Structural Equation Modeling (SEM)

Outline

1. *Introduction*
2. The rational
3. The complete structural equation model
4. Types of models
5. The strategy of analysis
6. The evaluation of models

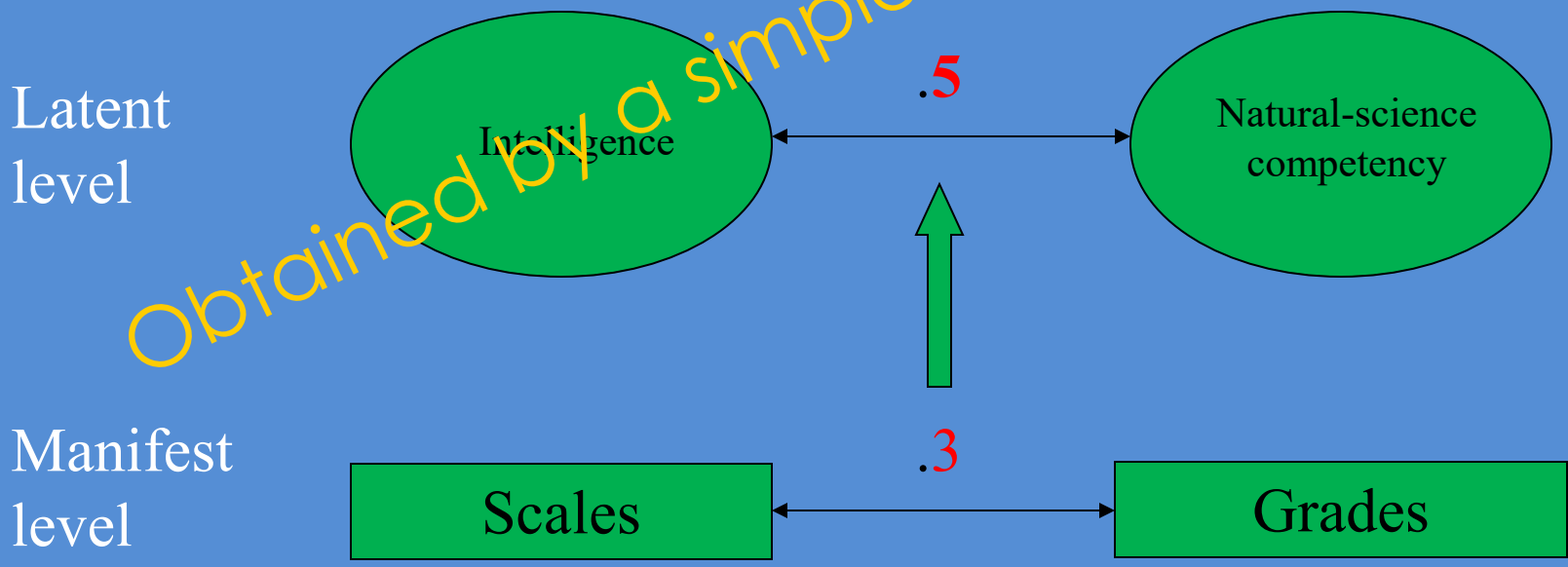
Introduction

- Structural equation modeling (SEM) is a procedure for conducting *structural investigations on the latent level*.
- The separation of error from the true component of measurement implies the switch from the manifest to the latent level.
- There are consequences for the relationships on the latent level!

Slide from the first course unit

The step from the manifest to latent levels

An example: the relationship between *intelligence* and *n-competency* with the transfer to the latent level



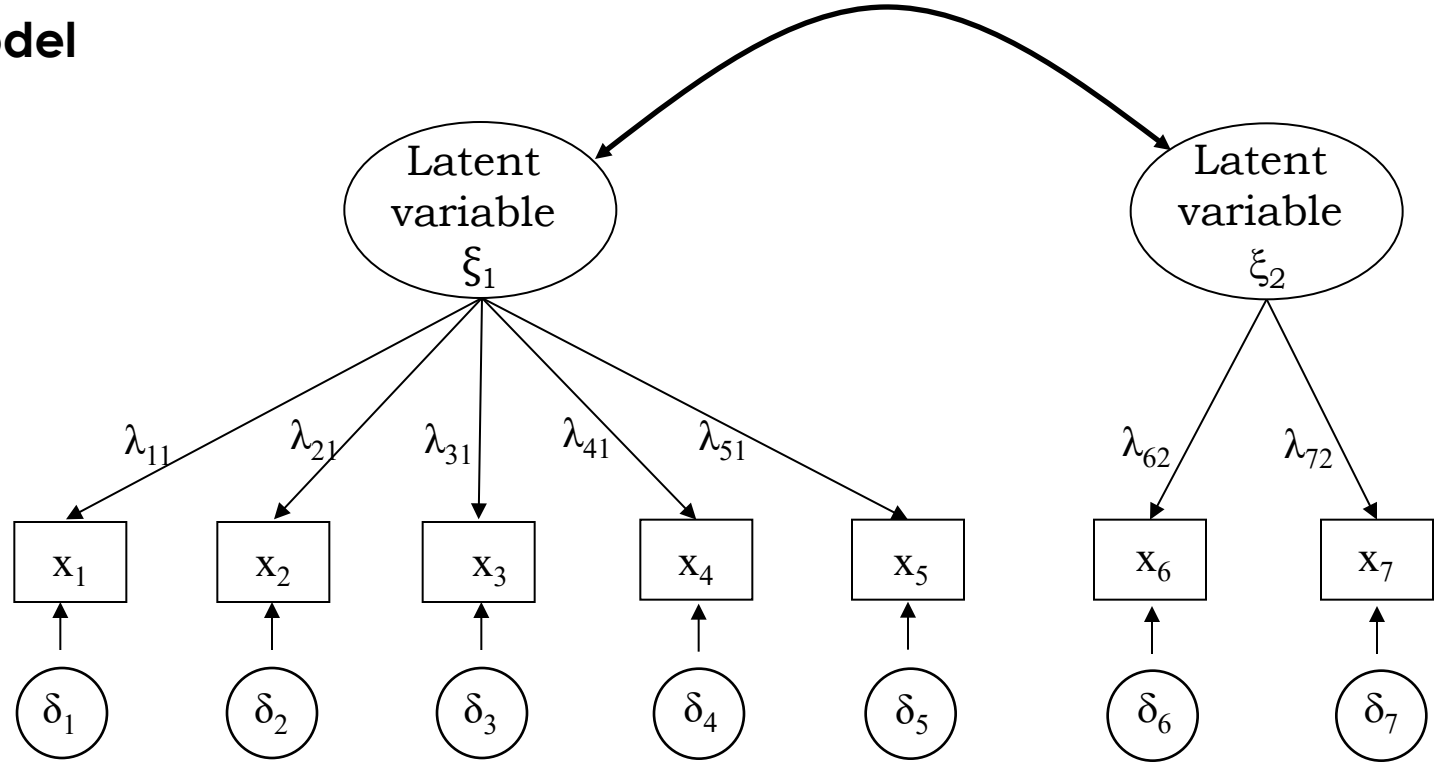
Introduction

- Modern estimation methods provide estimates of factor loadings and **residual variances**
- Elimination of **residual variances** enables the investigation of relationships on the latent level!

Introduction

An example using a sophisticated estimation method

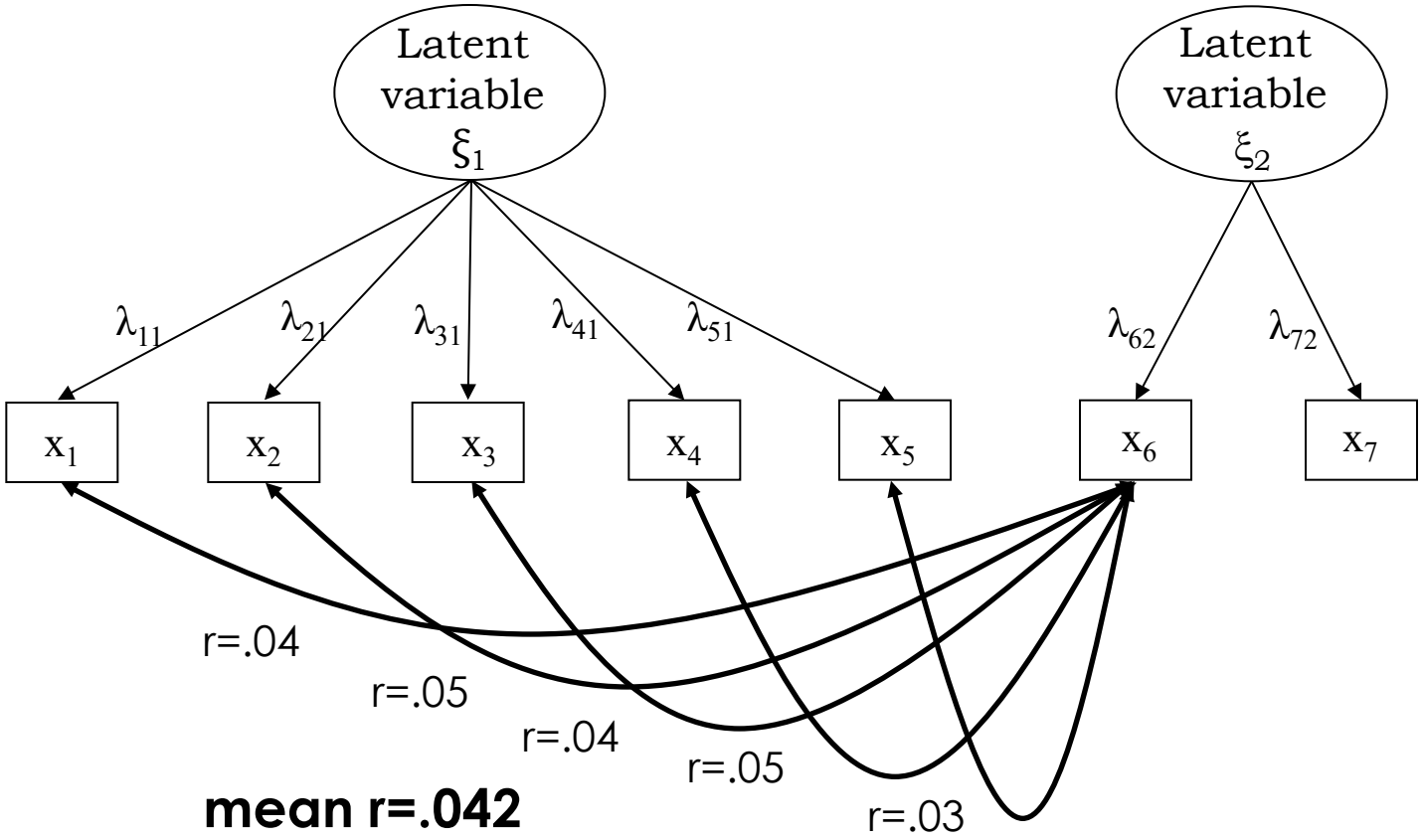
The model



Introduction

An example using a sophisticated estimation method

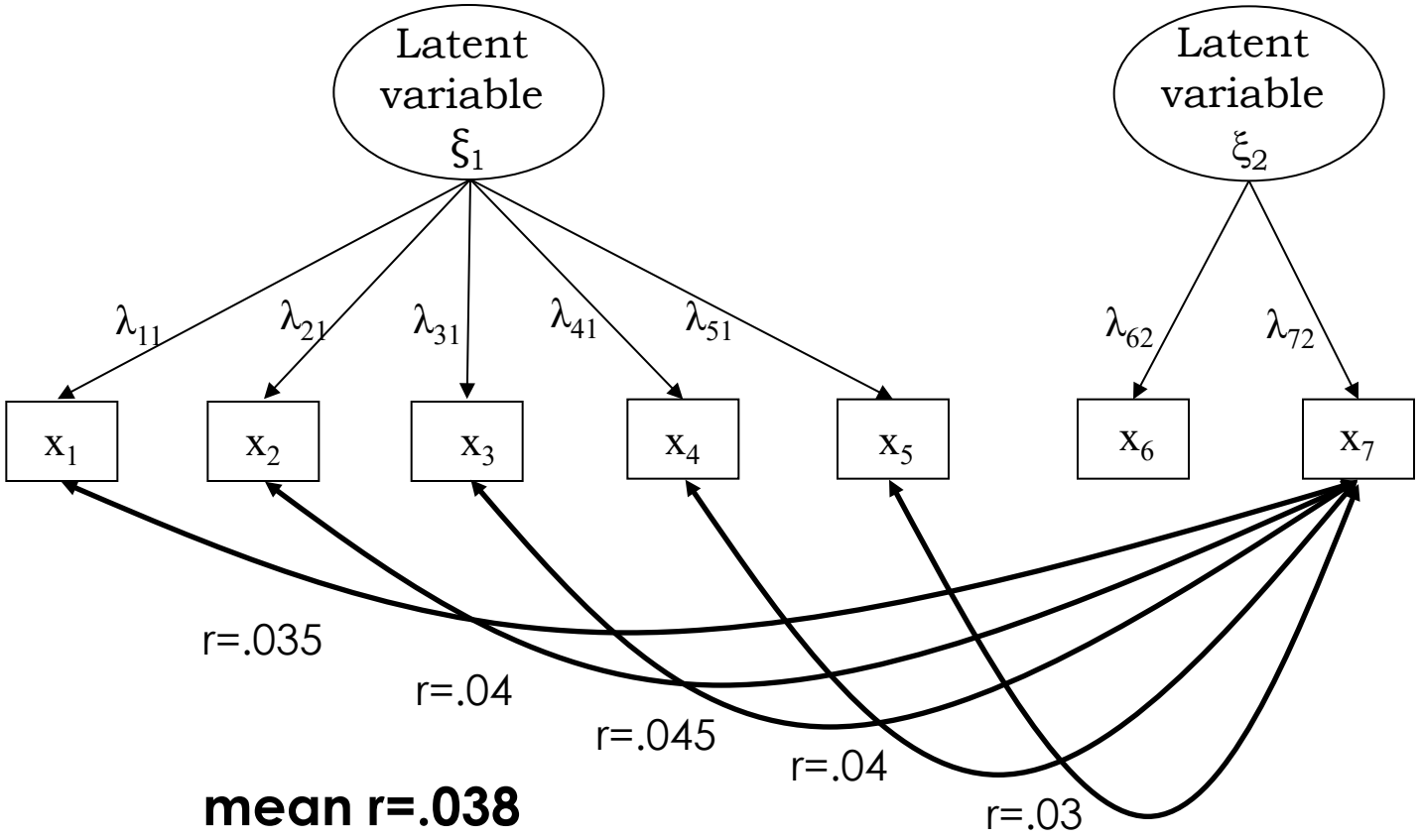
Correlations on the manifest level



Introduction

An example using a sophisticated estimation method: ML estimation

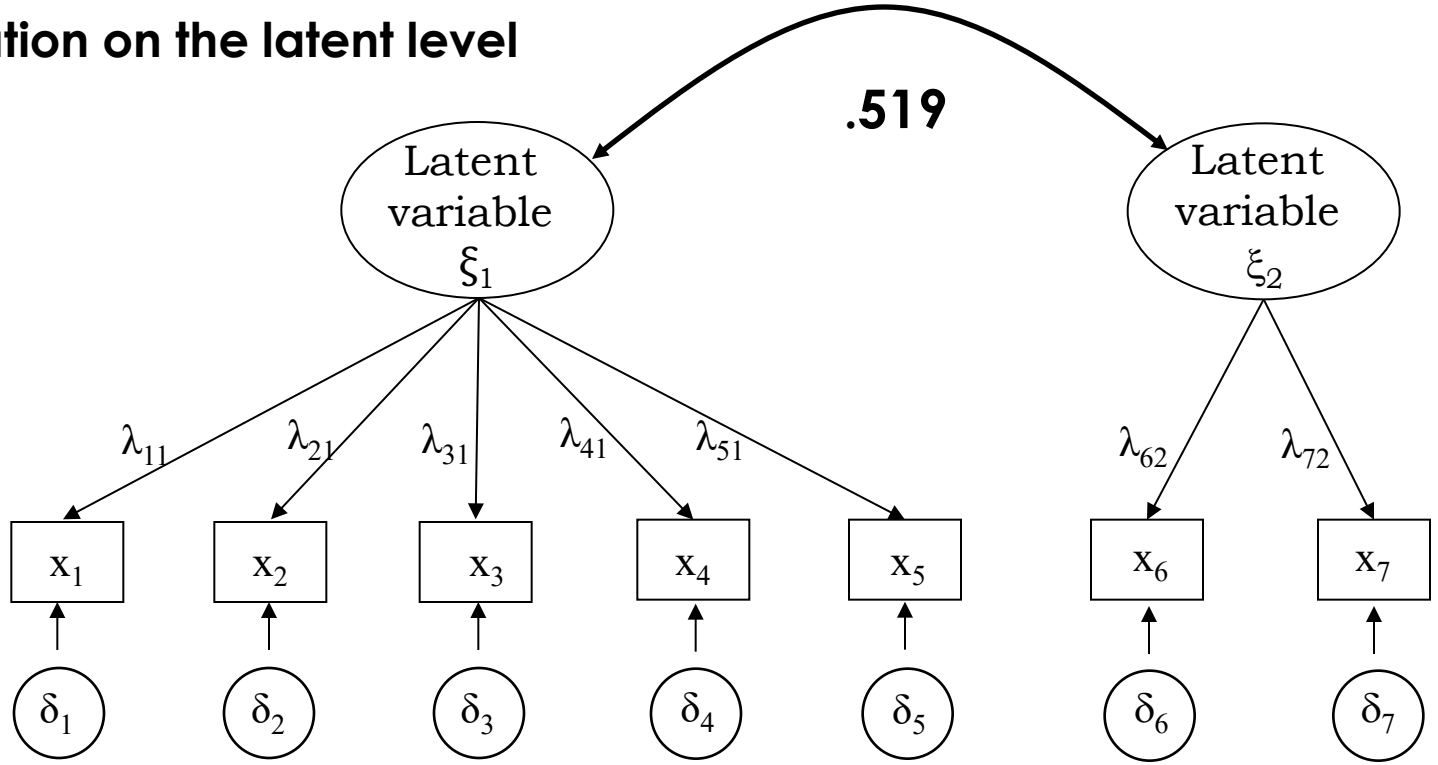
Correlations on the manifest level



Introduction

An example using a sophisticated estimation method：ML estimation

Correlation on the latent level



Overall mean correlation: r=.040

Introduction

- SEM combines several methods; the application of these methods is not apparent to the user.

- Since there is hardly any influence on these methods for the user, SEM teaching is frequently focused on explaining how to do structural equation modeling only.

- Nevertheless, it is good to have an idea of how the results are achieved.

Outline

1. Introduction
2. *The rational*
3. The complete structural equation model
4. Types of models
5. The strategy of analysis
6. The evaluation of models

The rationale

(„**Rationale**“ means „basic idea of something or is a (rough) explanation of **how it works *basically***)

In this section structural equation modeling is described by taking different perspectives for providing an idea of how it works.

Four stages are considered: …

The rationale

(„**Rationale**“ means „basic idea* of something or is a (rough) explanation of **how it works basically**)

1. stage: preparation of Input
2. stage: investigation of measurement models
3. stage: investigation of structure
4. stage: investigation of model fit

Three *perspectives* are taken!

