Introduction to Structural Equation Modeling

Enrico Perinelli

Department of Psychology and Cognitive Science
University of Trento

Who am I?

Academic

Researcher in Tenure Track (RTT) *July 1 2024 – ongoing*
Work and Organizational Psychology (M-PSI/06)

Research Fellow (RTD-a) *July 1, 2021 – June 30, 2024*
Work and Organizational Psychology (M-PSI/06)

Postdoctoral Researcher – University of Trento *Sept 15, 2018 – June 30, 2021*
Work and Organizational Psychology (M-PSI/06)

Education

Ph.D. – Sapienza University of Rome *Feb 2018*
Personality and Organizational Psychology

M.S. – University of Bologna *Feb 2013*
Clinical Psychology

B.S. – University of L'Aquila *Oct 2010*
Psychology

Who am I?

Research Areas

1. Organizational Psychology, Personality Psychology, Psychometrics
2. Educational Psychology and Individual Differences

Research Topics

1. Social desirability scales
2. Individual differences in personality from childhood (moral disengagement, effortful control) to Adulthood (big five, positivity)
3. Methodological and theoretical issues in the study of self-esteem
4. Personality factors and organizational outcomes (e.g., Well-Being at work)
5. Methodological issues in organizational psychology and management
6. Non-cognitive skills in junior high school students

Who am I?

Research Projects and Practical Applications

1. Guardia di Finanza (Sapienza University of Rome)
2. Mobbing (Autonomous Province of Trento and Univ of Trento)
3. Nudging and Unemployment (Autonomous Province of Trento and Univ of Trento)
4. Non-cognitive skills in Junior high school students (Fondazione Caritro and Univ of Trento)
5. Member of the WebeWo LAB -> Interdisciplinary lab (Law, Psychology, Sociology) that aims to study interventions, antecedents, and outcomes of well-being at work
6. PRIN PNRR 2022 – P.I. of the project «*Loneliness, non-cognitive skills, and academic achievement in Junior High School: Assessment and intervention based on Big Data, longitudinal, and intensive methods*»

More info at

https://webapps.unitn.it/du/it/Persona/PER0209265/Curriculum

Who Are You?

….please, introduce your

- *name*
- *background*
- *skills*
- *acquired knowledge*
- *expectations about this course….. and the PhD in general*

This is useful to start creating a collaborative, positive, and multidisciplinary organizational climate…

Schedule

- 05 December 2024 – Aula 3
- 06 “ “ – “
- 12 “ “ – “
- 13 “ “ – “
- 18 “ “ – “ (*Project Presentation*)

Your project for this PhD course will involve one of the following options:

- A theoretical or practical argument focusing on a Structural Equation Modeling (SEM) technique (e.g., mediation with SEM, Longitudinal SEM, Latent Growth Modeling, etc.)
- The examination and discussion of one or more articles that report SEM findings.
- Data analysis using M*plus* or R. You have the flexibility to choose whether to analyze your own dataset, an open-access dataset, or a simulated dataset.
- Hypothesize a SEM pipeline – from Model Formulation to Model Evaluation – accompanied by hypothetical codes (in R or M*plus*)

Let’s start our journey...

The need for Latent Variable Models

Problems with *Measurement*

The focus of Psychology is the study of human behaviors and human cognitive processes. Thus, it is difficult to think at tools capable to assess them.

The need for Latent Variable Models

Problems with *Measurement*

In 1940 a committee established by the British Association for the Advancement of Science concluded that measurements taken in psychological methods could not be considered “scientific measurements”.

The need for Latent Variable Models

Solutions for improving *Measurement*

Stevens (1946)

MEASUREMENT is defined as the assignment of numerals to objects or events according to rules

SCIENCE
Vol. 103, No. 2684 Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens
Director, Psycho-Acoustic Laboratory, Harvard University

FOR SEVEN YEARS A COMMITTEE of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant by the term measurement. An interim report in 1938 ... by the formal (mathematical) properties of the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered.

