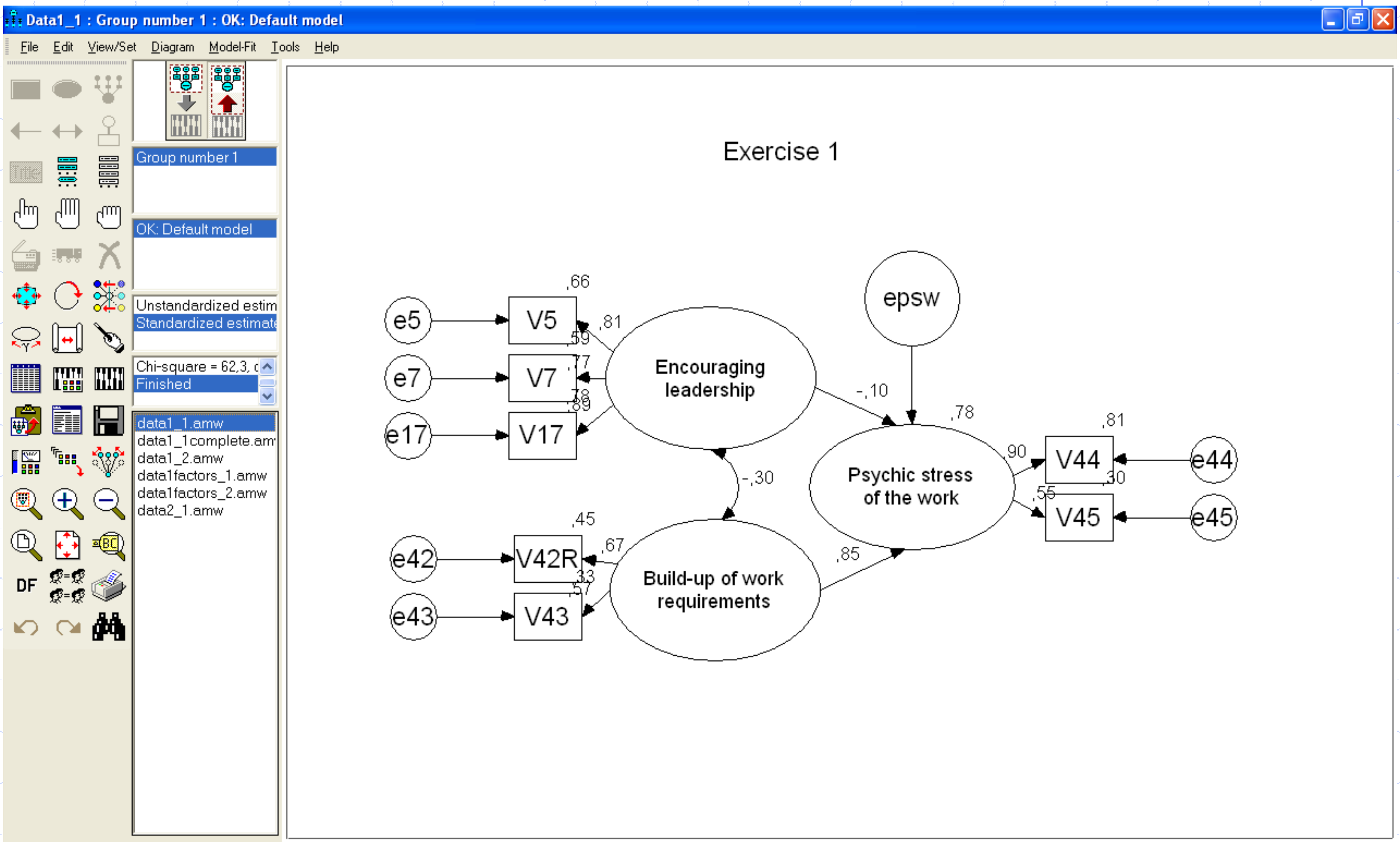
Conclusions

- Overgeneralization is always a problem – but specifically with SEM one must pay extra attention when interpreting causal relationships since *multivariate normality* of the data is assumed.
  - This is a severe limitation of linear analysis in general because the reality is seldom linear.
- We must also point out that SEM is based on covariances that are not stable when estimated from small (<200 observation) samples.
- On the other hand, too large (>200 observations) sample size is also a reported problem (e.g., Bentler et al., 1983) of the significance of χ^2.

Conclusions

- SEM programs allow calculation of modification indices which help researcher to fit the model to the data.
  - Added or removed dependencies must be based on theory!
  - Overfitting model to the data reduces generalizability!
- Following slides demonstrate the effect of sample size and model modification (according to modification indices).
  - Example 2 in the course exercise booklet.