The rationale

The *perspectives* …

1. ***Input-output perspective*** … considering the input and output of a sequence of stages
2. ***EFA/MR perspective*** … doing it by EFA followed by MR (*poor man's SEM*)
3. ***SEM algorithm-based perspective*** … doing it by standard SEM software (*integrated SEM*)

The rationale

.....

1. **<u>Input-output perspective</u>**: the rationale concentrating on **input** and **output**

• It starts with the *data matrix* **M** (input 1) and leads to the *covariance matrix* **S** (output 1):

$$\mathbf{M} = [(\text{Data})] \Rightarrow \mathbf{S} = [(\text{Covariances})]$$

The rationale

.....

1.The rationale concentrating on **input** and **output** of the steps

- ... the *covariance matrix* **S** (input 2) is used for obtaining the *matrix of correlations of factors* **ϕ** (output 2):

$$\mathbf{M} = [(\text{Data})] \Rightarrow \mathbf{S} = [(\text{Covariances})] \Rightarrow \mathbf{\Phi} = [(\text{factor correlations})]$$

The rationale

.....

1.The rationale concentrating on **input** and **output** of the steps

- ... the *matrix of correlations of factors* Φ (input 3) is used for obtaining the *matrix of gamma coefficients* Γ (output 3):

...

$$\Phi = [(\text{factor correlations})] \Rightarrow \Gamma = [(\text{regression weights / path coefficients})]$$

The rationale

.....

1.The rationale concentrating on **input** and **output** of the steps

- ... **all parameter estimates** (input of various steps) for estimation of model fit available (final output):

... $\Gamma = [(\text{regression weights / path coefficients})]$ ➡ Σ so that 。 。 。

Fit estimation

(indices of model fit)

is possible

The rationale

.....

2. **EFA/MR perspective**: explanation concentrating on EFA and MR as methods for investigating the data

The more basic way of conducting EFA:

- it yields factors and factor scores
- correlating the factor scores yield factor correlations

$$\mathbf{M} = [(\text{Data})] \quad \Rightarrow \quad \Xi = [(\text{factor score})]$$

The rationale

.....

2. The rationale using on EFA and MR as methods for investigating the data

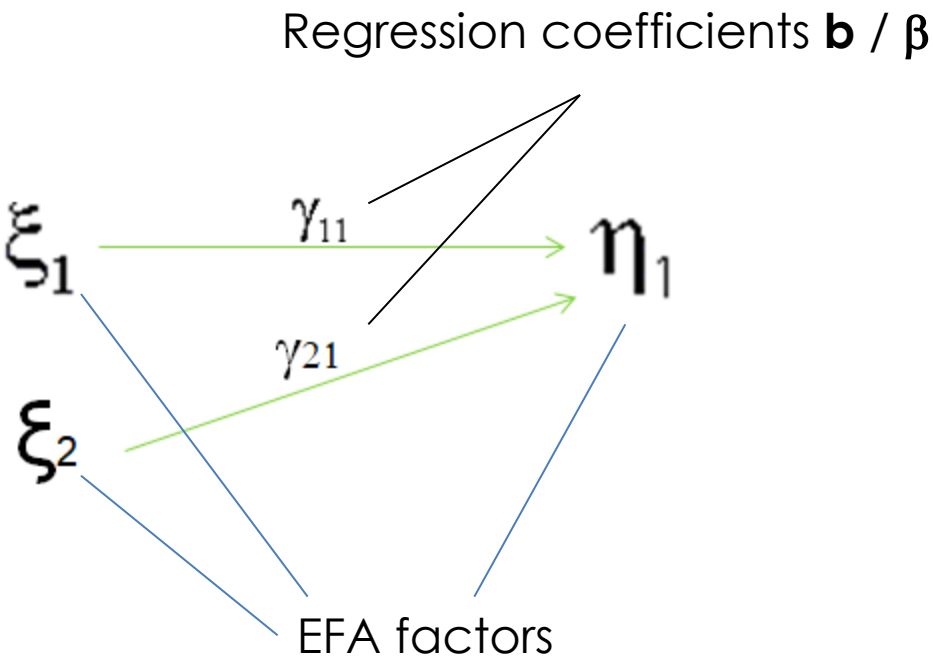
The **factor scores of EFA** can be used as basis for conducting multiple regression (MR). MR yields regression weights.

$\mathbf{M} = [(\text{Data})]$ ⇨ ..... ⇨ $\mathbf{\Gamma} = [(\text{regression weights})]$

The rationale

.....

An illustration:



The rationale

.....

The parameters of the following matrices can be made available this way: $\boldsymbol{\Lambda}$, $\boldsymbol{\theta}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}$ ⇨ $\boldsymbol{\Sigma}$

What remains is the investigation of model fit.

The rationale

2. Finally, the *model-implied covariance matrix* $\boldsymbol{\Sigma}$ [$\boldsymbol{\Sigma}(\boldsymbol{\theta})$] ...

.... for reproducing the *empirical covarince matrix* **S**

- and -

.... for estimating the fit indices

are available

But, there are no easily available procedures for doing it outside of SEM software.

The rationale

.....

3. **<u>SEM algorithm-based perspective</u>**: an algorithm performs *simultaneous* parameter estimation for the whole model

i.e. at first the structure of $\boldsymbol{\Sigma}$ is specified according to the information provided by the path diagram and afterwards all the parameters are estimated *simultaneously*

The rationale

The matrices $\boldsymbol{\Sigma}$ and S are considered as the input to the fit function F( ) :

$$F(\boldsymbol{\Sigma} (\boldsymbol{\theta}), \mathbf{S}) \longrightarrow \text{min}$$

that is minimized. It is minimized by modifying the values of the parameters included in $\boldsymbol{\theta}$.

The rationale

It is an iterative algorithm starting with initial values.

In some software the user can/could provide starting values.

The reached minimum is transformed into a χ^2 statistic.

The rationale

In all perspectives information on parameter estimates is provided.

But, only in the 3th perspective fit indices are made available that enable model evaluation!

The rationale: a demonstration

… from cognitive psychology. It is investigated whether (visual) *working memory* and *executive processes* (shifting) explain *reasoning*.

… using EFA and MR respectively SEM (2. and 3. rationals)

The rationale: a demonstration

… from cognitive psychology. It is investigated whether (visual) *working memory* and *executive processes* (shifting) explain *reasoning*.

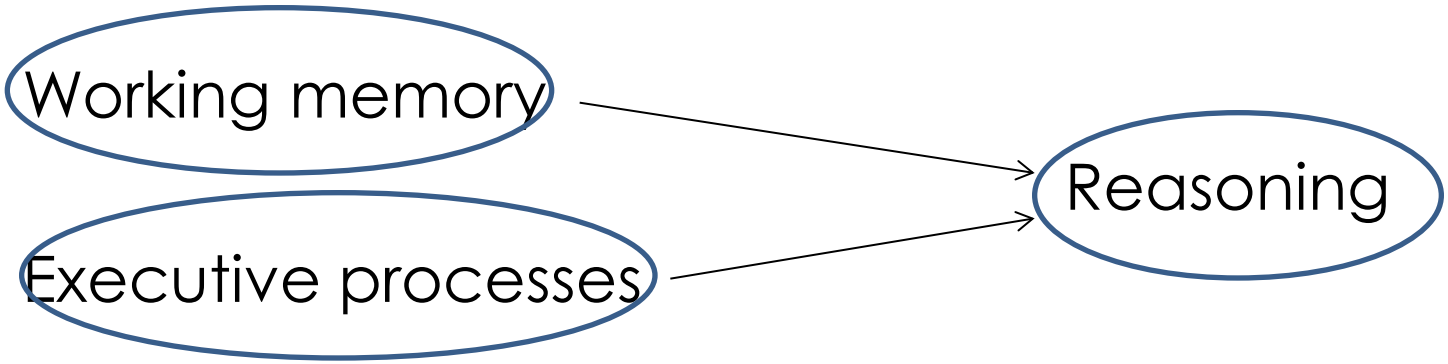
The available data (➡ manifest variables):

-three reasoning scores
-six (visual) working memory scores
-four executive process (shifting) scores

The rationale: a demonstration

… from cognitive psychology. It is investigated whether (visual) *working memory* and *executive processes* (shifting) explain *reasoning*.

The core model structure



The rationale: a demonstration

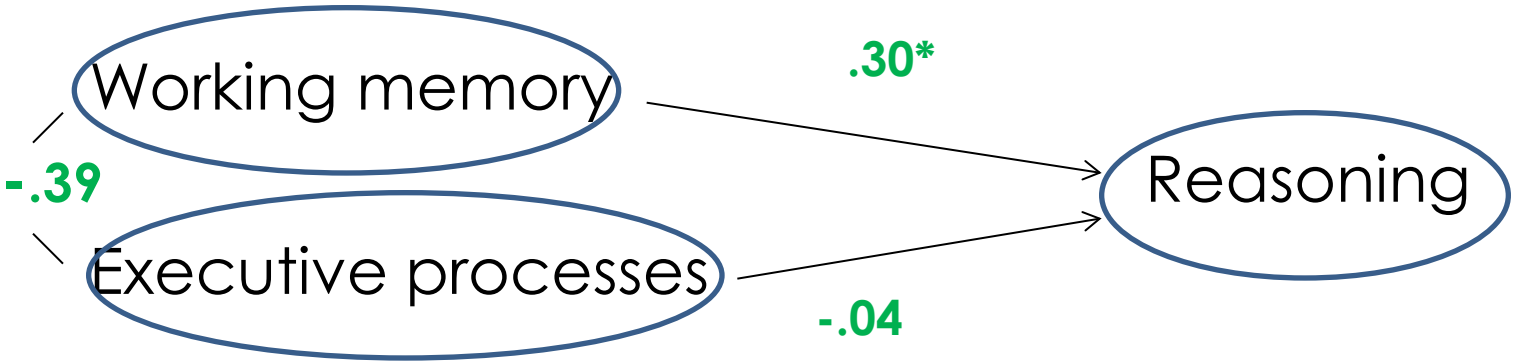
… from cognitive psychology. It is investigated whether (visual) *working memory* and *executive processes* (shifting) explain *reasoning*.

< the parameters were estimates by EFA and MR using SPSS: poor man‘s SEM >

The rationale: a demonstration

… from cognitive psychology. It is investigated whether (visual) *working memory* and *executive processes* (shifting) explain *reasoning*.

The outcome obtained by SPSS:



The rationale: a demonstration

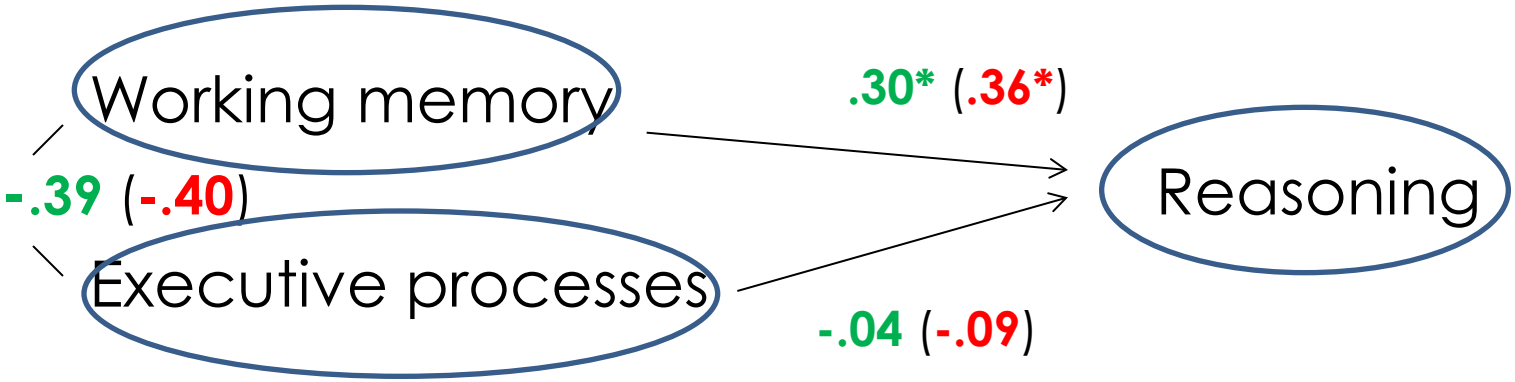
… from cognitive psychology. It is investigated whether (visual) *working memory* and *executive processes* (shifting) explain *reasoning*.

< the parameters were estimates by SEM software using LISREL >

The rationale: a demonstration

… from cognitive psychology. It is investigated whether (visual) *working memory* and *executive processes* (shifting) explain *reasoning*.

The outcome obtained by SPSS and *SEM procedure (in parentheses)*:



The rationale: a demonstration

… from cognitive psychology. It is investigated whether (visual) *working memory* and *executive processes* (shifting) explain *reasoning*.

SEM output additionall signifies that the *model fit* is ***not good*** because of inhomogeneity of the working memory scores.

… ***a measurement model was not good***

(this information is not available from SPSS)

Outline

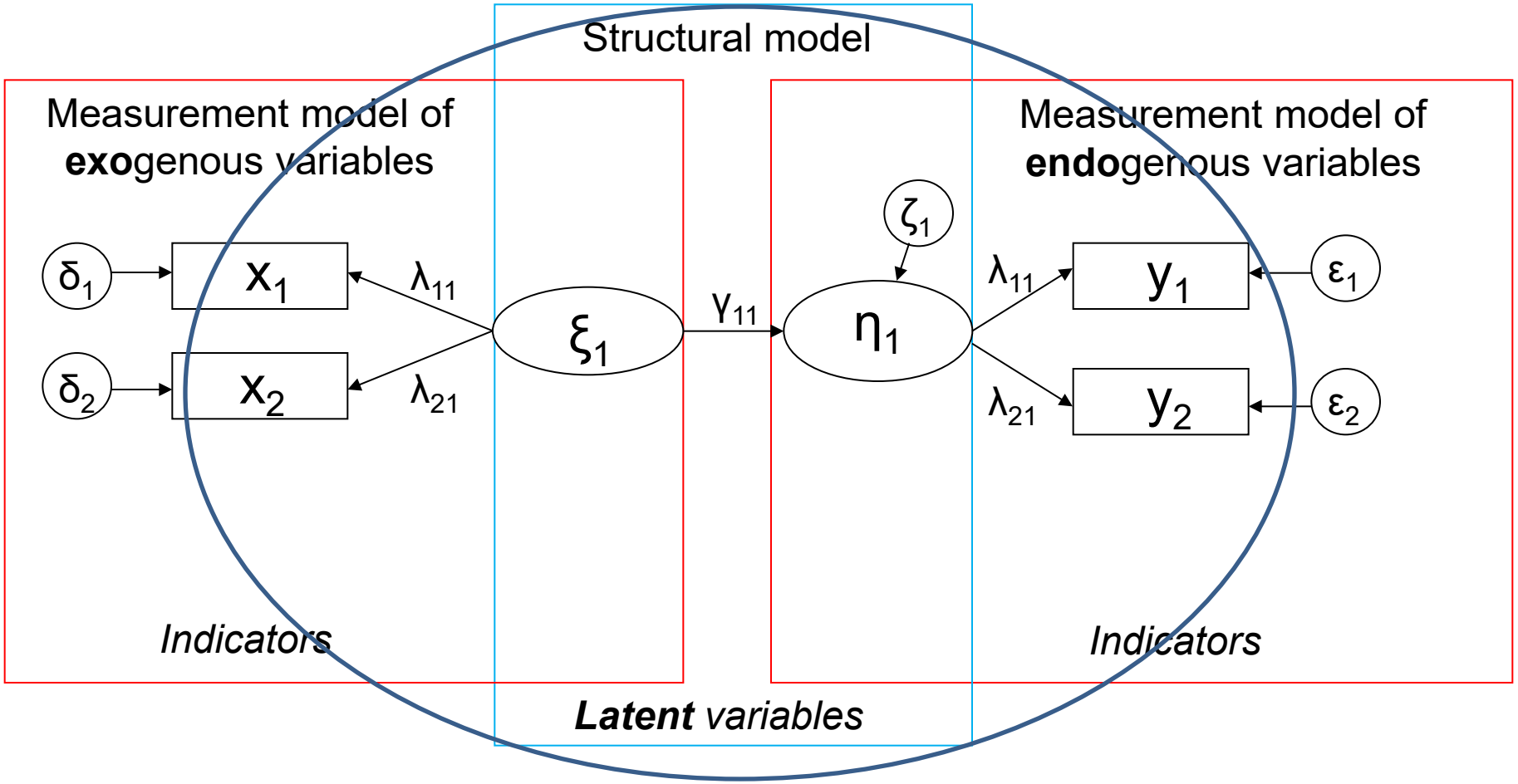
1. Introduction
2. The rational
3. *The complete structural equation model*
4. Types of models
5. The strategy of analysis
6. The evaluation of models

The complete structural equation model

The complete structural equation model combines models of measurement with a structural model:

The behind the sceen perspective

The complete structural equation model: simple case



The complete structural equation model

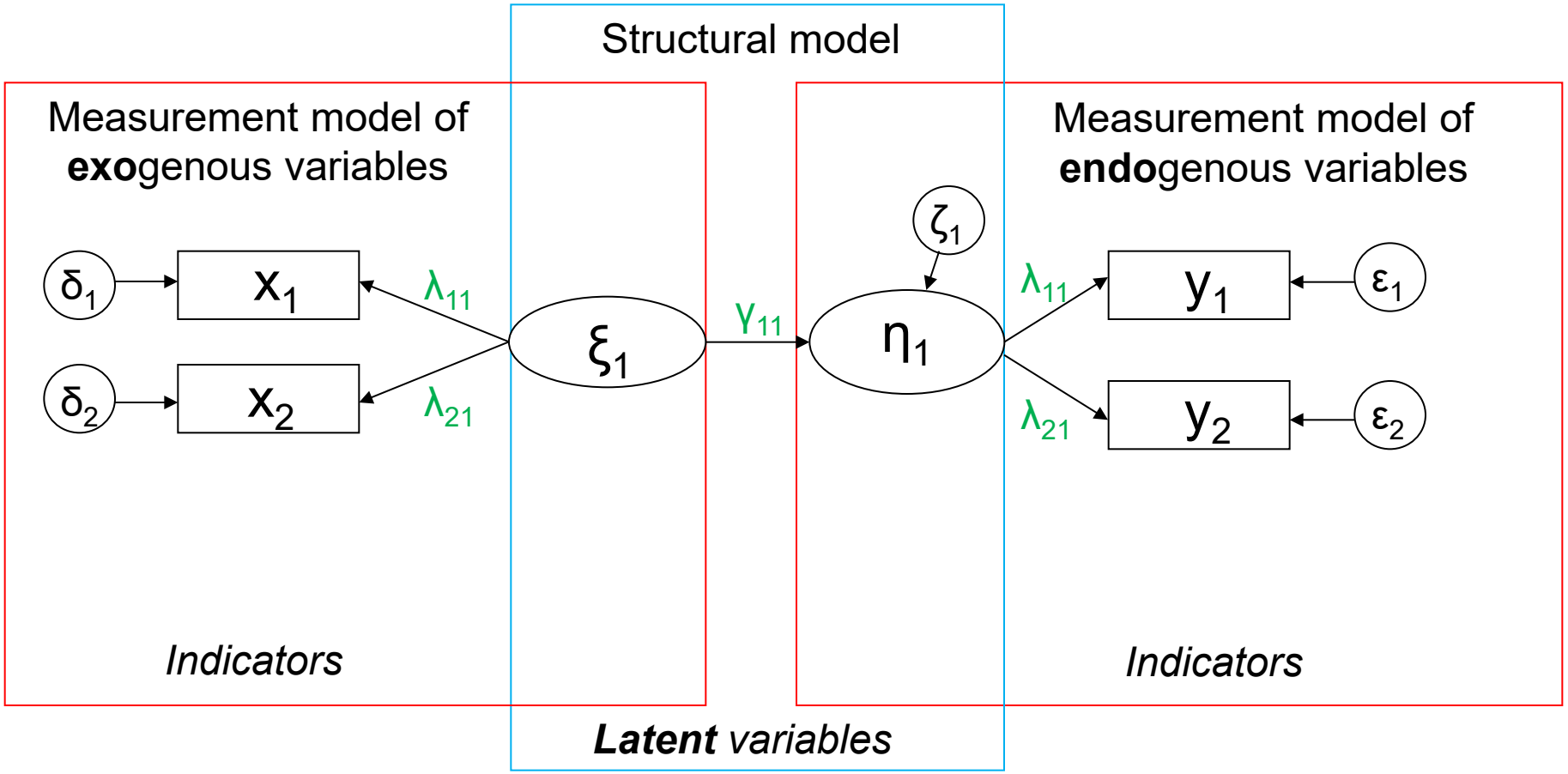
In preparing the path diagram both parts (measurement and structural models) have to be considered.