A CLASSIFICATION OF SCALES OF MEASUREMENT

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that

The need for Latent Variable Models

Solutions for improving *Measurement*

Stevens (1946)

SCALES
Nominal
Ordinal
Interval
Ratio

TABLE 1

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible Statistics (invariantive)
NOMINAL	Determination of equality	*Permutation group* $x' = f(x)$ $f(x)$ means any one-to-one substitution	Number of cases Mode Contingency correlation
ORDINAL	Determination of greater or less	*Isotonic group* $x' = f(x)$ $f(x)$ means any monotonic increasing function	Median Percentiles
INTERVAL	Determination of equality of intervals or differences	*General linear group* $x' = ax + b$	Mean Standard deviation Rank-order correlation Product-moment correlation
RATIO	Determination of equality of ratios	*Similarity group* $x' = ax$	Coefficient of variation

The need for Latent Variable Models

Solutions for improving *Measurement*

Self-report questionnaires, Other-report questionnaires, Objective measures, ecc.

<u>*Two pivotal concepts:*</u>

RELIABILITY → Precision of a measurement ($X_i = T_i + E_i$)

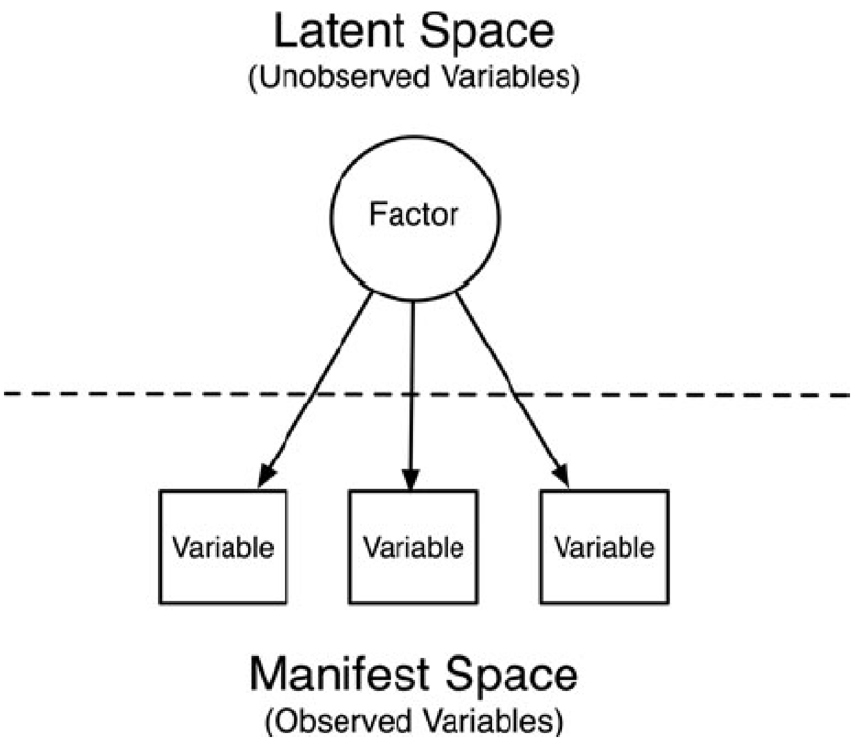
VALIDITY → The instrument is a proper tool for measuring what it aims to measure.

An instrument may be reliable, but not valid (e.g., I cannot use a "reliable" ruler for measuring intelligence)

The need for Latent Variable Models

Solutions for improving *Measurement*

The focus on statistical methods has been pivotal for the advancement of the concept of measurement in Psychology