Data1_1.amw (Exercise 2)



Data1_1.amw (Exercise 2)

- Large sample (n=447) produces biased χ^2/df and p values (both too large).
- Model fit indices are satisfactory at best (RMSEA > .10, TLI <.90).
  - As there are missing values in the data, calculation of modification indices is not allowed (in AMOS).

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	24	62,250	11	,000	5,659
Saturated model	35	,000	0		
Independence model	7	1129,189	28	,000	40,328

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	,945	,860	,954	,882	,953
Saturated model	1,000		1,000		1,000
Independence model	,000	,000	,000	,000	,000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	,102	,078	,128	,000
Independence model	,297	,282	,312	,000

Smaller randomized sample with no missing values, modified model

- Replace missing values with series mean:
  - SPSS: Transform – Replace missing values – Series mean.
- Produce a smaller (n=108) randomized subsample:
  - SPSS: Data – Select cases – Random sample of cases – Approximately 20% of cases.
- Produce modification indices analysis:
  - AMOS: View/set – Analysis properties – Modification indices.

Amos Output

data1_1.amw
- Analysis Summary
- Notes for Group
- Variable Summary
- Parameter summary
- Assessment of normality
- Observations farthest from th
- Sample Moments
- Notes for Model
- Estimates
- Modification Indices
- Minimization History
- Pairwise Parameter Comparis
- Model Fit
- Execution Time

Default model

Modification Indices (Group number 1 - Default model)

Covariances: (Group number 1 - Default model)

			M.I.	Par Change
e45	<-->	Encouraging_leadership	21,150	-,251
e44	<-->	Encouraging_leadership	5,575	,112
e42	<-->	Encouraging_leadership	5,704	-,120
e43	<-->	Encouraging_leadership	8,949	,148
e43	<-->	e45	9,722	-,144
e43	<-->	e44	5,604	,092
e7	<-->	Build-up of work_requirements	7,317	-,084
e7	<-->	e45	5,921	-,090
e7	<-->	e42	5,474	-,081
e17	<-->	Build-up of work_requirements	5,458	,072
e17	<-->	e43	7,604	,093

Variances: (Group number 1 - Default model)