Furthermore, the diagram must be prepared for being transformed into the corresponding model of the covariance matrix:

$$\boldsymbol{\Sigma} = \ldots\ldots$$

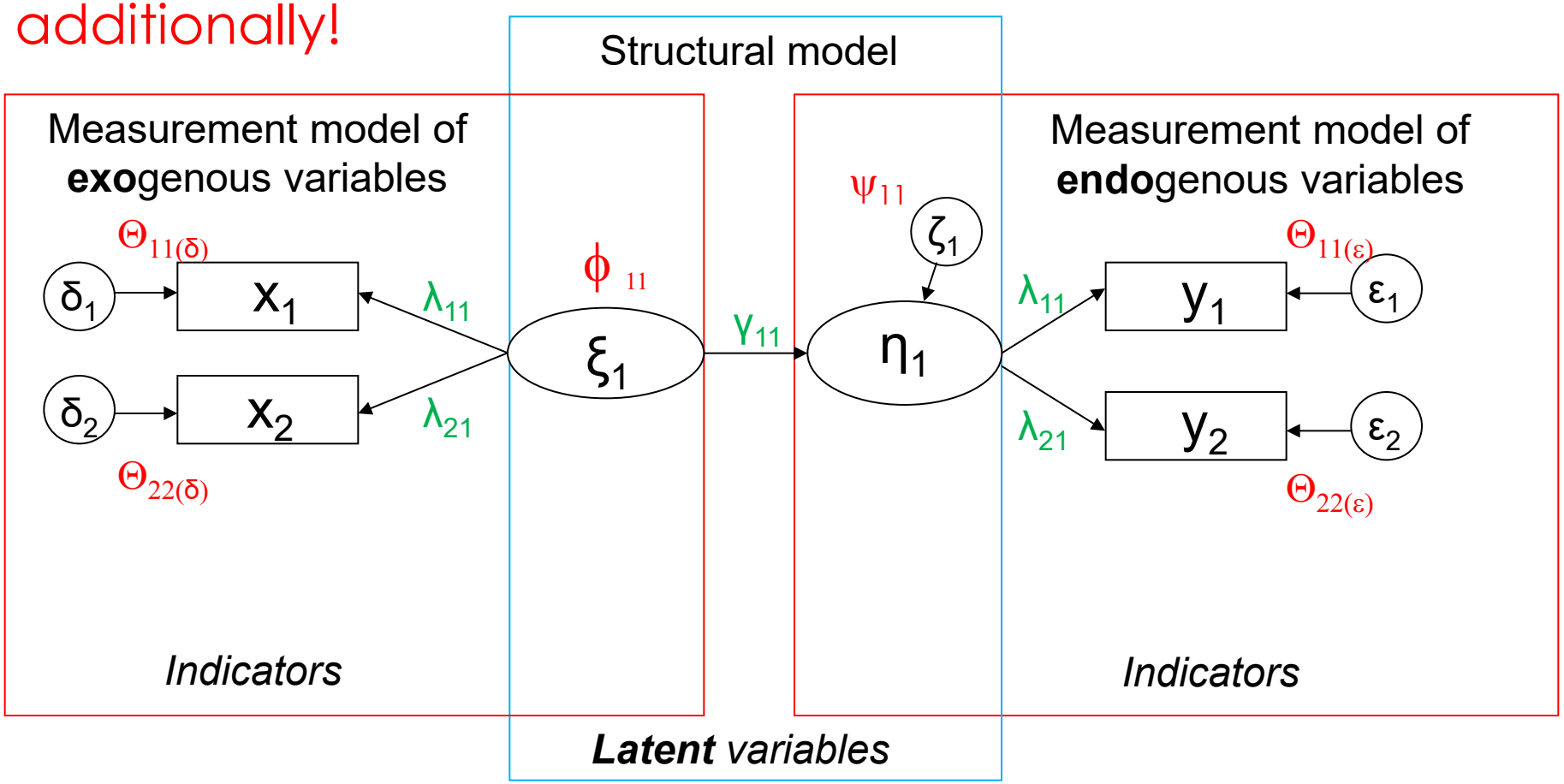
The complete structural equation model : simple case

What is important for the research aim!

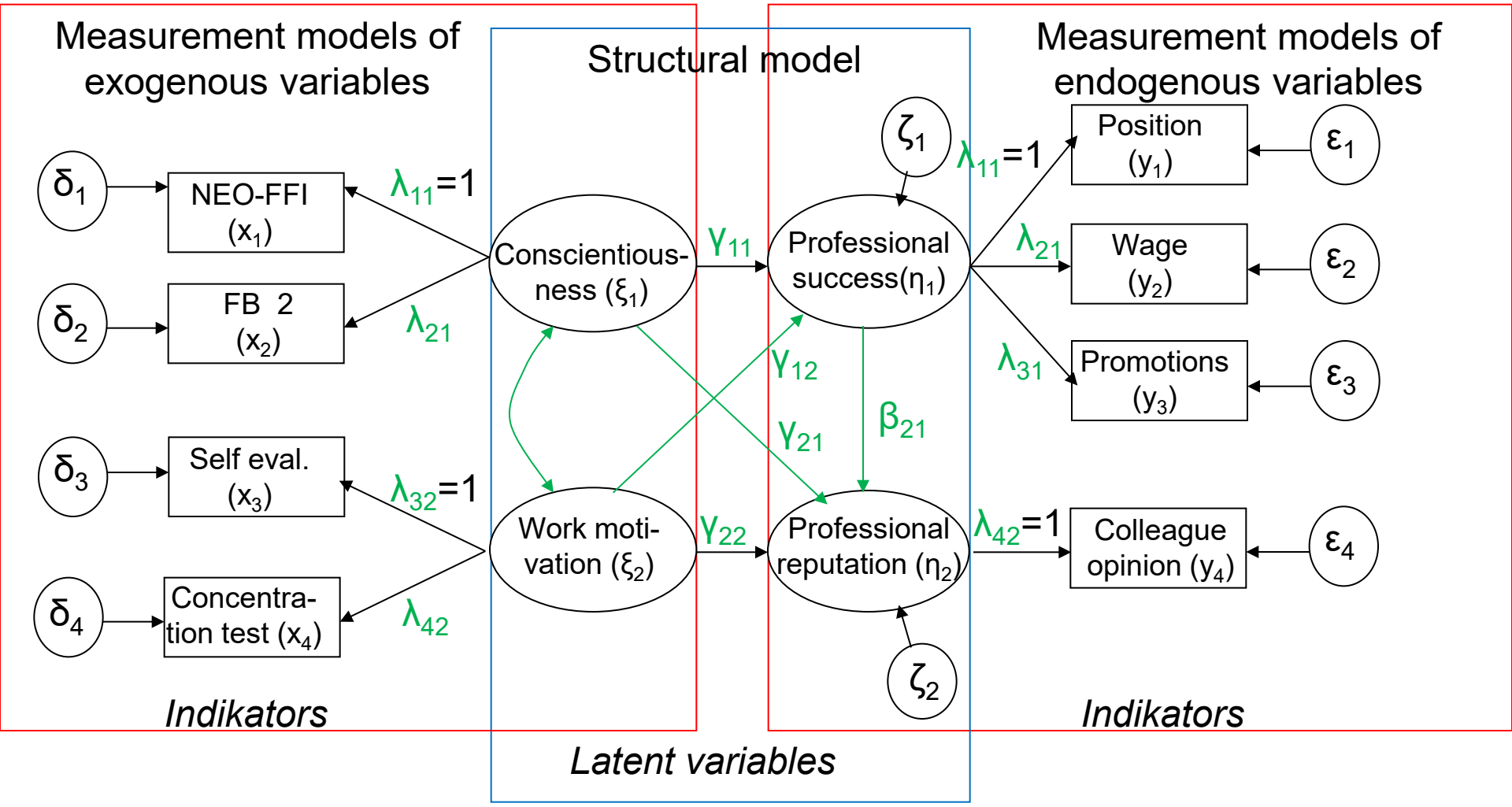


The complete structural equation model : simple case

What is needed additionally!

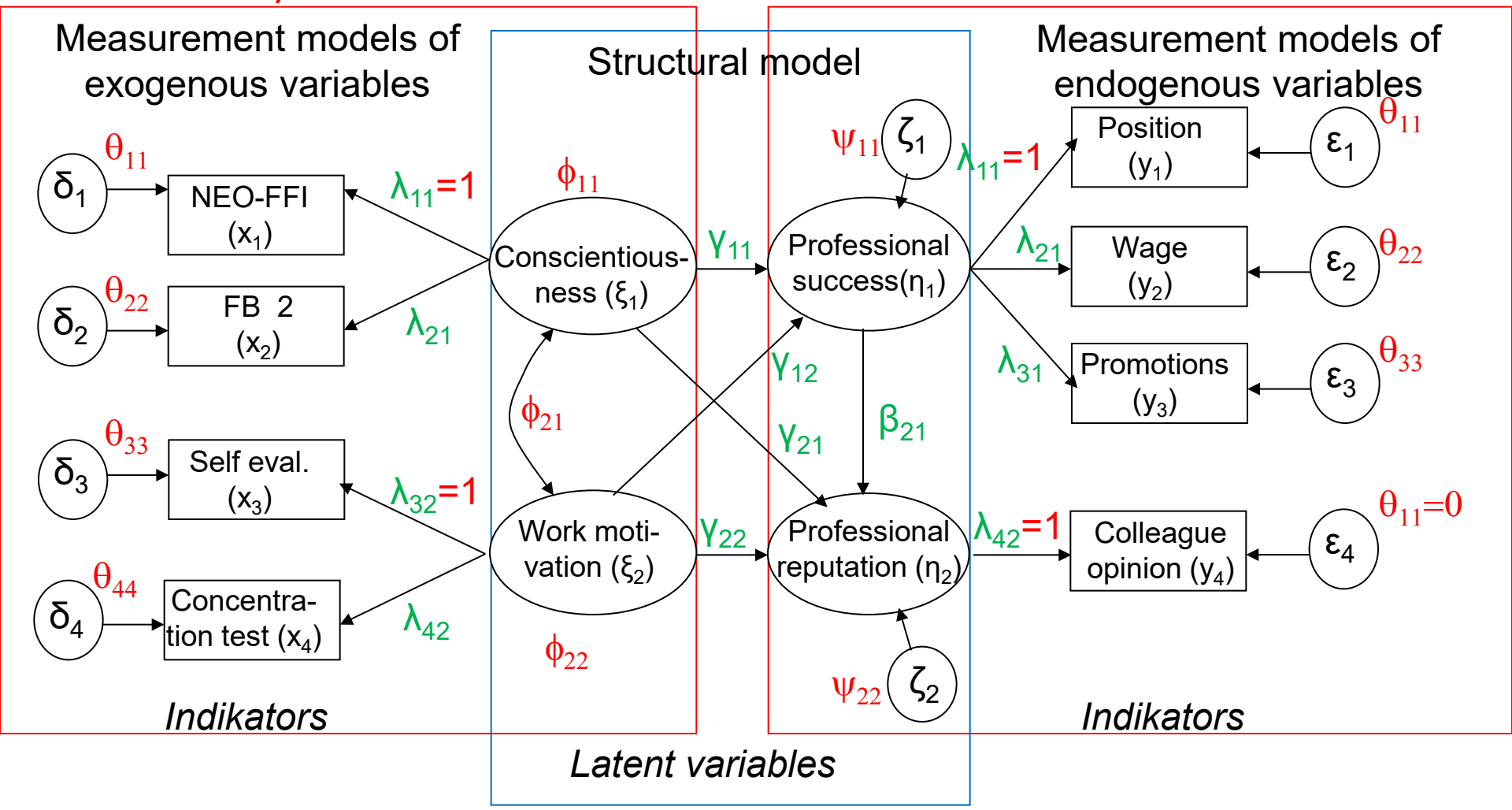


The complete structural equation model : complex case



The complete structural equation model : complex case

What is needed additionally!



The formal representation of *the simple case* *(What we know so far)*

- The endogenous model of measurement

$$\mathbf{y} = \Lambda_y \eta + \epsilon$$

- The exogenous model of measurement

$$\mathbf{x} = \Lambda_x \xi + \delta$$

- The structural model

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

The formal representation of *the simple case* *(What we know so far)*

- The model of the covariance matrix associated with the exogenous model of measurement

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'+\boldsymbol{\Theta}$$

… obtained from …

$$\mathbf{x} = \boldsymbol{\Lambda}\xi + \boldsymbol{\delta}$$

The model of the covariance matrix: *the simple case*

- The model of the covariance matrix associated with the complete structural equation model

$$\Sigma = \left(\begin{array}{c|c} \cdots & \cdots \\ \hline \cdots & \cdots \end{array} \right)$$

where $\mathbf{A} = \mathbf{I}$

The model of the covariance matrix: *the simple case*

- The model of the covariance matrix associated with the complete structural equation model

$$\Sigma = \left(\begin{array}{c|c} yy & yx \\ \hline xy & xx \end{array} \right)$$

where $\mathbf{A} = \mathbf{I}$

The model of the covariance matrix: *the simple case*

- The model of the covariance matrix associated with the complete structural equation model

$$\Sigma = \left(\begin{array}{c|c} \Lambda_y \mathbf{A}(\Gamma\Phi\Gamma' + \Psi)\mathbf{A}'\Lambda_y' + \Theta_\epsilon & \Lambda_y \mathbf{A}\Gamma\Phi\Lambda_x' \\ \hline \Lambda_x\Phi\Gamma'\mathbf{A}'\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta \end{array} \right)$$

where $\mathbf{A} = \mathbf{I}$

The model of the covariance matrix: *the simple case*

- The … $\mathbf{y} = \Lambda_y \eta + \epsilon$ … $\mathbf{x} = \Lambda_x \xi + \delta$

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

- The complete covariance model

$$\Sigma = \begin{pmatrix} \Lambda_y A(\Gamma \Phi \Gamma' + \Psi) A' \Lambda_y' + \Theta_\epsilon & \Lambda_y A \Gamma \Phi \Lambda_x' \\ \Lambda_x \Phi \Gamma' A' \Lambda_y' & \Lambda_x \Phi \Lambda_x' + \Theta_\delta \end{pmatrix}$$

where $A = I$

The formal representation of *the complex case*

- The endogenous model of measurement

$$\mathbf{y} = \Lambda_y \eta + \epsilon$$

- The exogenous model of measurement

$$\mathbf{x} = \Lambda_x \xi + \delta$$

- The structural model

$$\eta = \mathrm{B}\eta + \Gamma\xi + \zeta$$

$$\mathbf{B} = \begin{bmatrix} 0 & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & 0 & \beta_{23} & \beta_{24} \\ \beta_{31} & \beta_{32} & 0 & \beta_{34} \\ \beta_{41} & \beta_{42} & \beta_{43} & 0 \end{bmatrix}$$

The model of the covariance matrix: *the complex case*

- The complete covariance model

$$\Sigma = \begin{pmatrix} \Lambda_y \mathbf{A}(\Gamma\Phi\Gamma' + \Psi)\mathbf{A}'\Lambda_y' + \Theta_\epsilon & \Lambda_y \mathbf{A}\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma'\mathbf{A}'\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta \end{pmatrix}$$

where $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$.

The model of the covariance matrix: *the complex case*

- The … $\mathbf{y} = \Lambda_y \eta + \epsilon$ … $\mathbf{x} = \Lambda_x \xi + \delta$

  $\eta = \mathrm{B}\eta + \Gamma\xi + \zeta$

- The complete covariance model

$$\Sigma = \begin{pmatrix} \Lambda_y \mathrm{A}(\Gamma\Phi\Gamma' + \Psi)\mathrm{A}'\Lambda_y' + \Theta_\epsilon & \Lambda_y \mathrm{A}\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma'\mathrm{A}'\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta \end{pmatrix}$$

where $\mathrm{A} = (\mathrm{I} - \mathrm{B})^{-1}$.

The estimation problem:

Finde numbers for all parameters of $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Lambda}_y \mathbf{A}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})\mathbf{A}'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\epsilon & \boldsymbol{\Lambda}_y \mathbf{A}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' \\ \boldsymbol{\Lambda}_x \boldsymbol{\Phi}\boldsymbol{\Gamma}'\mathbf{A}'\boldsymbol{\Lambda}_y' & \boldsymbol{\Lambda}_x \boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta \end{pmatrix}$$

so that $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S}) \longrightarrow \min$

How is it done? See estimation methods

The complete structural equation model

All parameters of $\boldsymbol{\Sigma}$ are estimated simultaneously!

The complete structural equation model

The *correct preparation* of the path diagram is very important !!!

.... and is controlled by the user !!!

Regarding the additional parameters of the model of the covariance matrix: there are ***defaults*** (*pre set specifications*)
.... *but better not relay on defaults*

Outline

1. Introduction
2. The rational
3. The complete structural equation model
4. *Types of models*
5. The strategy of analysis
6. The evaluation of models

Types of models

In the following some types of complete structural equation models are discussed.