From *Nesselroade & Molenaar (2016)*

The need for Latent Variable Models

Solutions for improving *Measurement*

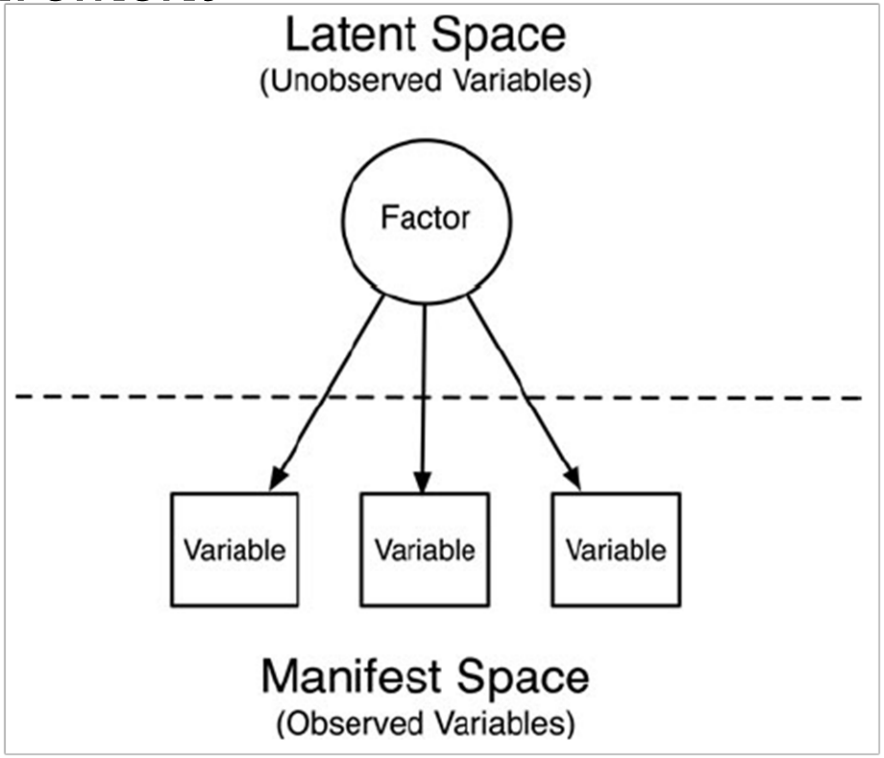
Operationalization Process

Construct

Instrument

Indicators

Variables



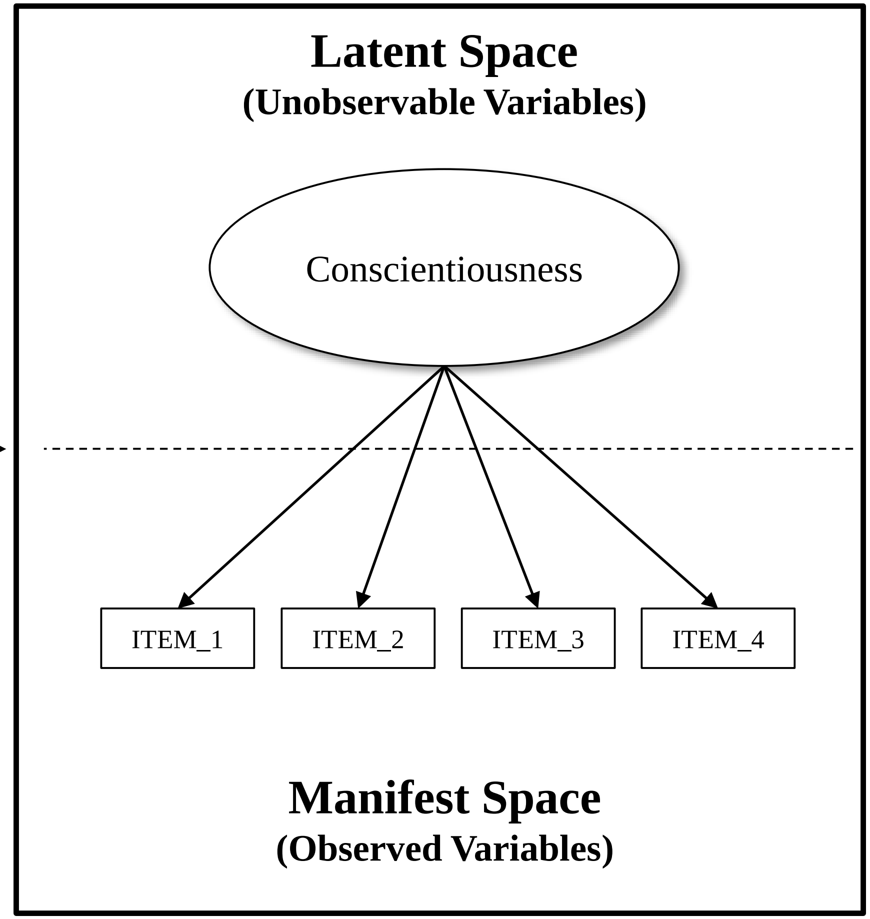
Operationalization Process: An Example

Conscientiousness

"Scrupulousness" (orderliness and precision) and "Perseverance" (capability of fulfilling one's own tasks and commitments)