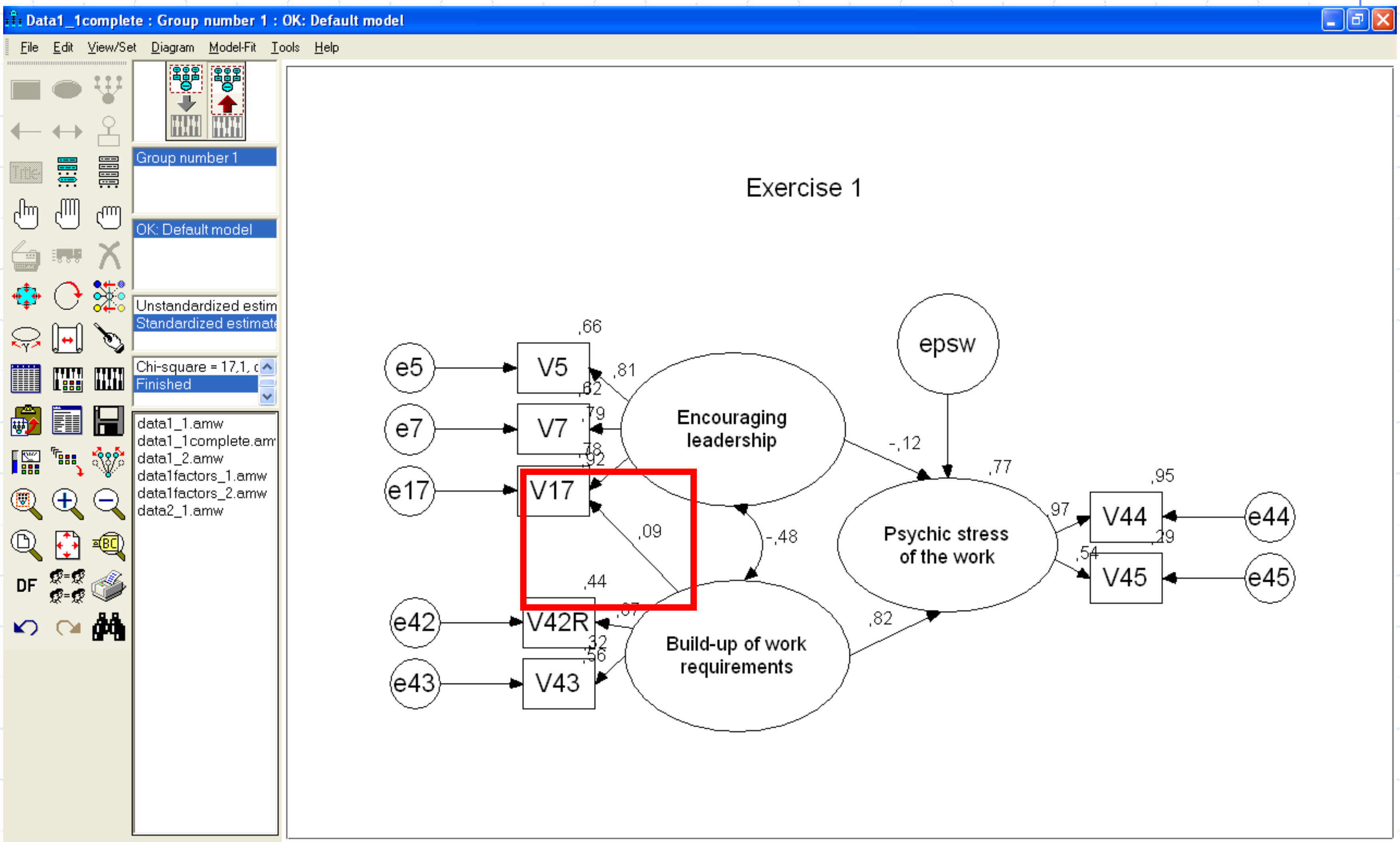
	M.I.	Par Change

Regression Weights: (Group number 1 - Default model)

			M.I.	Par Change
V45	<---	Encouraging_leadership	17,248	-,203
V44	<---	Encouraging_leadership	4,583	,092
V42R	<---	Encouraging_leadership	5,000	-,101
V43	<---	Encouraging_leadership	7,819	,124
V7	<---	Build-up of work_requirements	6,347	-,134
V17	<---	Build-up of work_requirements	4,773	,116

A new path is added to the model.

Modified model



Model Fit Summary

CMIN

NEW MODEL | OLD MODEL

Model	NPAR	CMIN	DF	P	CMIN/DF	NPAR	CMIN	DF	P	CMIN/DF
Default model	25	17,148	10	,071	1,715	24	62,250	11	,000	5,659
Saturated model	35	,000	0			35	,000	0		
Independence model	14	301,823	21	,000	14,373	7	1129,189	28	,000	40,328

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	,943	,881	,976	,947	,975	,945	,860	,954	,882	,953
Saturated model	1,000		1,000		1,000	1,000		1,000		1,000
Independence model	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE	RMSEA	LO 90	HI 90	PCLOSE
Default model	,082	,000	,146	,193	,102	,078	,128	,000
Independence model	,354	,319	,389	,000	,297	,282	,312	,000

Contents

References

- Arbuckle, J., & Wothke, W. (1999). *Amos 4.0 User's Guide*. Chicago: SPSS Inc.
- Bollen, K. (1989). *Structural Equations with Latent Variables.* New York: John Wiley & Sons.
- Hair, J. F., Anderson, R. E., Tatham R. L., & Black, W. C. (1995). *Multivariate Data Analysis* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

References

- Kaplan, D. (2000). *Structural Equation Modeling*. Thousand Oaks: Sage.
- Long, J. (1983). *Confirmatory Factor Analysis*. California: Sage.
- Muthén, L., & Muthén, B. (2000). *MPLUS User Manual*. Los Angeles: Muthén & Muthén.
- Nokelainen, P., & Ruohotie, P. (1999). Structural Equation Modeling in Professional Growth Research. In P. Ruohotie, H. Tirri, P. Nokelainen, & T. Silander (Eds.), *Modern Modeling of Professional Growth, vol. 1* (pp. 121-154). Hämeenlinna: RCVE.

References

- Nokelainen, P., & Ruohotie, P. (2009). Non-linear Modeling of Growth Prerequisites in a Finnish Polytechnic Institution of Higher Education. *Journal of Workplace Learning, 21*(1), 36-57.
- Pearl, J. (2000). *Causality.* New York: Cambridge University Press.
- Ruohotie, P. (1996). Professional Growth and Development. In K. Leithwood et al. (Eds.), *International Handbook of Educational Leadership and Administration* (pp. 419-445). Dordrecht: Kluwer Academic Publishers.
- Raykov, T., & Marcoulides, G. (2000). *A First Course in Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

References

- Sager, J., Griffeth, R., & Hom, P. (1998). A Comparison of Structural Models Representing Turnover Cognitions. *Journal of Vocational Behavior*, *53*, 254-273.
- Schumacker, R. E., & Lomax, R. G. (2004). *A Beginner's Guide to Structural Equation Modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tabachnick, B ., & Fidell, L. (1996). *Using Multivariate Statistics*. New York: HarperCollins.
- Wright, S. (1921). Correlation and Causation. *Journal of Agricultural Research*, *20*, 557-585.
- Wright, S. (1934). The Method of Path Coefficients. *The Annals of Mathematical Statistics*, *5*, 161-215.