Types of models

There are types of complete structural equation models that apply to a number of different research questions

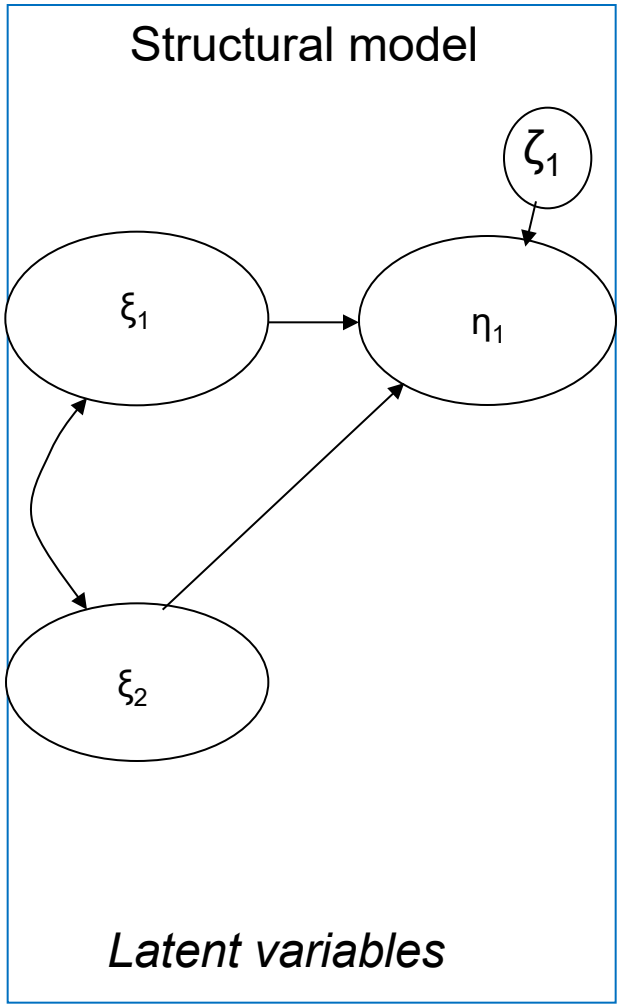
Types of models

The linear prediction model

Types of models

Basic model structure:

Linear prediction model
(see multiple regression)



Types of models: ***application***

Research question:

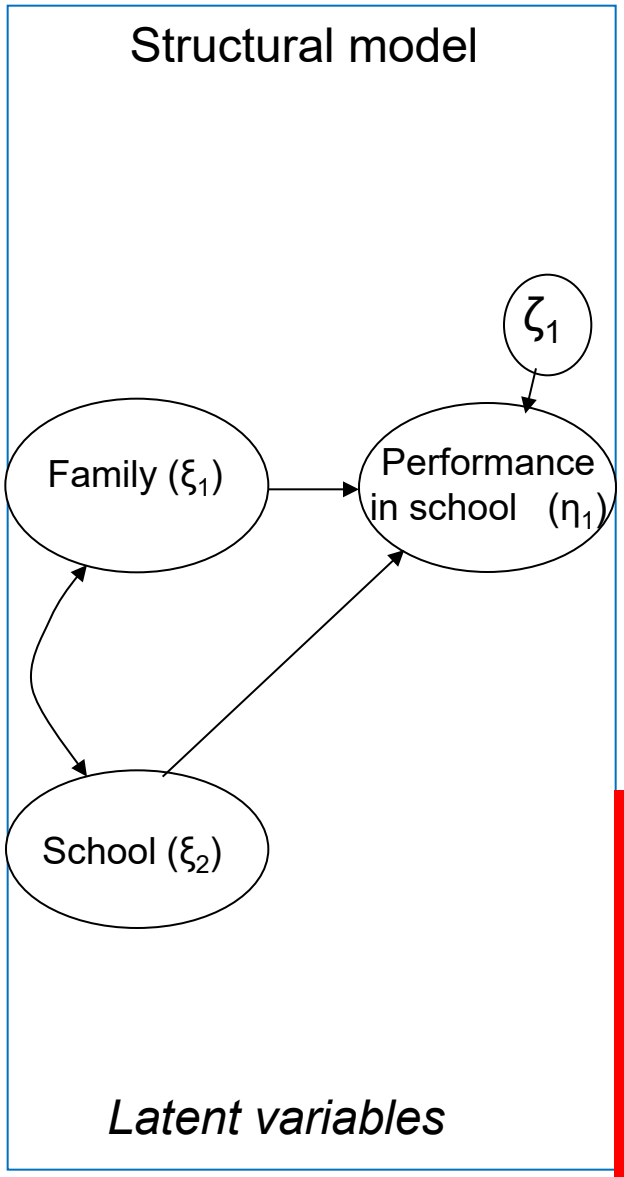
How do family and school (*the teacher, teaching style, etc.***) influence the child‘s performance in school?**

Hypothesis 1
Family influences the child‘s performance in school

Hypothesis 2
Characteristics of school influence the child‘s performance in school

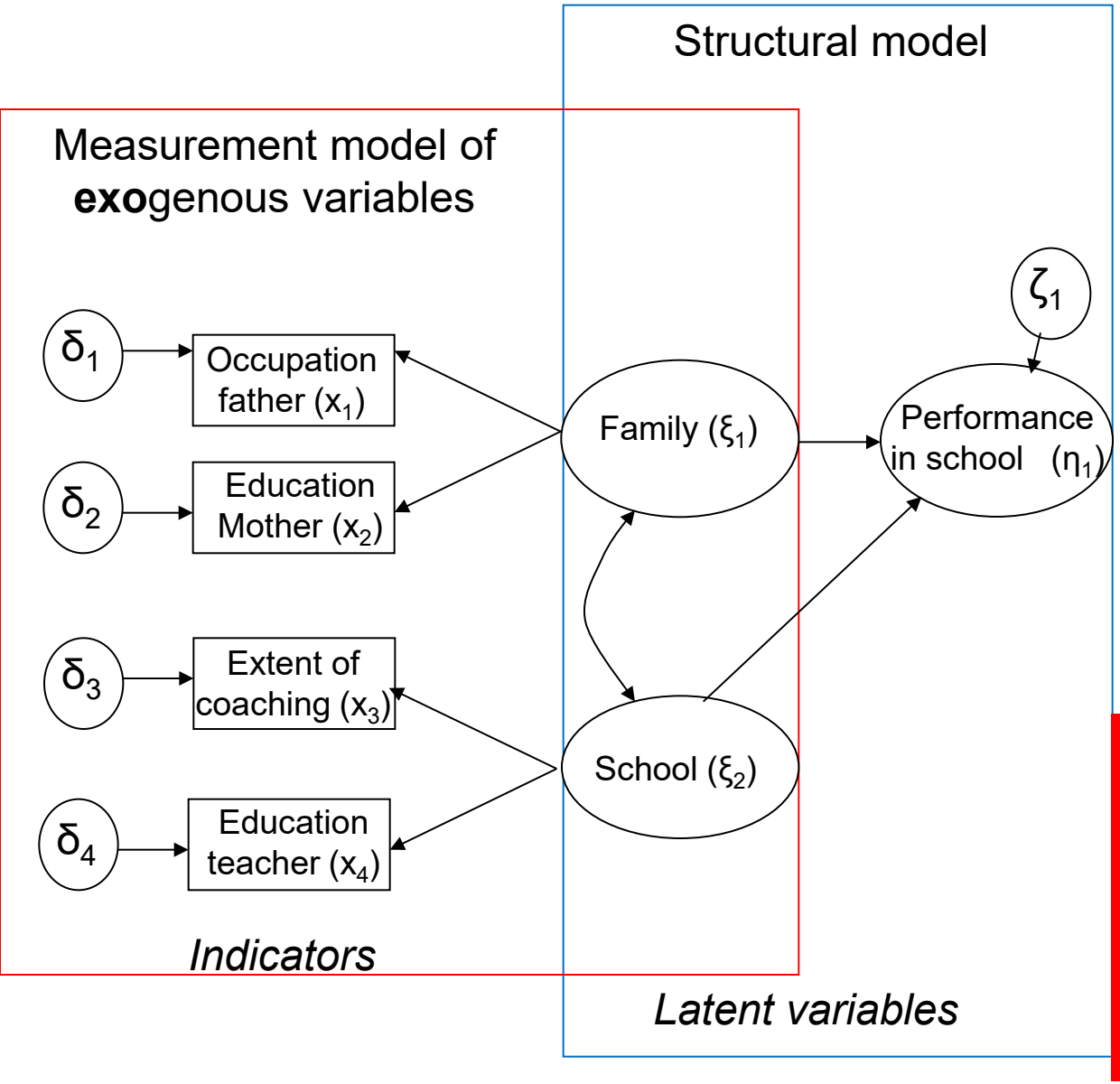
(Hypothesis 3
Family and school depend on each other **)**

Path diagram for example



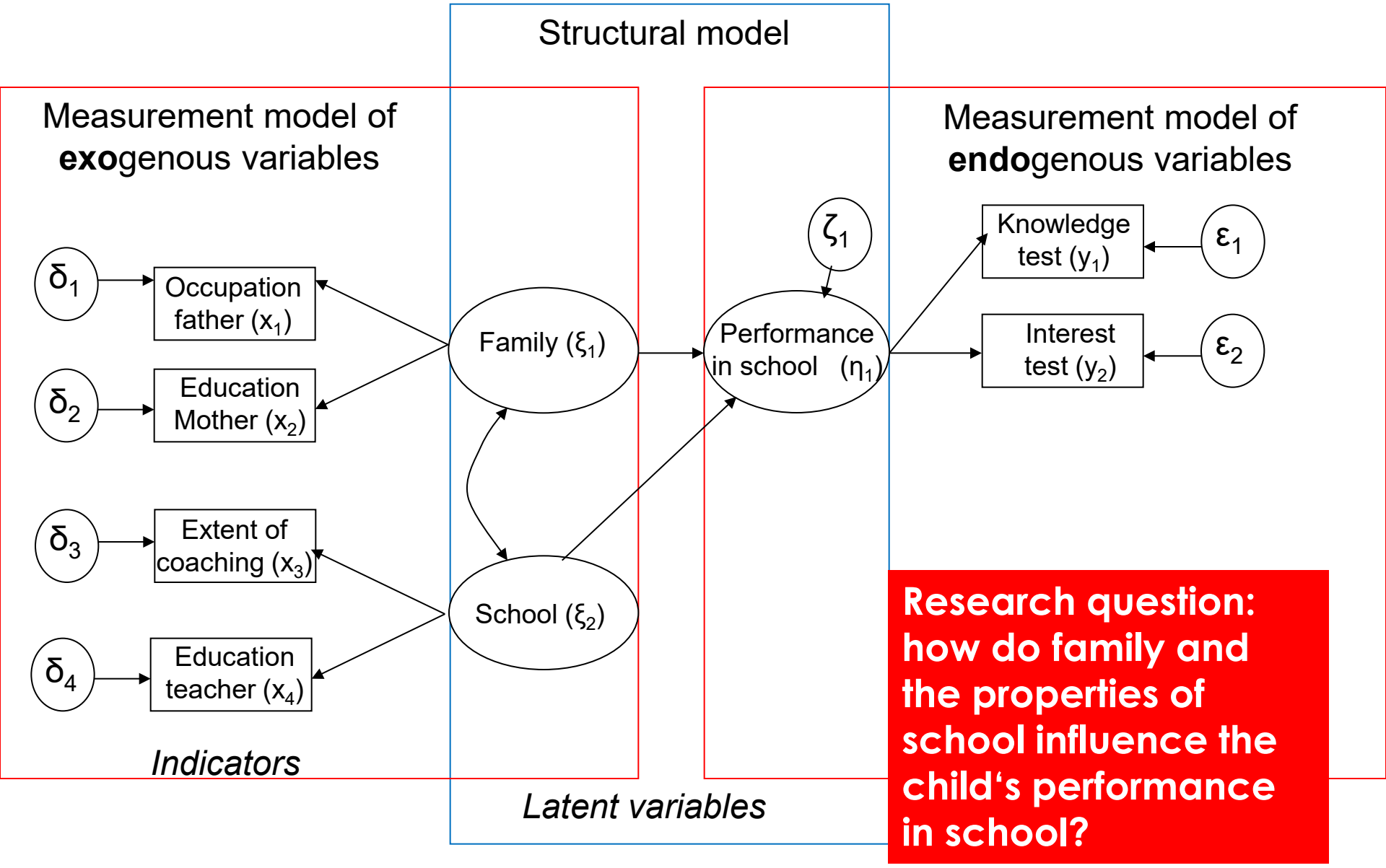
Research question: how do family and the properties of school influence the child's performance in school?

Path diagram for example



Research question: how do family and the properties of school influence the child's performance in school?

Path diagram for example



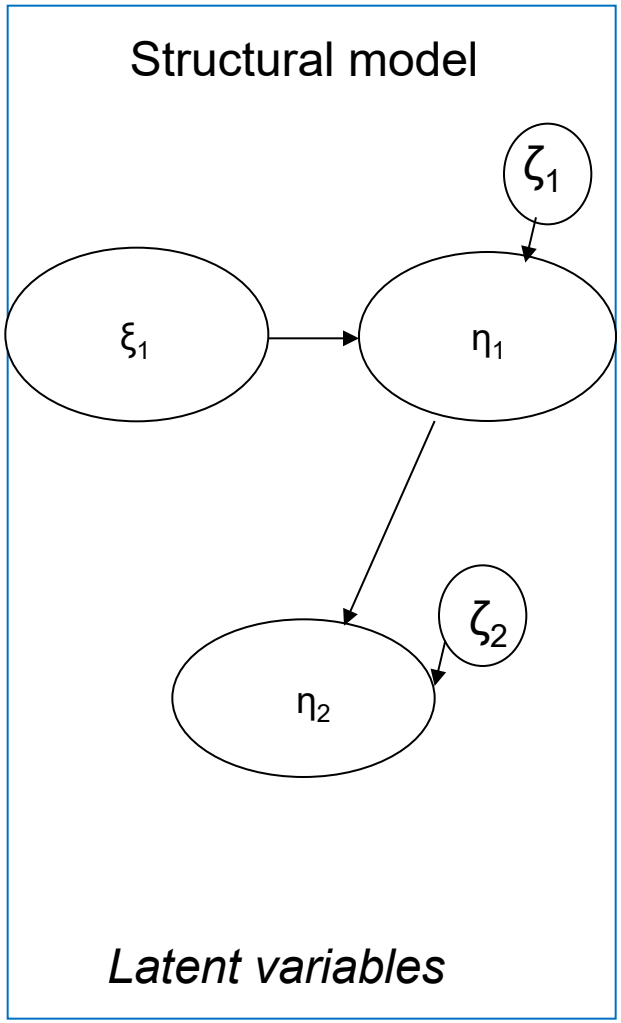
Types of models

The mediation model *(in combination with the linear prediction model)*

Types of models

Basic model structure:

Mediation model
(see path analysis)



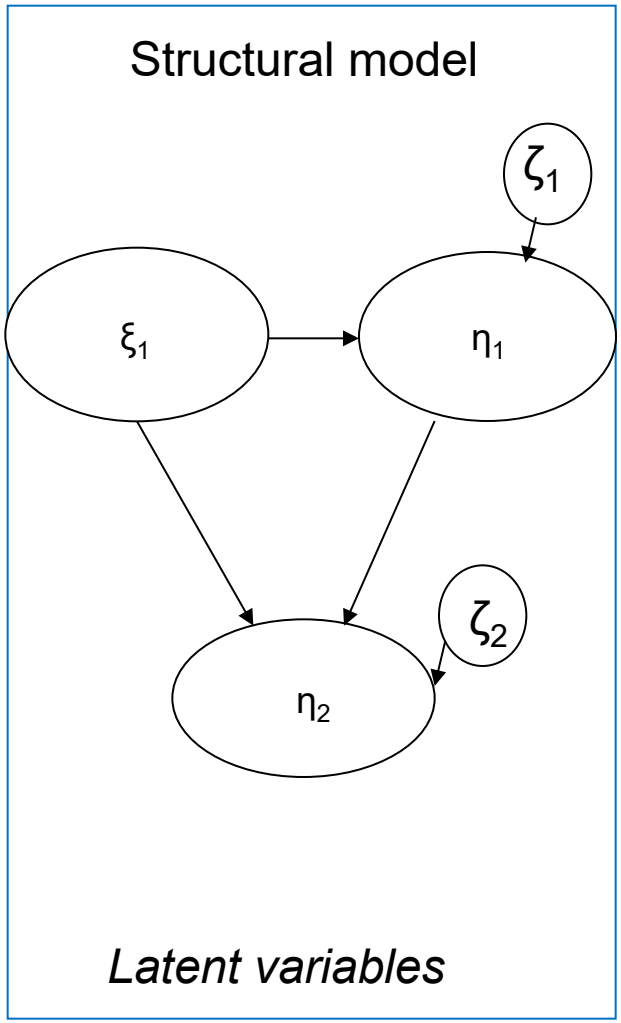
Types of models

Basic model structure:

Mediation model
(see path analysis)

mostly in combination with

Linear prediction model



This enlarged model is selected to find out whether η_1 is really important!

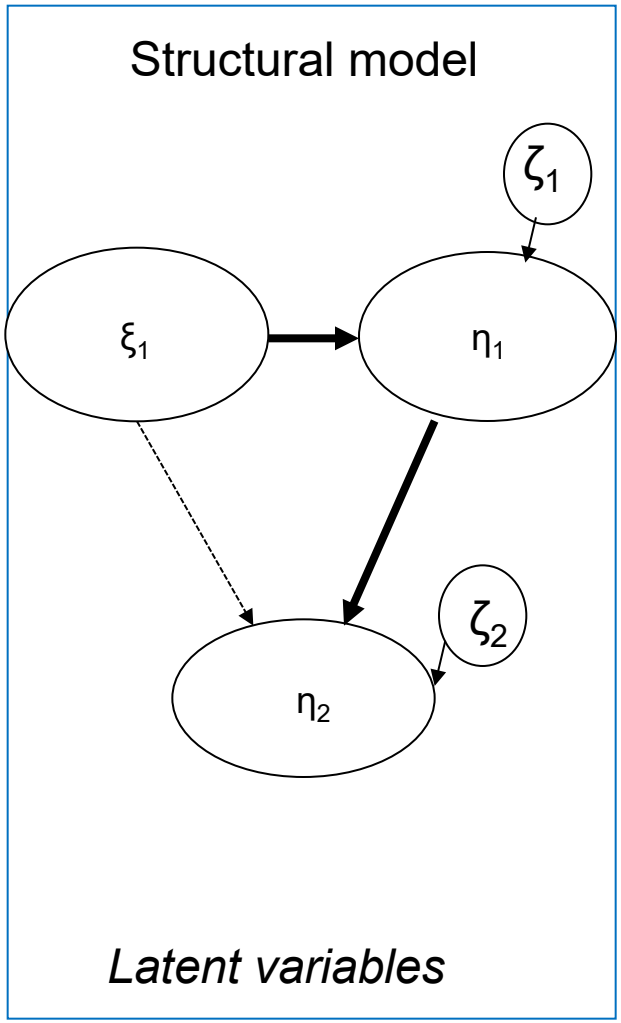
Types of models

Basic model structure:

Mediation model
(see path analysis)

mostly in combination with

Linear prediction model



This enlarged model is selected to find out whether η_1 is really important!