Big Five Questionnaire
(only selected items that refer to conscientiousness)

ITEM_1 *I tend to be very thoughtful*
ITEM_2 *Before completing a job I spend a lot of time revising it*
ITEM_3 *It's difficult for me to give up an activity I've undertaken*
ITEM_4 *If I fail in a task, I keep trying until I succeed*



The need for Latent Variable Models

Reliability: Consistency (precision) of a measure

- Reliability (General formula)

ρ_{xx} = [$Var(T)$ / $Var(T + E)$], with values ranging from 0 to 1

- Cronbach's Alpha

$X = Y_1 + Y_2 + Y_3 + \ldots + Y_k$

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} \sigma^2_{Y_i}}{\sigma^2_X}\right)$$

K = number of components (e.g., items)
$\sigma^2{}_{Yi}$ = variance of the component i
$\sigma^2{}_X$ = variance of the observed total score

- Omega total

$$\omega_{Total} = \frac{\left(\sum_{i=1}^{k} \lambda_i\right)^2}{\left(\sum_{i=1}^{k} \lambda_i\right)^2 + \sum_{i=1}^{k} \theta_{ii}}$$

λ_i = factor loading of the component i;
a factor loading represents the strenght of the relationship between an indicator and its latent variable
θ_{ii} = Residual variance of the component i

- AVE (Average Variance Extracted)

$$AVE = \frac{\left(\sum_{i=1}^{k} \lambda_i^2\right)}{\left(\sum_{i=1}^{k} \lambda_i^2\right) + \left(\sum_{i=1}^{k} \theta_{ii}\right)}$$

> .50

The need for Latent Variable Models

Validity: A test is valid if it measures what it purports to measure" (Borsboom et al., 2004, p. 1061)

"The problem of validity cannot be solved by psychometric techniques or models alone. On the contrary, it must be addressed by substantive theory"

(Borsboom et al., 2004, p. 1062).

Psychological Review
2004, Vol. 111, No. 4, 1061–1071

0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.4.1061

The Concept of Validity

Denny Borsboom and Gideon J. Mellenbergh
University of Amsterdam

Jaap van Heerden
Maastricht University

The need for Latent Variable Models

Validity: "A test is valid if it measures what it purports to measure" (Borsboom et al., 2004, p. 1061)

Validity

Qualitative/substantive

Evaluation of item content, comprehensiveness, and adeguacy

- Face validity (non-experts[a potential test-taker])
- Content validity (experts)

Internal (or structural)

Internal structure of the measure (e.g., number of factors, reliability)

External

Relationship between a measure and other constructs or measures

Compared to

- Construct validity
- Criterion validity

Size of effect

- Covergent validity (*large*)
- Discriminant validity (*low* or *null*)

The need for Latent Variable Models

Some references about validity

Some references to deepen Validity (particular focus on Work and Organiz Psych):

- Substantive Validity
  - Face Validity (Allen et al., 2023)
  - Content Validity (Colquitt et al., 2019; Rossiter, 2008)
- Structural (Internal) Validity
  - Reliability (Cortina et al., 2020)
  - Measurement Invariance (Somaraju et al., 2022)
- External Validity
  - Construct/Nomological/Criterion Validity (Lambert & Newman, 2023)
  - Convergent Validity (Cheung et al., 2024)
  - Discriminant Validity (Cheung et al., 2024; Rönkkö & Cho, 2022)