If yes

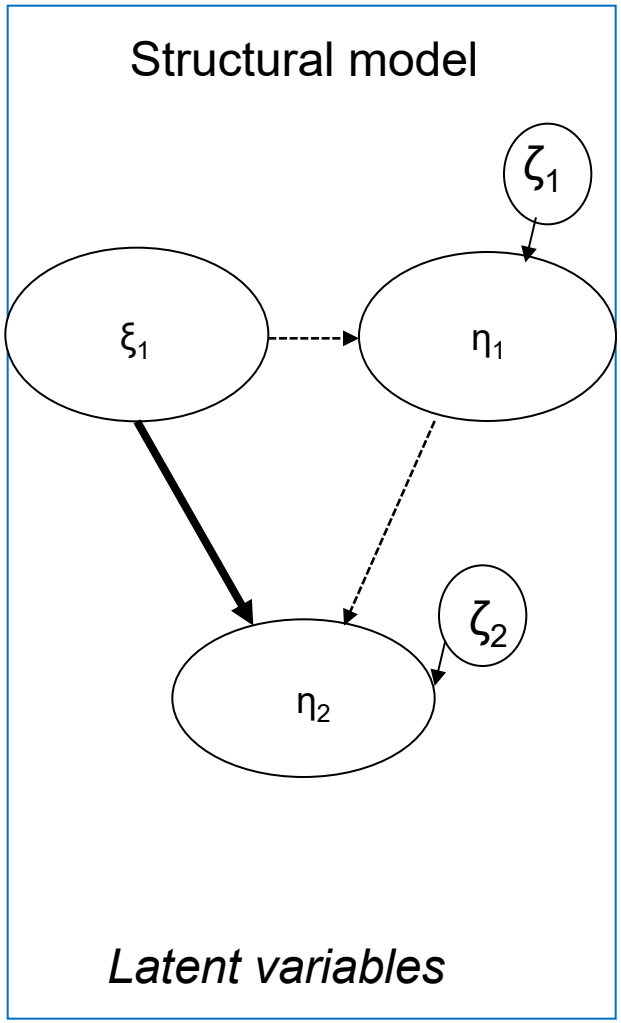
Types of models

Basic model structure:

Mediation model
(see path analysis)

mostly in combination with

Linear prediction model



This enlarged model is selected to find out whether η_1 is really important!

If no

Types of models: ***application***

Research question:

Does executive control predict performance in problem solving directly or indirectly via working memory?

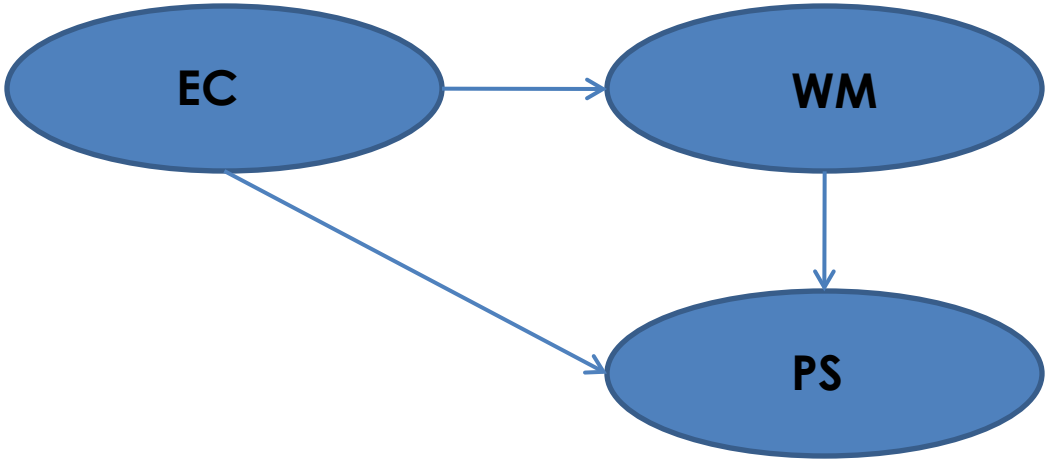
Hypothesis
Working memory mediates the effect of executive control on problem solving

Alternative Hypothesis
Executive control influences problem solving directly

Types of models: *application*

Research question:

Does executive control predict performance in problem solving directly or indirectly via working memory?



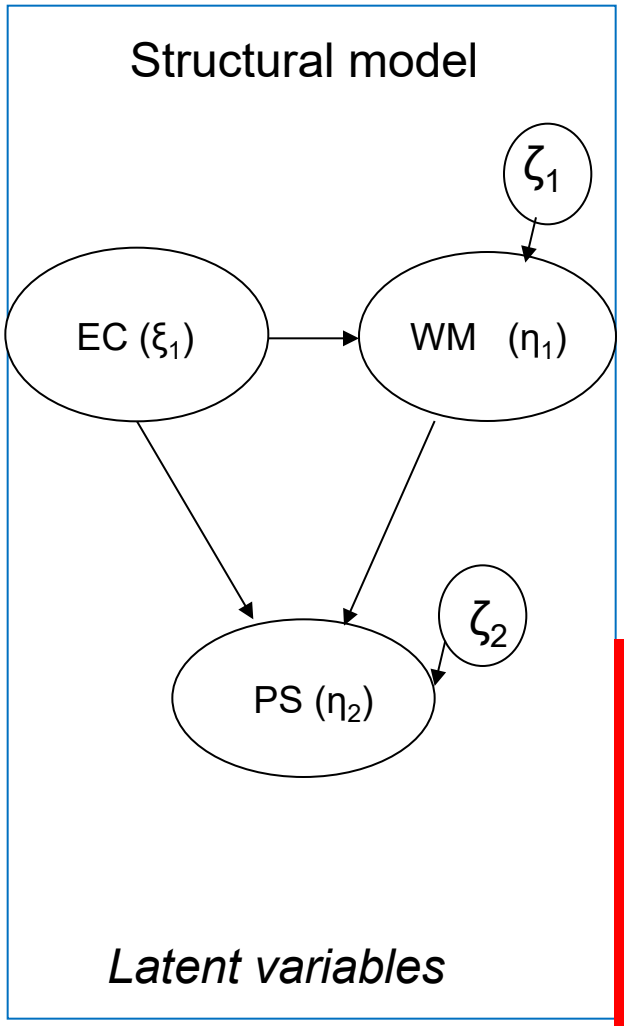
Types of models

Basic model structure:

Mediation model
(see path analysis)

mostly in combination with

Linear prediction model



Research question: Does executive control predict performance in problem solving directly or indirectly via working memory?

Types of models

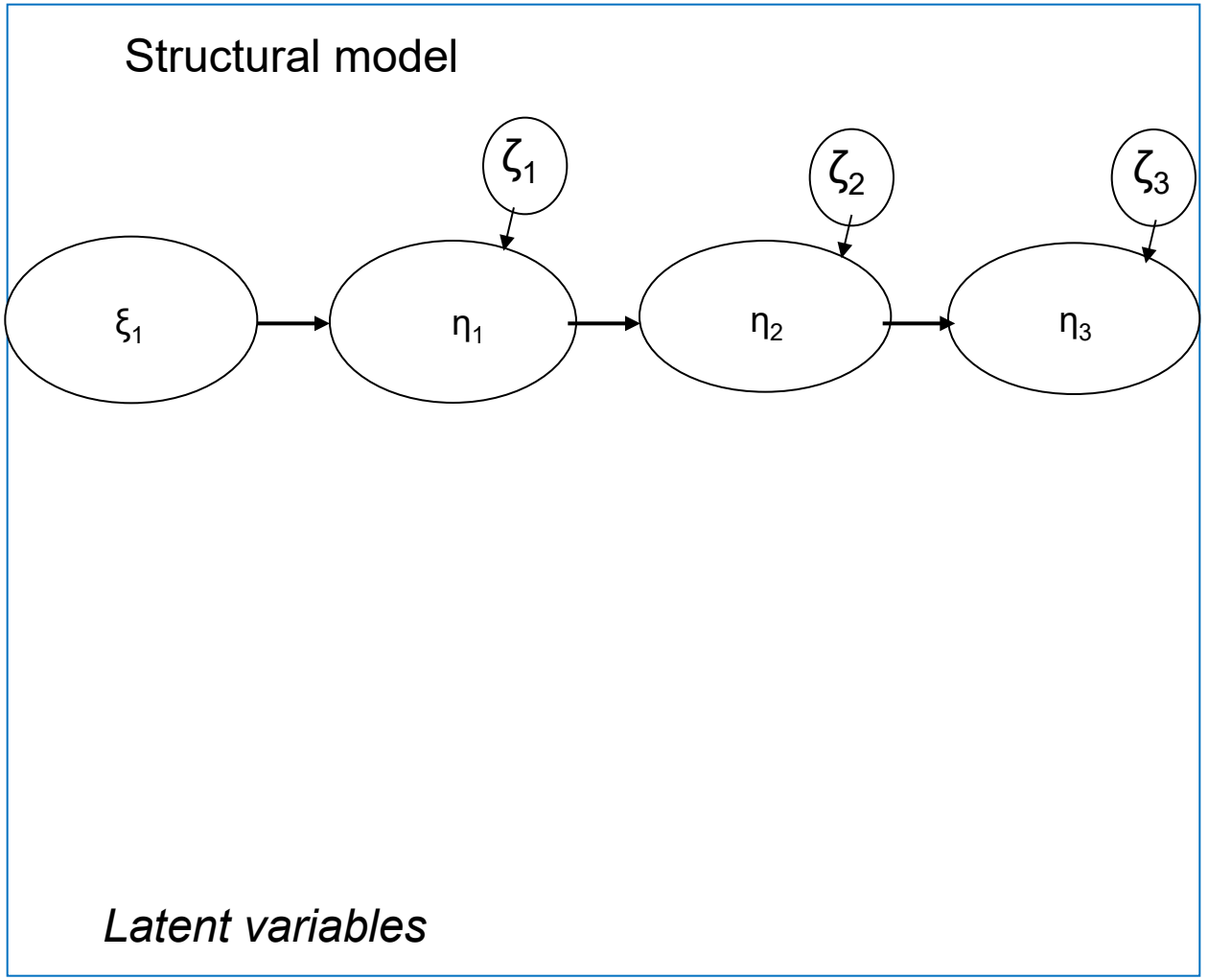
The developmental model *(as extended mediation model)*

Types of models

Basic model structure:

Developmental model

… with check for direct effect

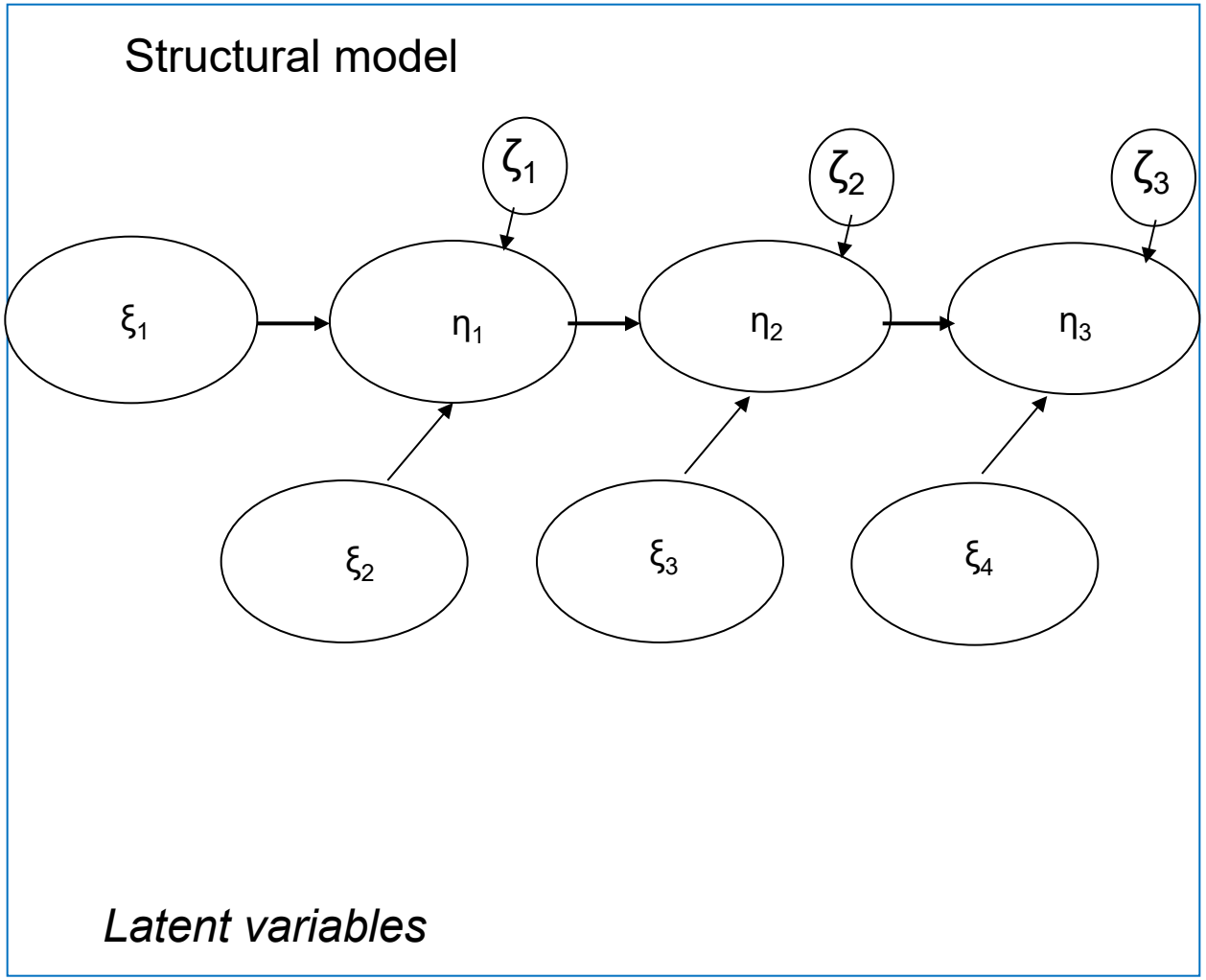


Types of models

Basic model structure:

Developmental model

… with check for additional influences



Types of models: ***application***

Research question:

How is the ability to write text (w-ability) influenced by teaching language grammar over a period of four years?

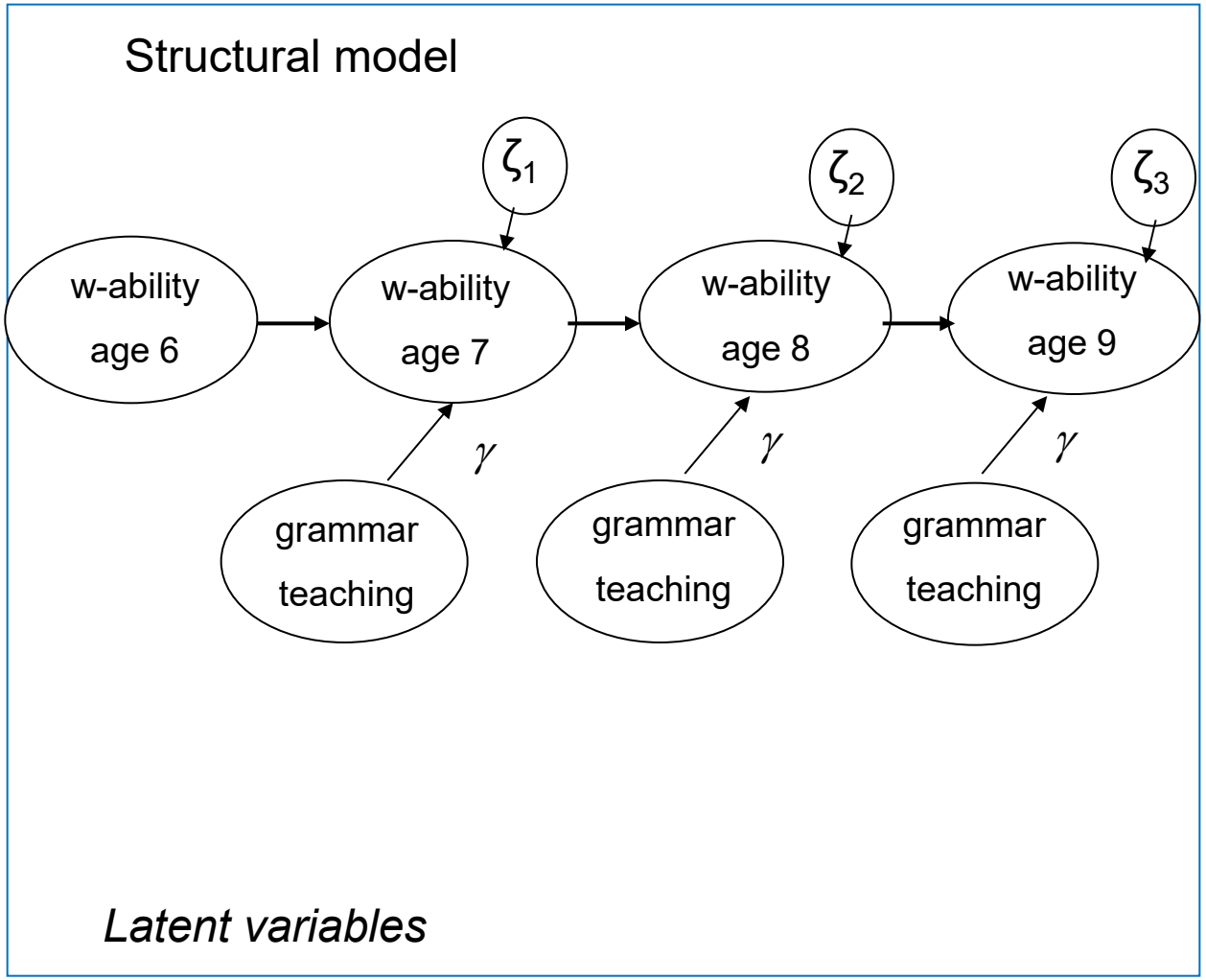
Hypothesis
There is a constant direct effect of teaching
Alternative hypothesis
The effect of teaching increases with childrens' age

Types of models

Basic model structure:

Developmental model

… with check for direct effect

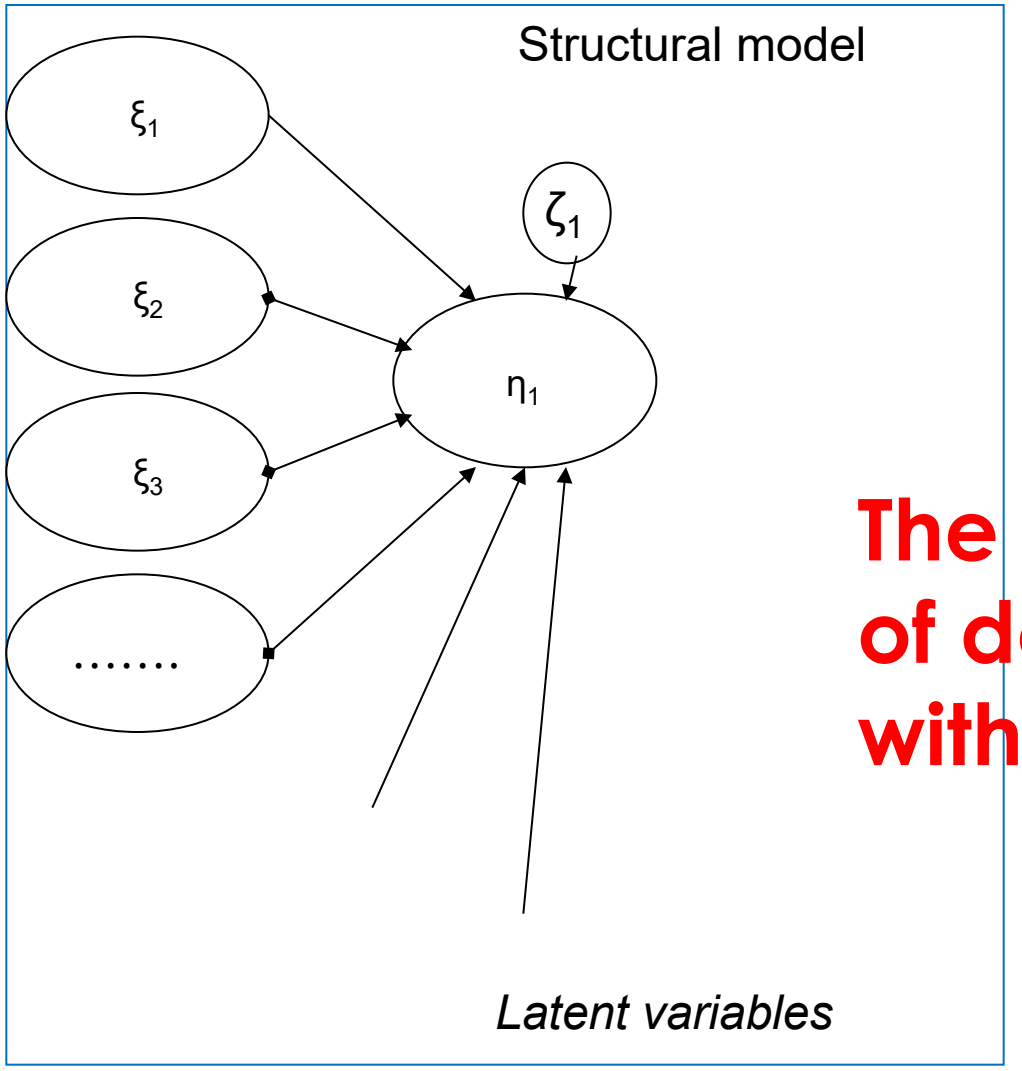


Types of models

The hierarchical linear prediction model

... *for dealing with many predictors*

Types of models



Basic model structure:

Linear prediction model

The problem of dealing with many …

Types of models: ***application***

Research question:

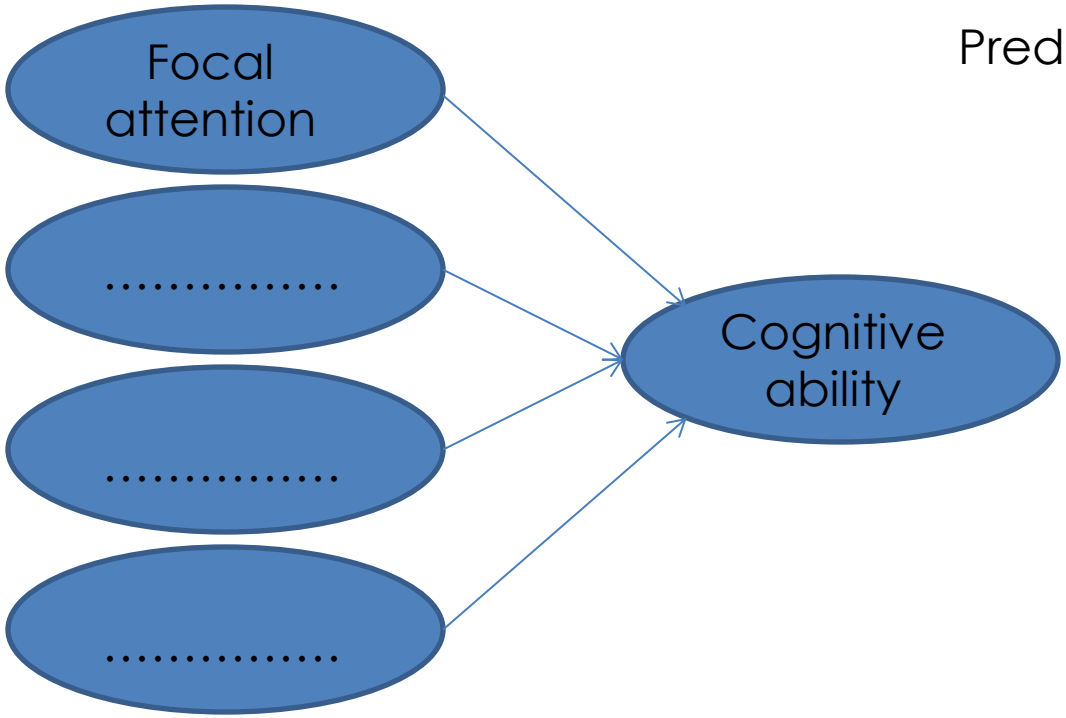
Does attention (focal attention, concentration, sustained attention, vigilance, etc.) predict cognitive ability?

Hypothesis

Attention predicts cognitive ability

Types of models: ***application***

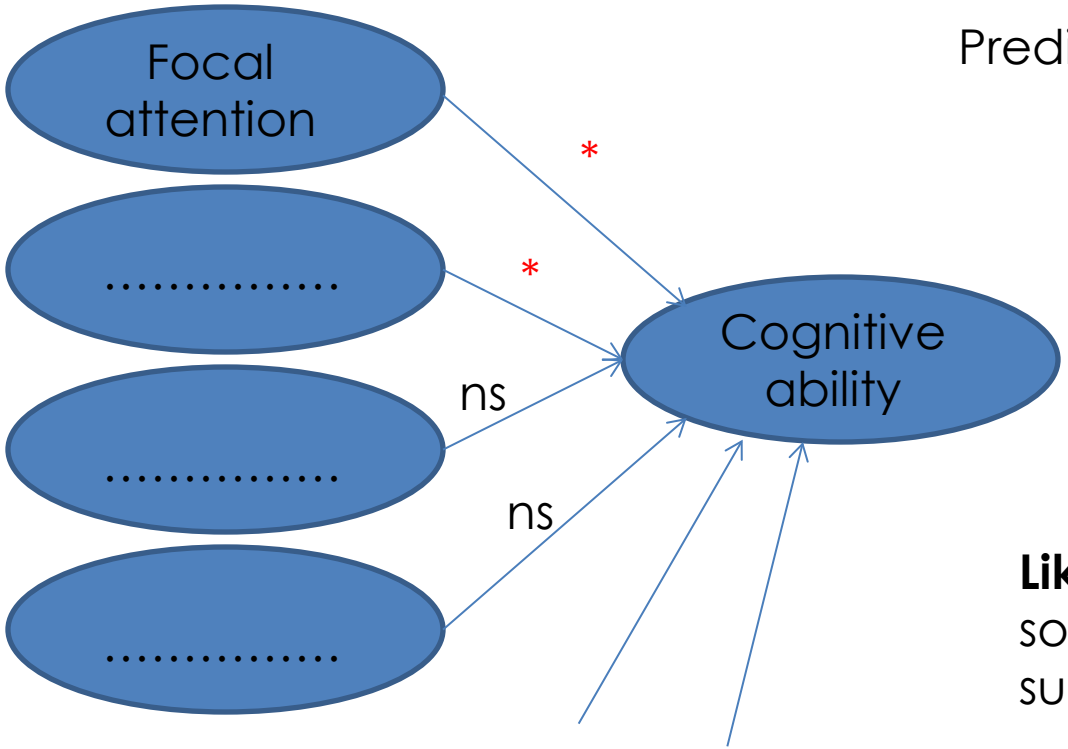
Model 1 (attention types as predictors)



Prediction model!

Types of models: ***application***

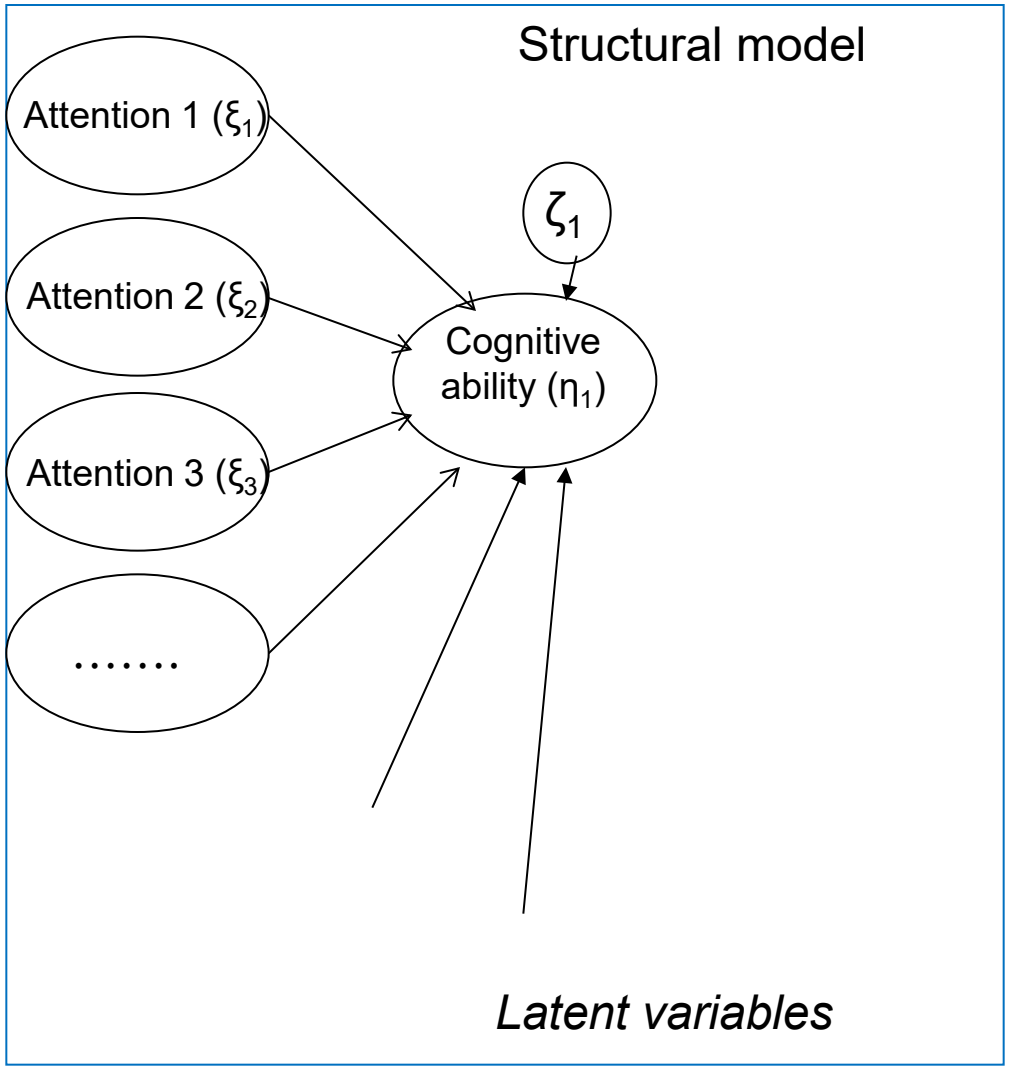
Model 1 (attention types as predictors)



Prediction model!

Likely disadvantage: only some types contribute substantially (*)

Types of models

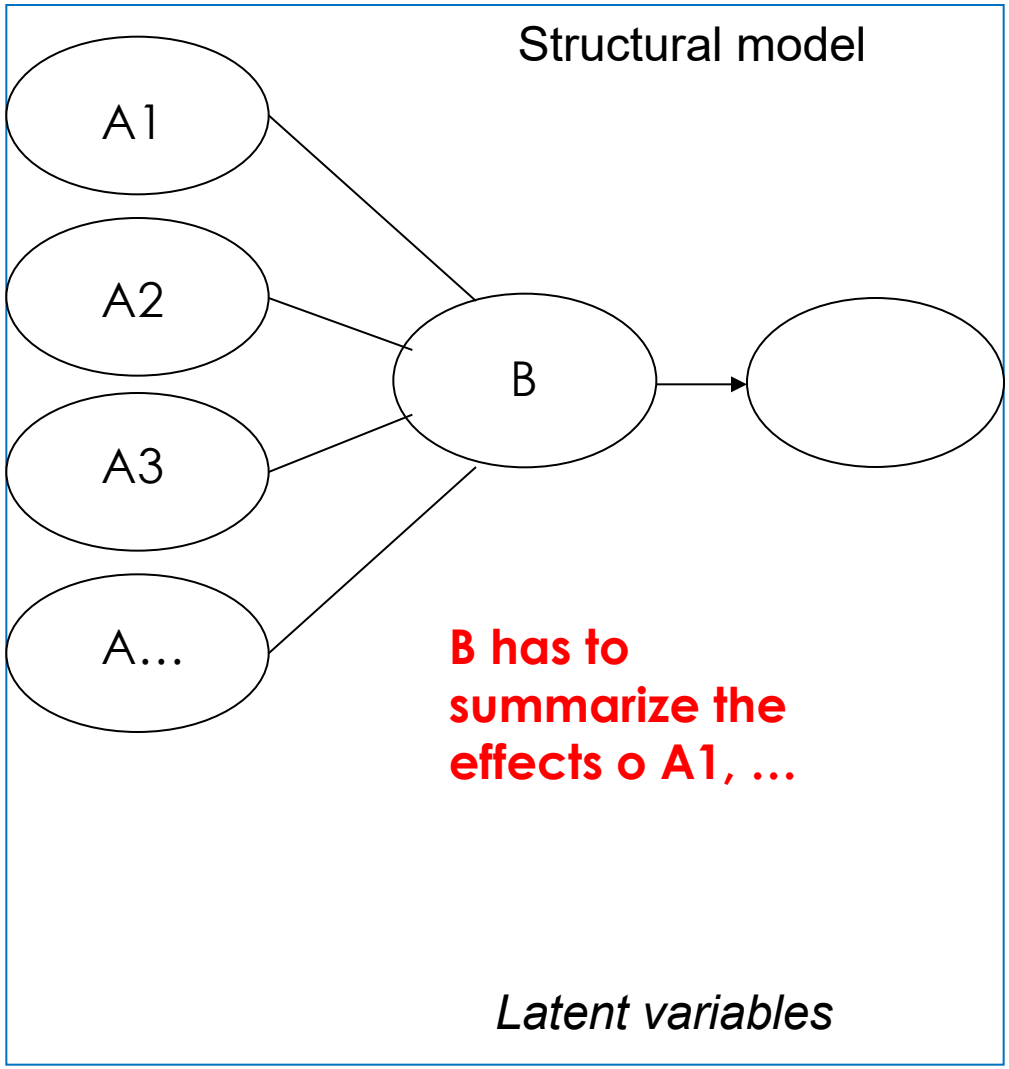


Basic model structure:

Linear prection model

Disadvantages: (1) only some predictors predict

Types of models



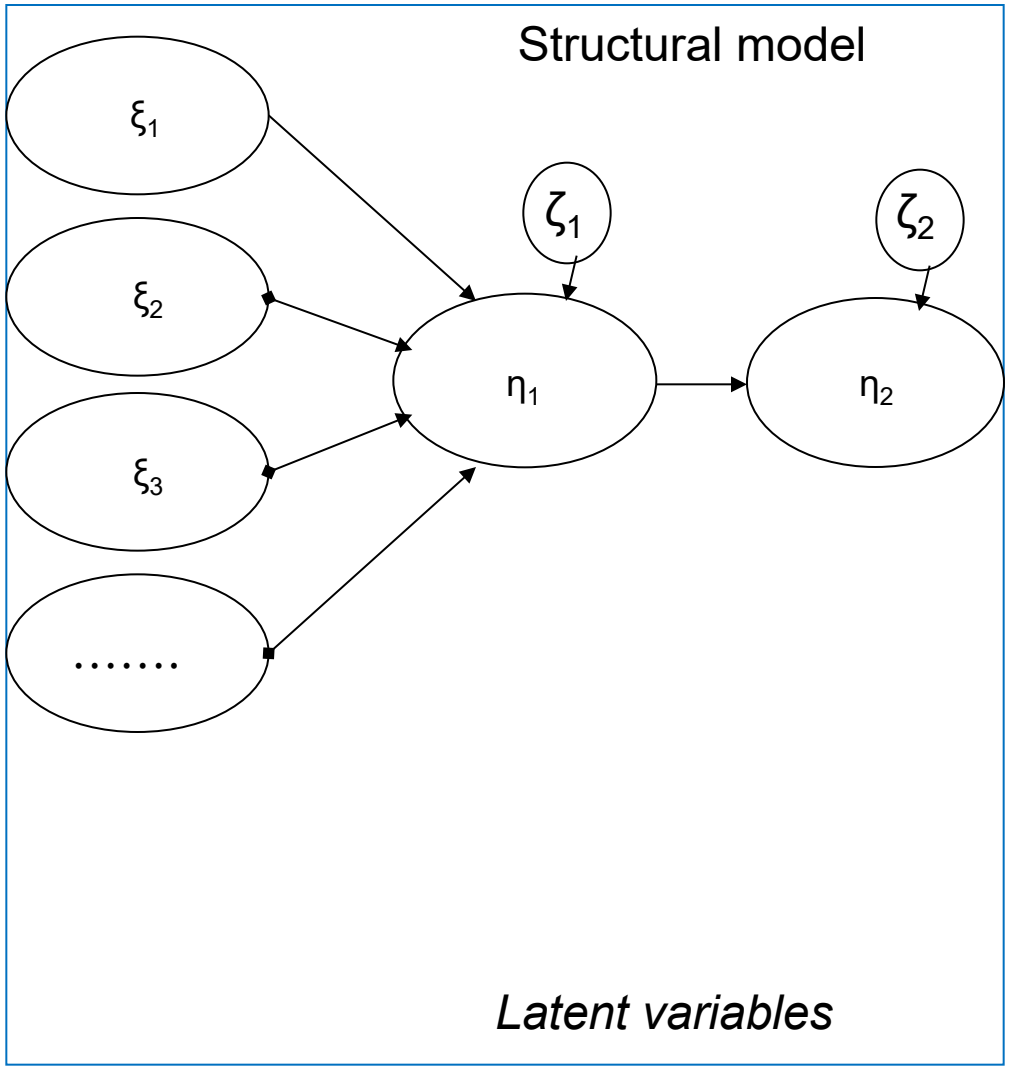
Basic model structure:

Linear prection model with mediation

- or -

Hierarchical linear prection model

Types of models



Basic model structure:

Linear prection model with mediation

Types of models: ***application***

Research question:

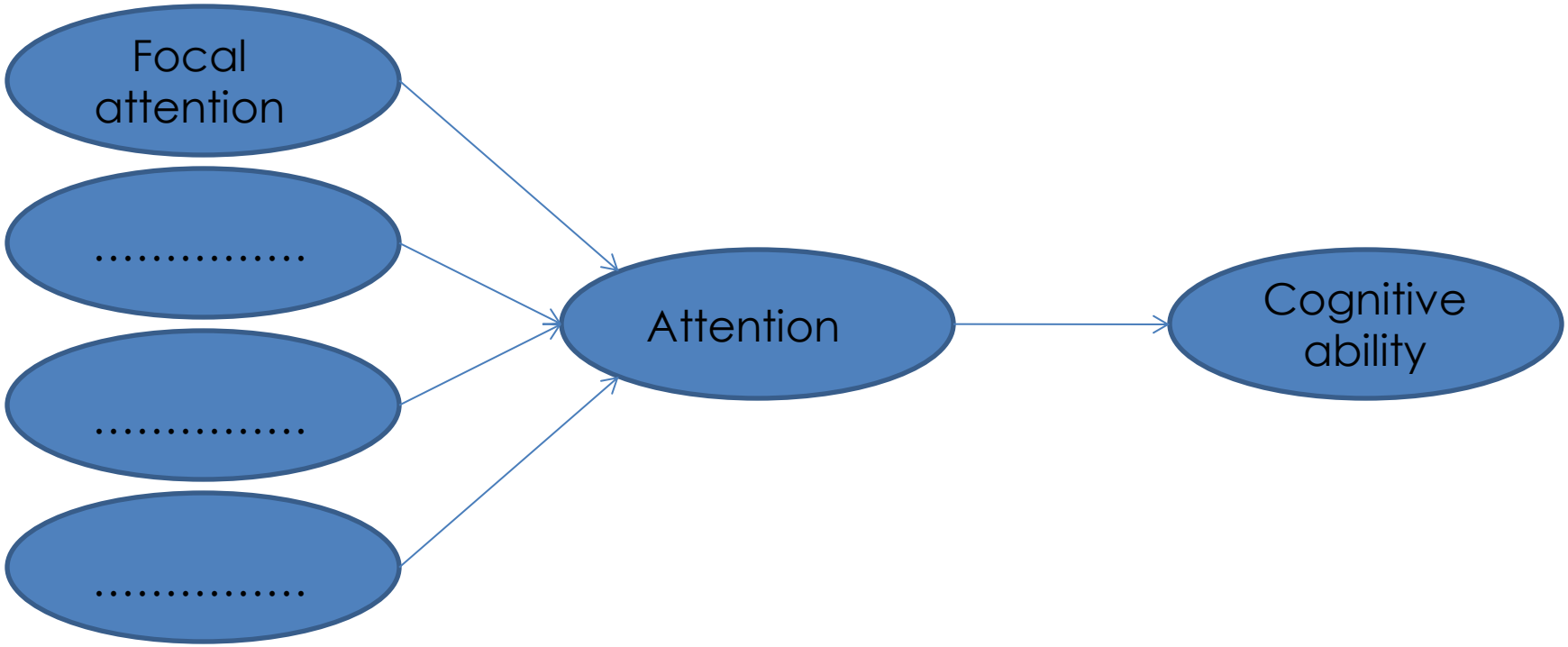
Does attention based on attention types (focal attention, concentration, sustained attention, vigilance, etc.) predict cognitive ability?