The need for Latent Variable Models

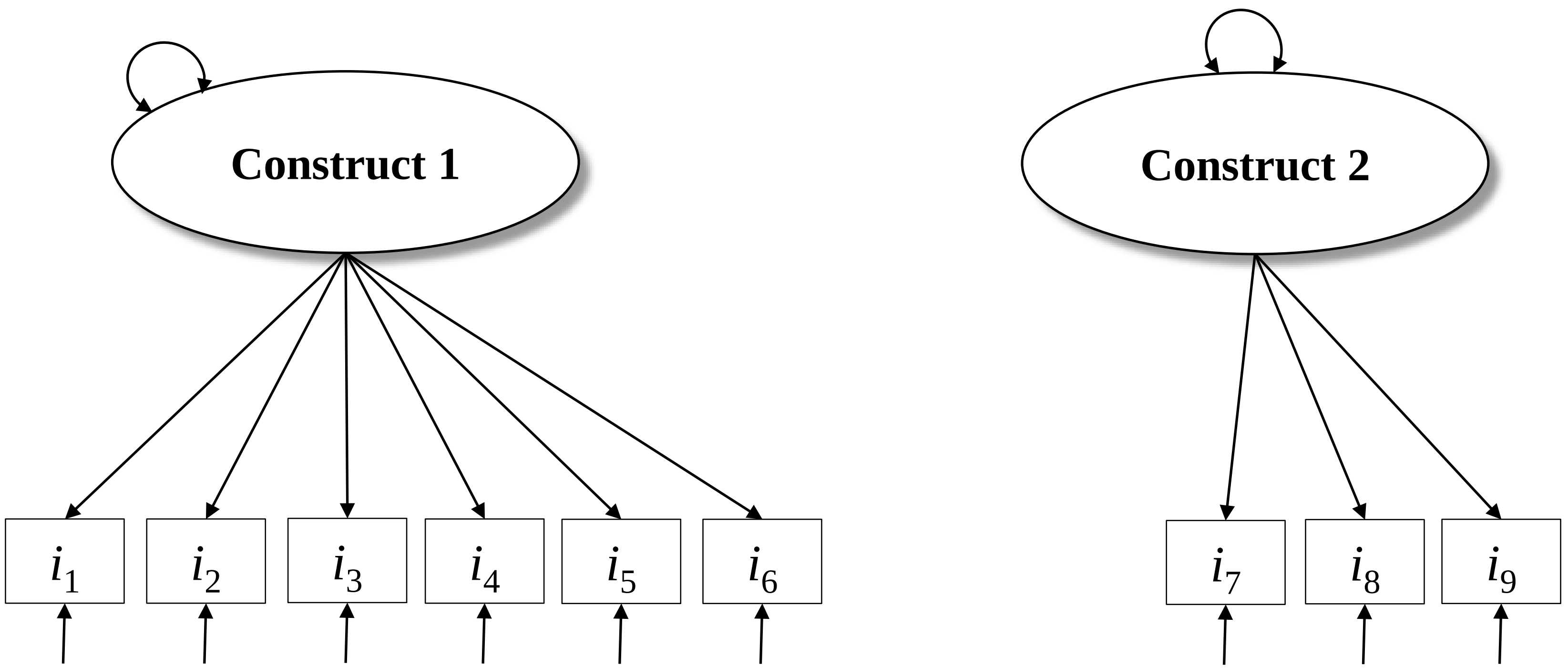
Researcher's main tasks in Personality Psychology

Measuring constructs
(Exploratory Factorial Analysis; EFA)

Connecting constructs
(Regression Models)