Hypothesis

Attention predicted by all attention types (…) predicts cognitive ability

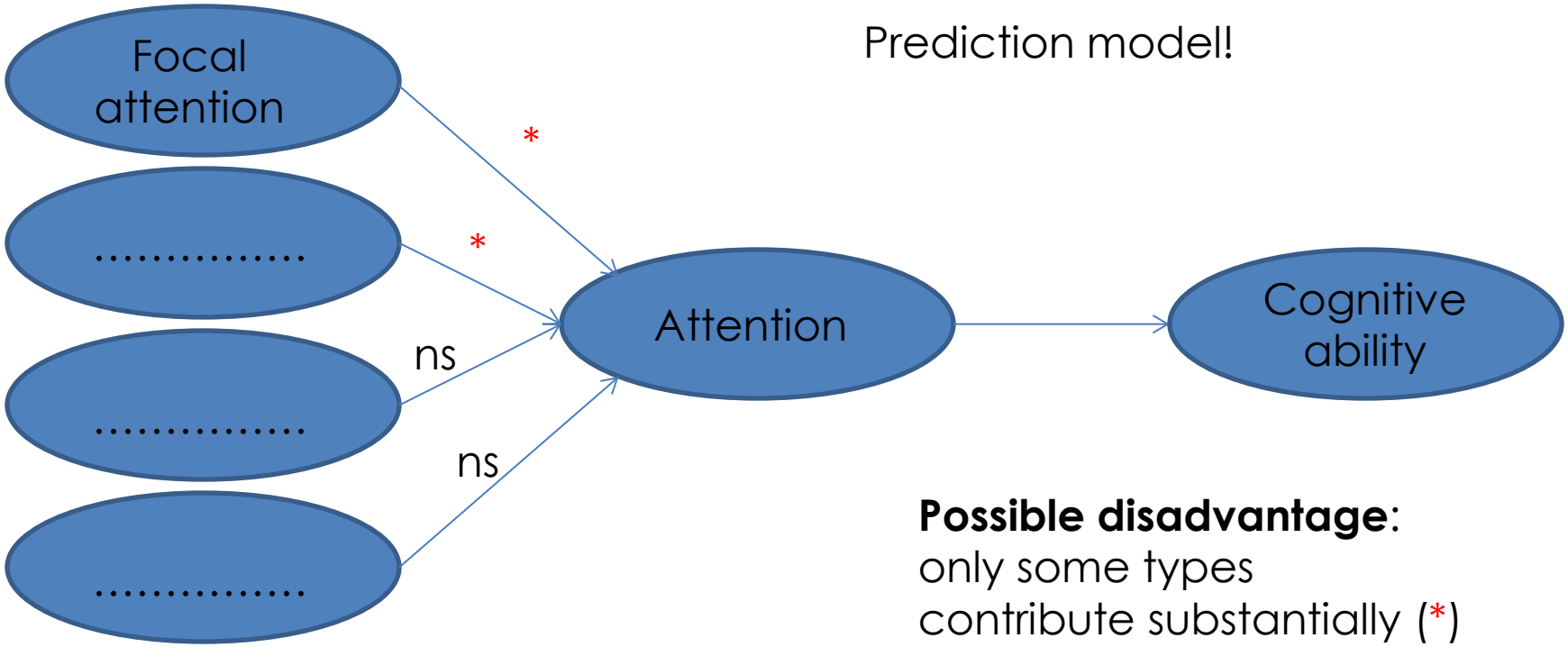
Types of models: ***application***

Model 2

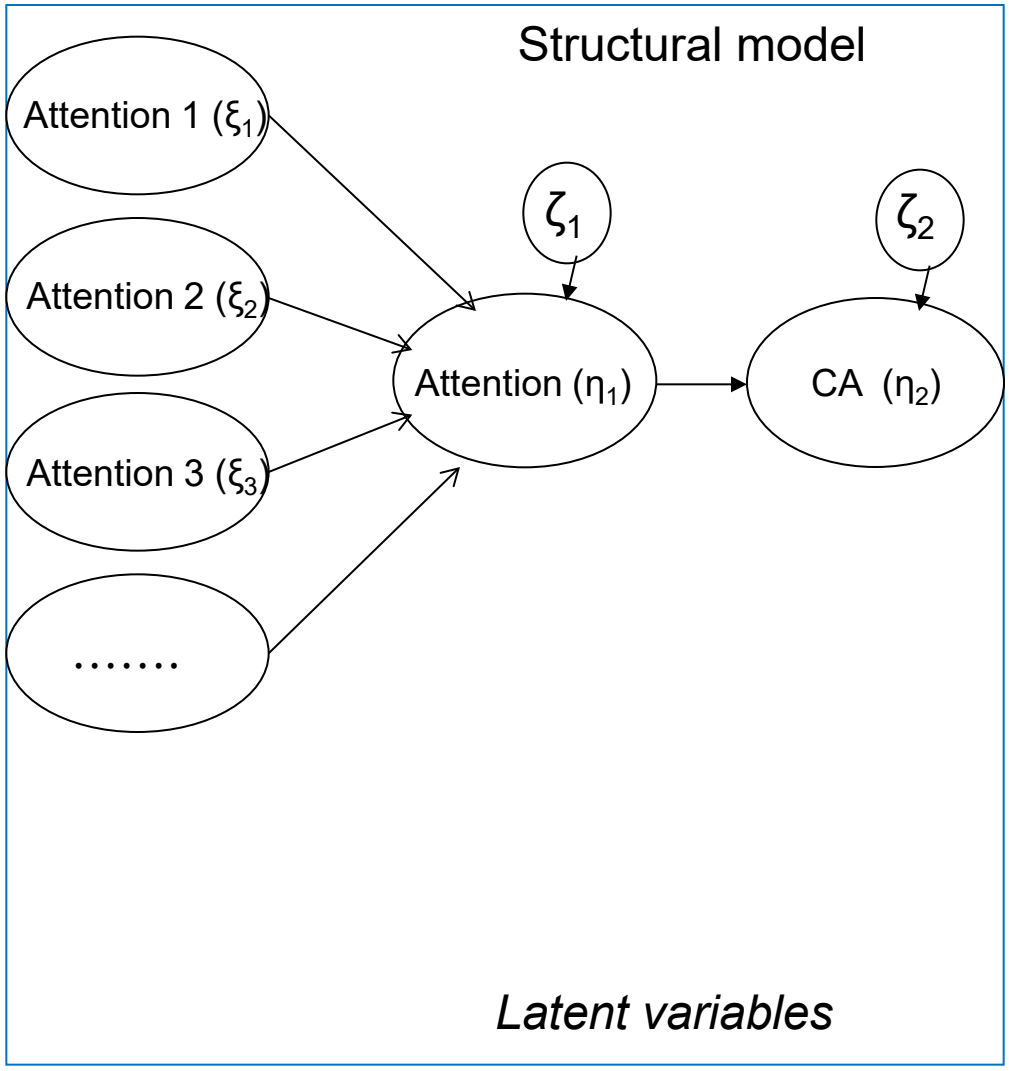


Types of models: ***application***

Model 2 (attention types as predictors)



Types of models: ***application***

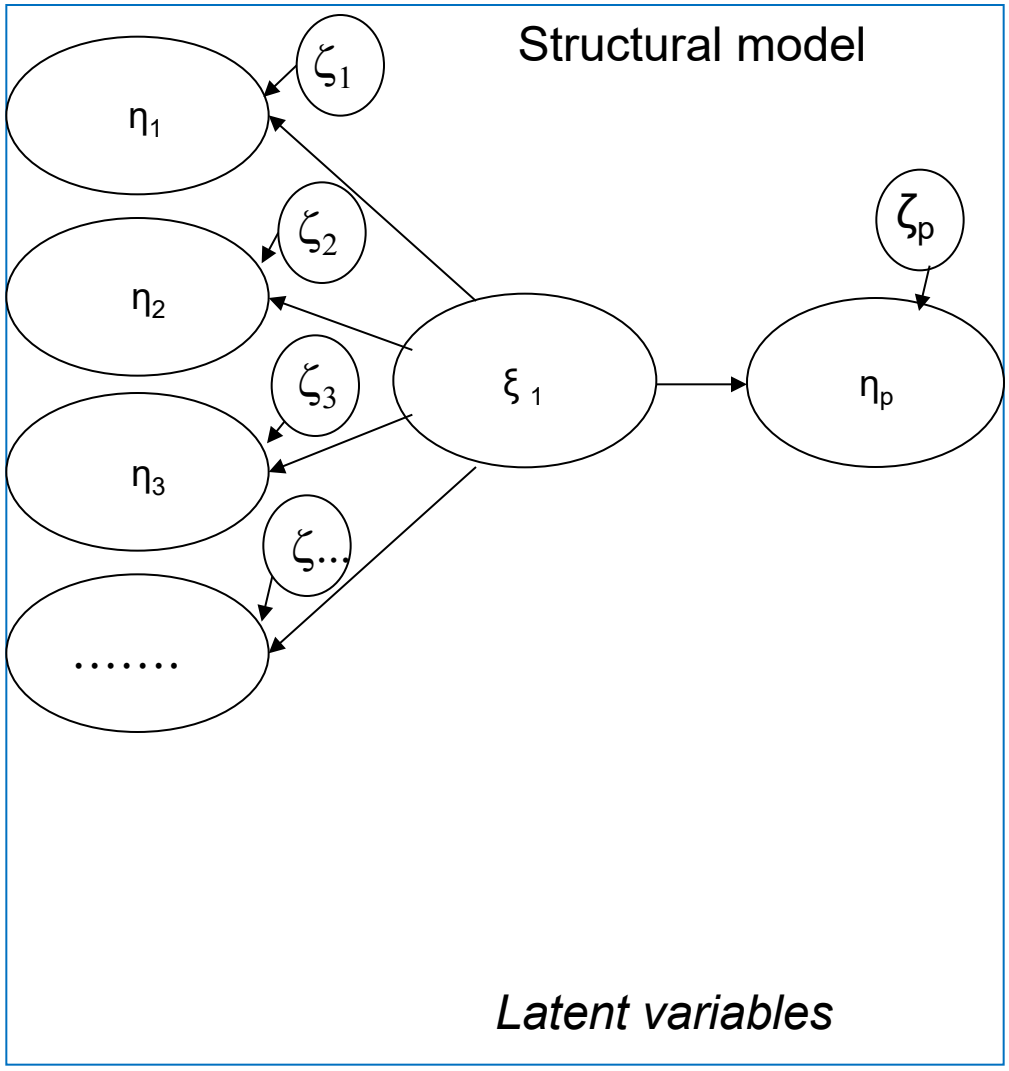


Basic model structure:

Linear prection model with mediation

Disadvantages: (1) only some predictors predict and (2) the hierarchical structure of attention is not considered!

Types of models



Basic model structure:

Hierachical linear prection model
(see hierachical model of measurement and linear prediction model)

Types of models: ***application***

Research question:

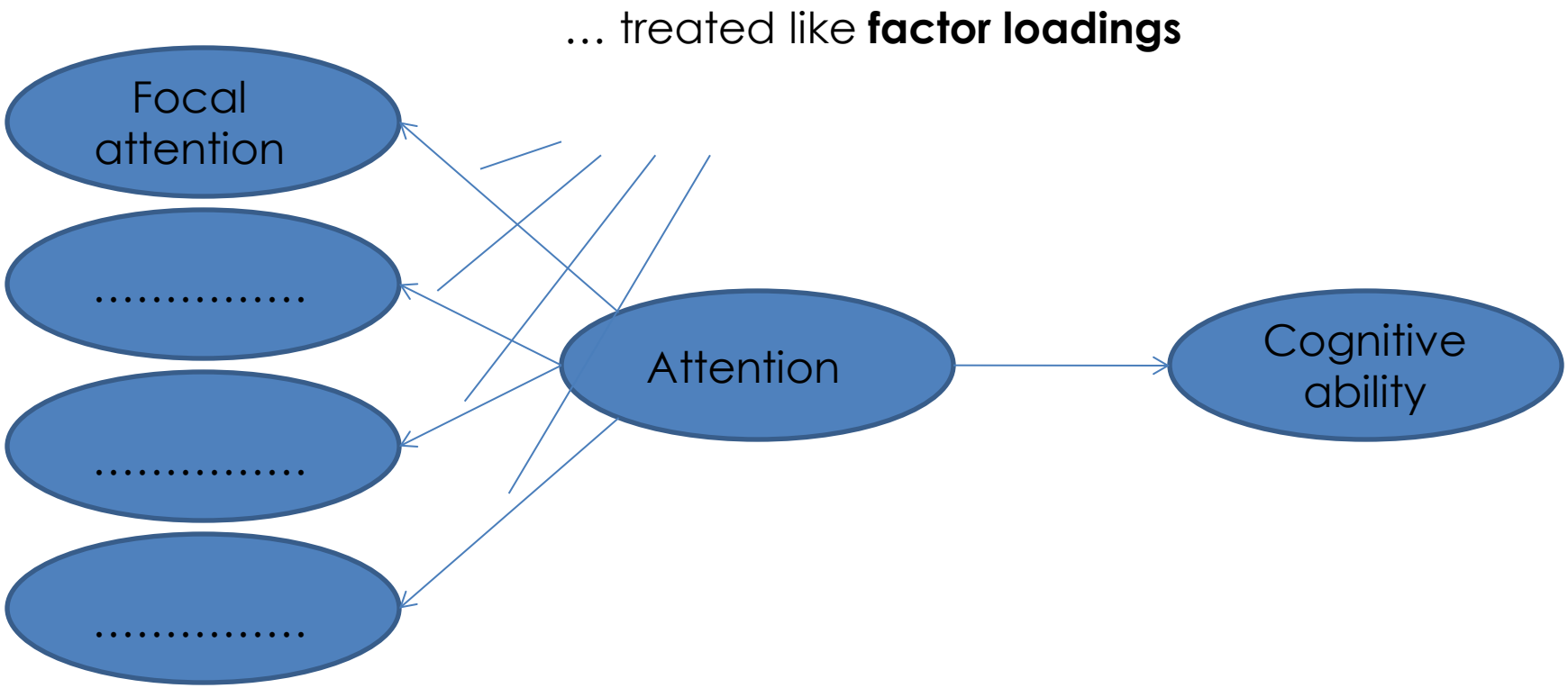
Does attention based on attention types (focal attention, concentration, sustained attention, vigilance, etc.) as *upper-level latent variable* predict cognitive ability?

Hypothesis

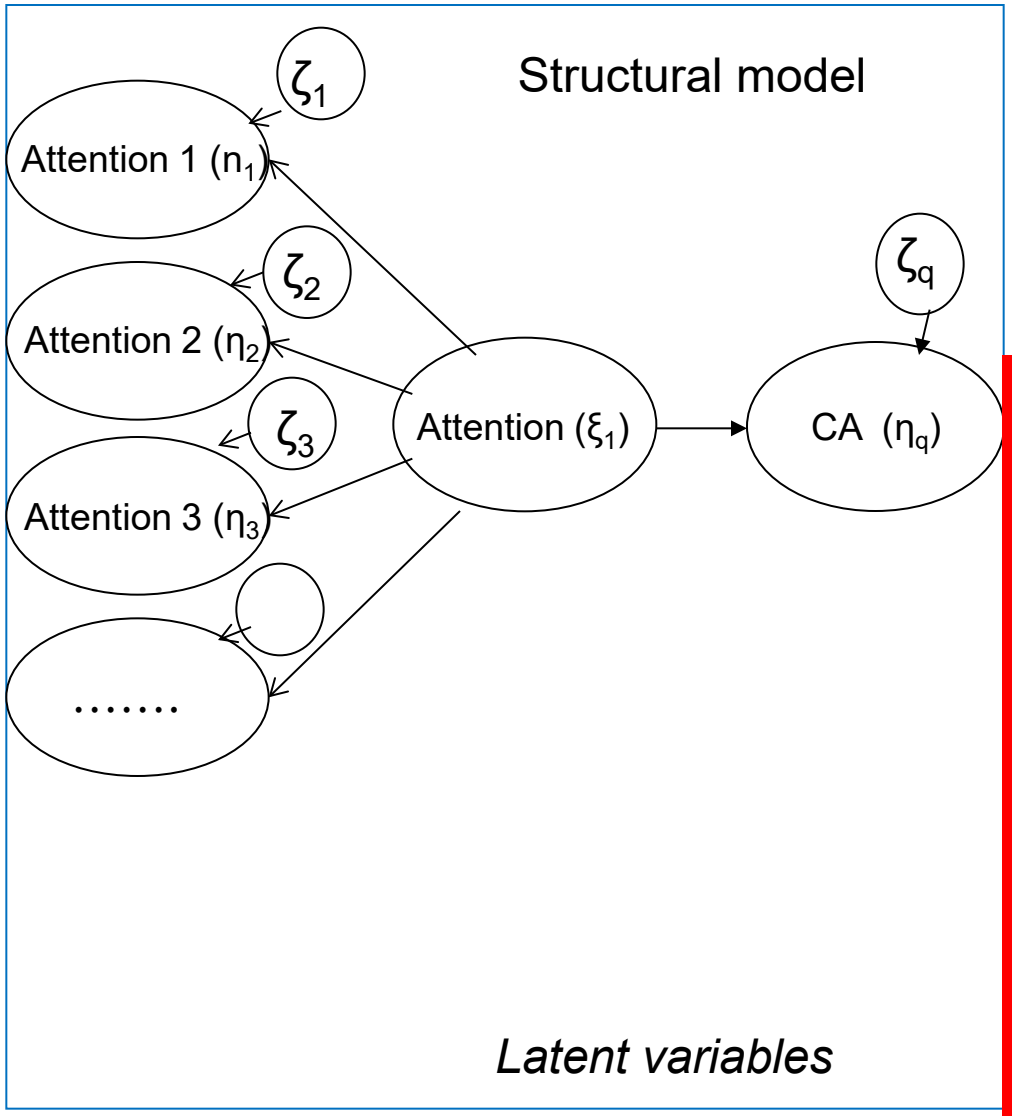
Attention including all attention types (…) predicts cognitive ability

Types of models: ***application***

Model 3



Types of models: *application*



Research question:
Does attention based on attention types (focal attention, concentration, sustained attention, vigilance, etc.) as *upper-level latent variable* predict cognitive ability?