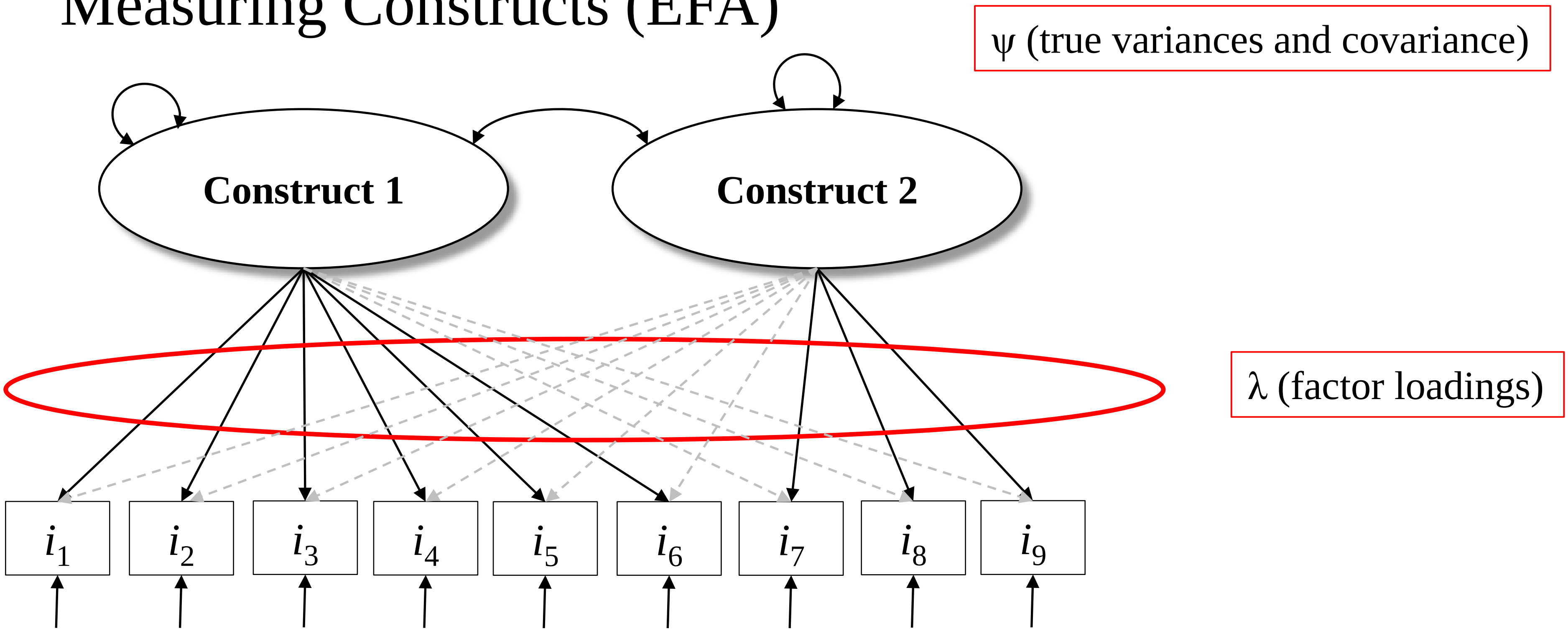
The need for Latent Variable Models

Measuring Constructs



The need for Latent Variable Models

Measuring Constructs (EFA)



Seeking Resources

1. I ask other feedback on my performance
2. I ask colleagues for advice [I ask other educational superintendents for advice]
3. I ask my supervisor for advice [I ask other managers for advice (not educational superintendents]
4. I try to learn new things at work
5. I contacted other people from work (e.g., colleagues, supervisors) to get the necessary information for completing my tasks
6. When I have difficulties or problems at my work, I discuss them with people from my work environment

Seeking Challenges

7. I ask for more tasks if I finish my work [I look for other projects to carry out, as soon as I conclude those planned]
8. I ask for more responsibilities
9. I ask for more odd jobs [I try to engage in projects/assignments that test myself]

The Likert scale ranged from 1 = *never* to 5 = *often*.

Hypothesized Construct	Item	1	2
Seeking Resources (α = .71)	1	**.34*****	.19***
	2	**.59*****	-.10***
	3	**.53*****	.18***
	4	**.46*****	.13***
	5	**.38*****	-.09***
	6	**.76*****	-.06**
Seeking Challenges (α = .83)	7	-.05*	**.61*****
	8	.09**	**.79*****
	9	.02$^{n.s.}$	**.82*****