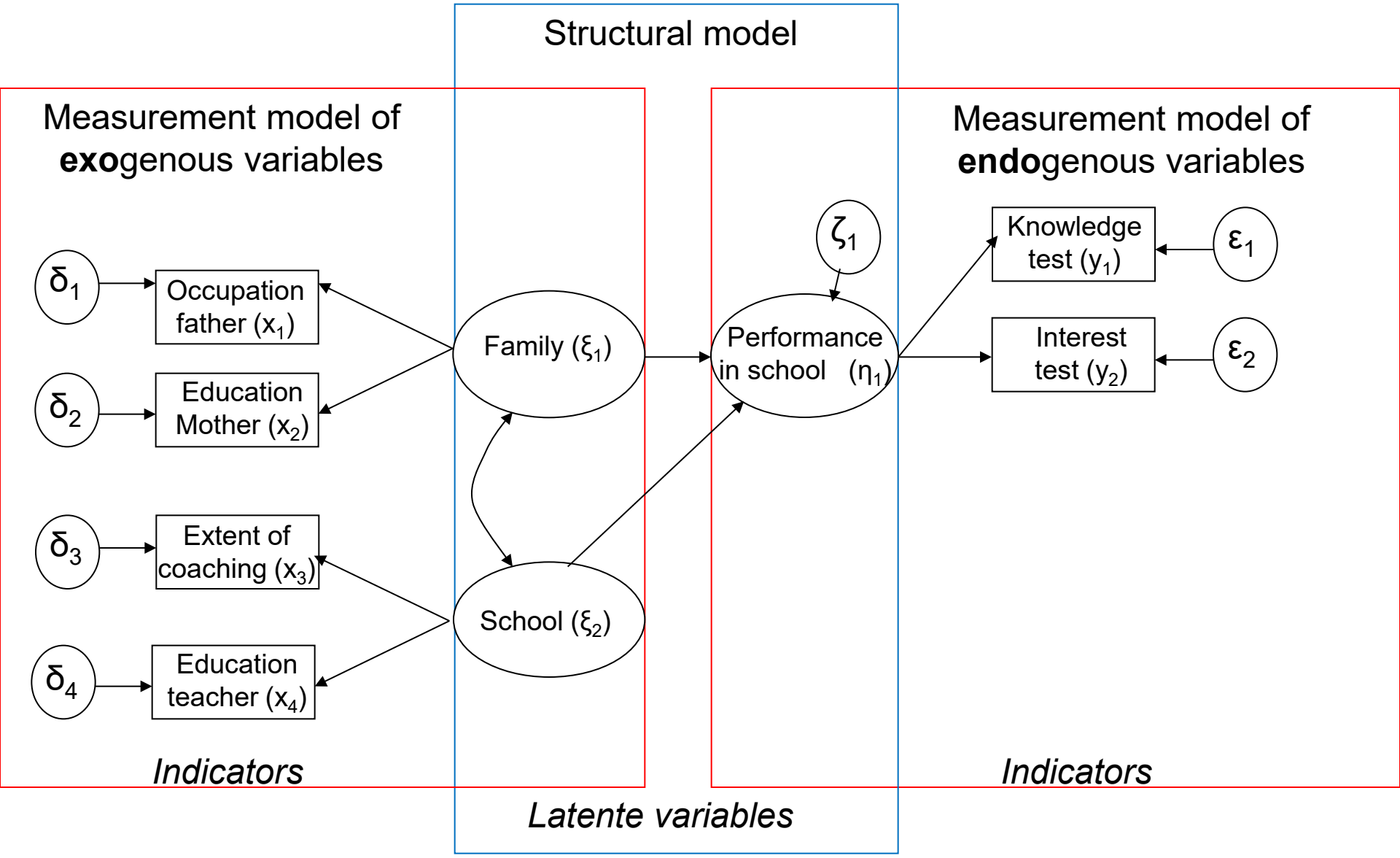
Outline

1. Introduction
2. The rational
3. The complete structural equation model
4. Types of models
5. *The strategy of analysis*
6. The evaluation ofmodels

The strategy of analysis

The strategy of analysis refers to the successive steps in data analysis and the decisions to be made

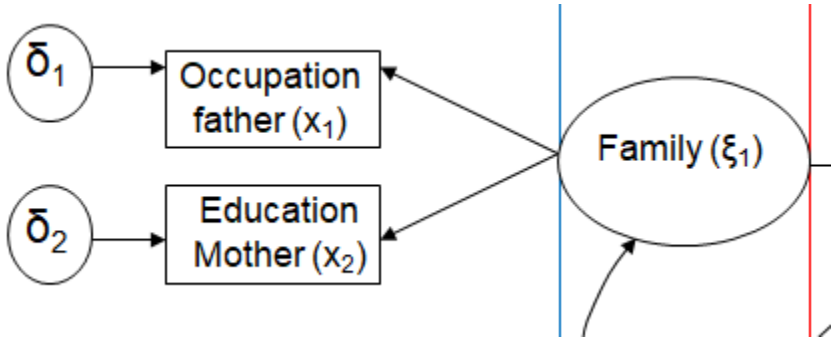
The to-be-investigated model



The strategy of analysis

- At first **model fit** of each model of measurement *of each latent variable* is investigated

  e.g.



The strategy of analysis

- At first model fit of each model of measurement *of each latent variable* is investigated
- If one of these models of measurement does not fit, the whole process is **terminated**

  … *if there is* ***no way out***!

The strategy of analysis

- …. *if there is no way out*!
- … such ways are:

  ♦ *combining indicators (parceling)*
  ♦ *equalizing residual variances*
  ♦ *equalizing factor loadings*
  ♦ *eliminating this measurement model from the complete model*

  …

The strategy of analysis

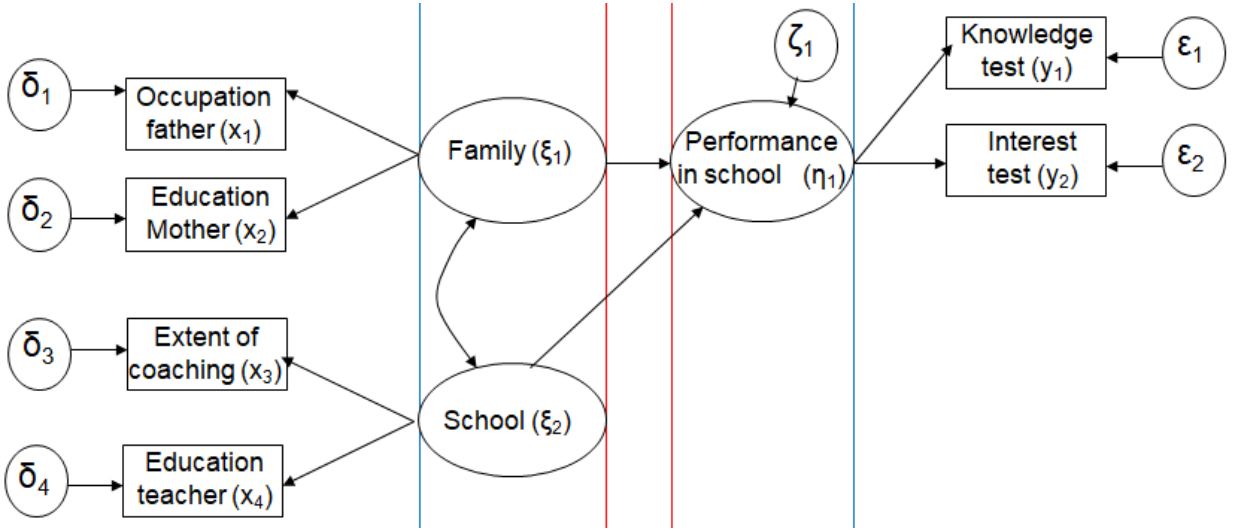
- …. ***if there is no way out***!
- … such ways are:

… … … … …

Note … These ways only apply to complete SEM models but not to CFA models

The strategy of analysis

- If all models of measurement show sufficient model fit, the complete model is investigated, e.g.



The strategy of analysis

- If all models of measurement show sufficient model fit, the complete model is investigated

- The strategy ofCFA requiring the investigation of alternative models does **not apply** to the complete model

- The complete model represents a hypothesis that requires …
  … either confirmation or rejection

The strategy of analysis

Comments regarding measurement models:

- Formerly in the case of misfit, **modifications** based on modification indices were recommended
- Today modifications are considered as **inadmissible manipulations** since the confirmatory character of the approach is lost

… if the hypothesis is modified.

Outline

1. Introduction
2. The rational
3. The complete structural equation model
4. Types of models
5. The strategy of analysis
6. *The evaluation of models*

The evaluation of models

- After conducting the statistical analysis by means of a software program (that simply has to be started after the specification of the model) an output becomes available that includes two types of information:

  - info on *parameter estimates*
  - info on *model fit*

The evaluation of models

- Fit information:
  … is used for evaluating whether there is a good degree of model fit (or at least an **acceptable** one).

  … if there is model misfit instead: the model is **given up**!

The evaluation of models

- Fit information:
  … is used for evaluating whether there is a good degree of model fit (or at least an **acceptable** one).

  … in the positive case: the evaluation proceeds to what is the meaning for the research question / hypothesis

The evaluation of models

- Example:

  $\chi 2$: 209.5

  RMSEA: .048 ✓ (=<.06)

  SRMR: .075 ✓ (=<.08)

  CFI: .96 ✓ (>=.95)

  NNFI: .95 ✓ (>=.95)

  AIC: 310.5

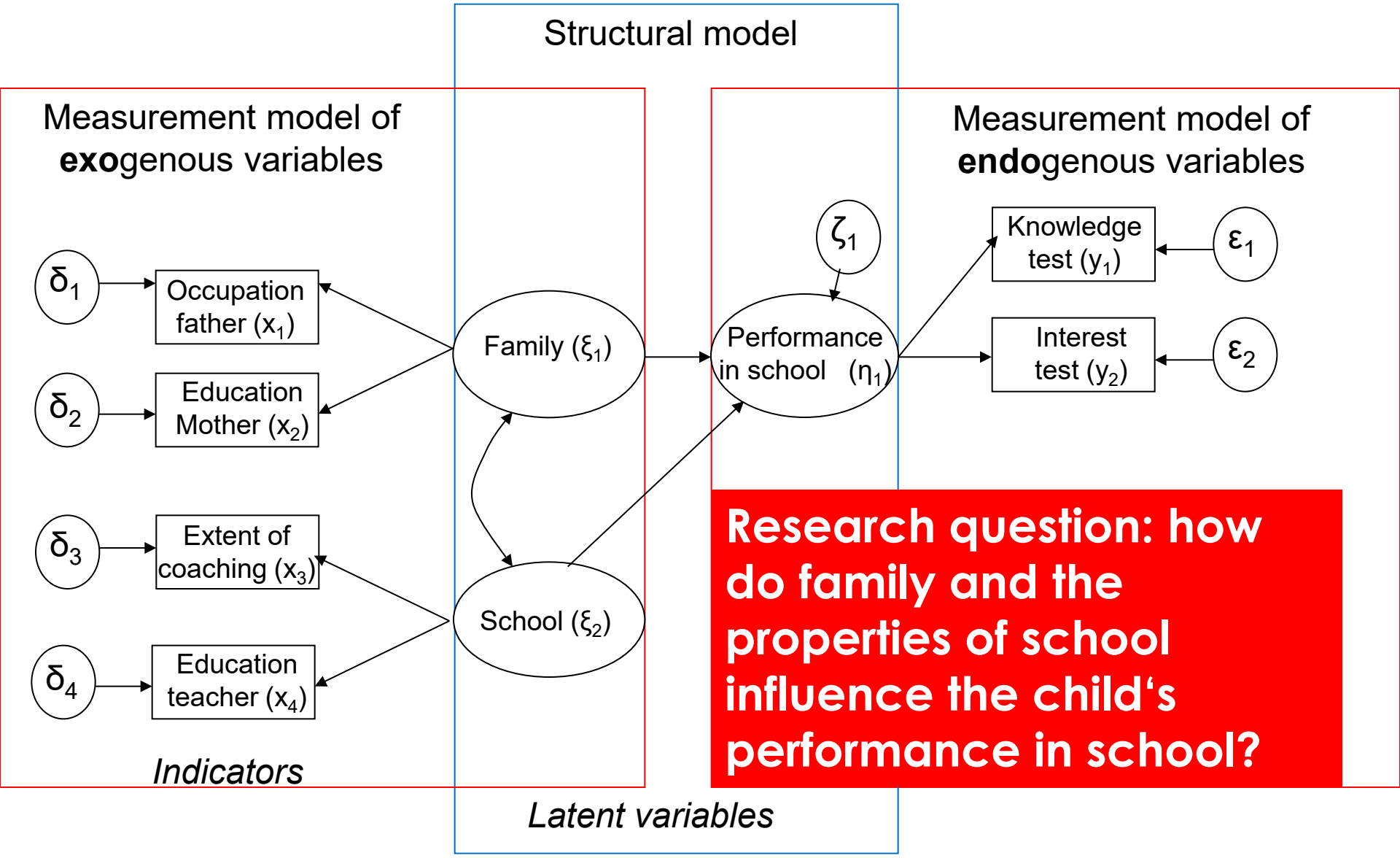
The evaluation of models

… in the positive case: the evaluation proceeds to what is the meaning for the research question / hypothesis

- Parameter information:

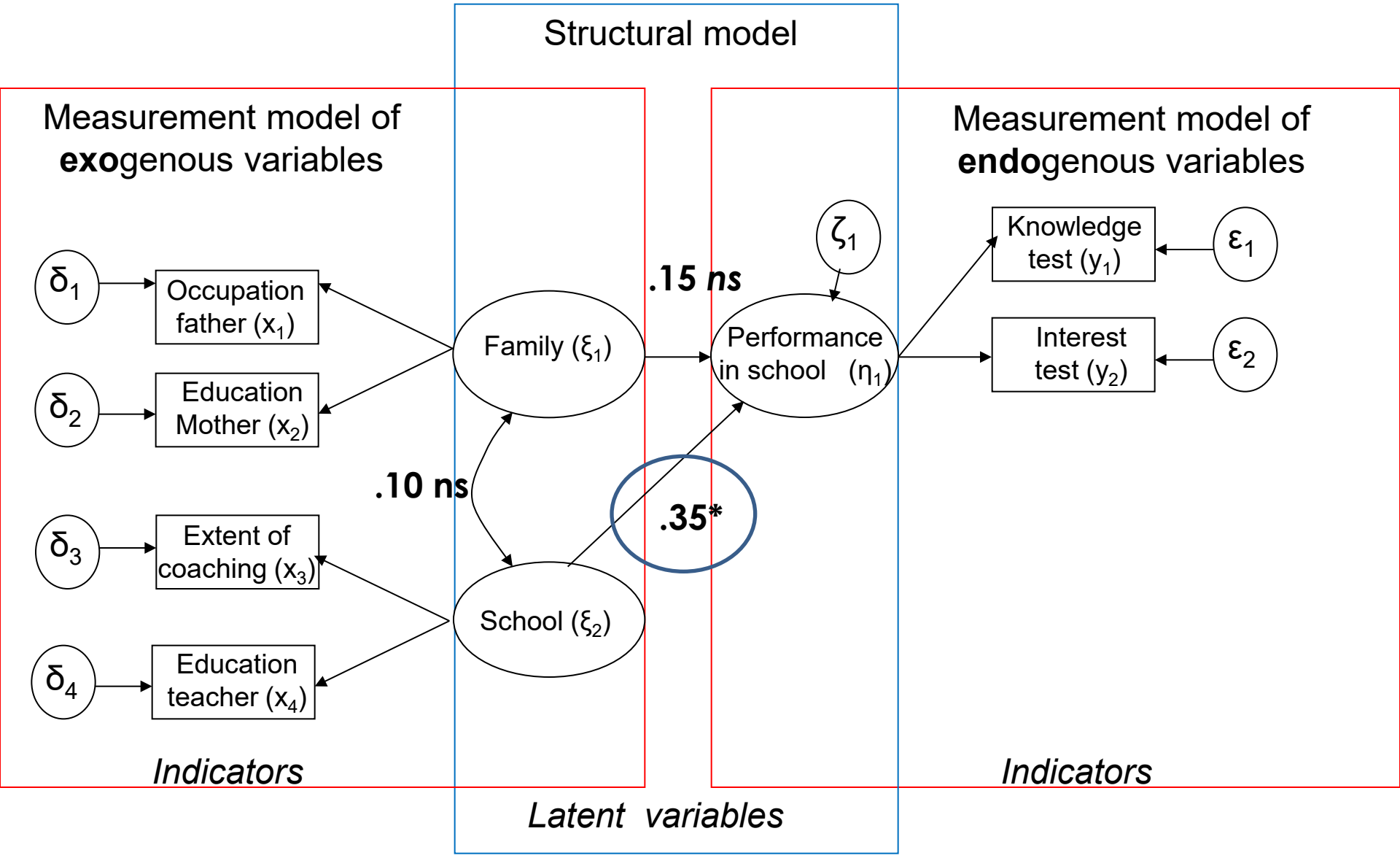
  The parameter estimates are used for providing an answer to the research question / ***for evaluating the hypothesis***.

Remember the following example



Research question: how do family and the properties of school influence the child's performance in school?

Remember the following example



The evaluation ...
The example

The gamma parameters indicate that the family has no influence on the childrens‘s school performance whereas the school has one.

In the case of no support of a hypothesis: What are possible reasons?

The evaluation ...
The example

The gamma parameters indicate that the family has no influence on the childrens‘s school performance whereas the school has one.

Possible reasons can eventually be found in the selection of indicators for family ...

Summary and brush up:

1. Introduction … confirmation that the elimination of error influences parameter estimates on the latent level

2. The rational … SEM can be perceived as combination of FA and MR / Path Analysis

3. The complete structural equation model … gives rise to a complex model of the covariance matrix

4. Types of models

5. The strategy of analysis … find out about confirmation / rejection of hypotheses

6. The evaluation of models … compare observed fit statistics and cutoffs / check estimates

QUESTIONS REGARDING COURSE UNIT 7

- Why are model modifications in SEM not acceptable for achieving good model fit?
- What has to be done in SEM if one model of measurement indicates bad model fit?
- List types of structural equation models
- What is the meaning of the step from the manifest to the latent level

Literature

- **Kline, R. B. (2011). *Principles and practices of structural equation modeling* (3rd edition) (Chapter 1: Introduction). New York, NY: The Guilford Press.**