N = 1060 superintendents
Latent Correlation = .40***

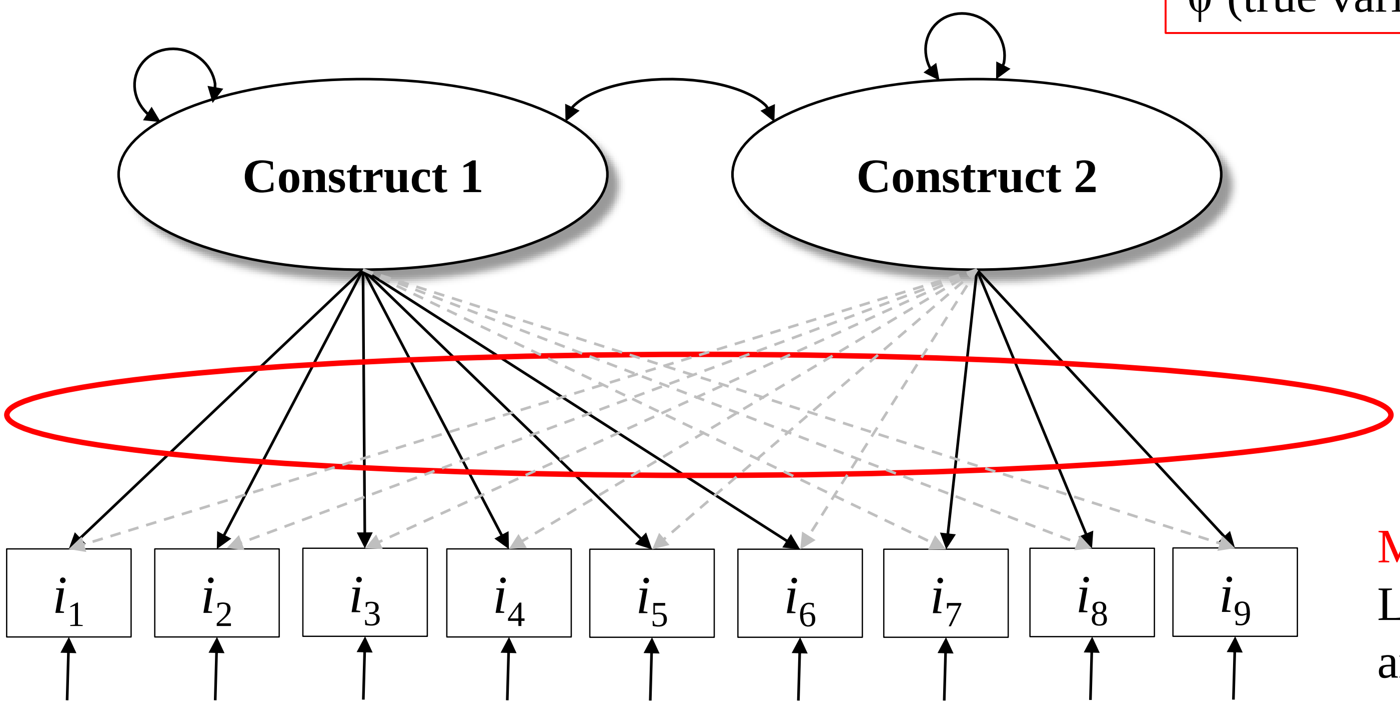
Recommended thresholds:

- $|\text{Target } \lambda| > .30$
- $|\text{Non-target } \lambda| < .30$

The need for Latent Variable Models

Measuring Constructs (EFA)

ψ (true variances and covariance)



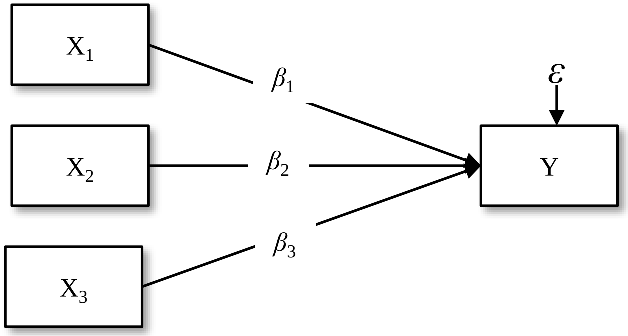
λ (factor loadings)

Main Limitation
Limited covariance(s) among contructs

The need for Latent Variable Models

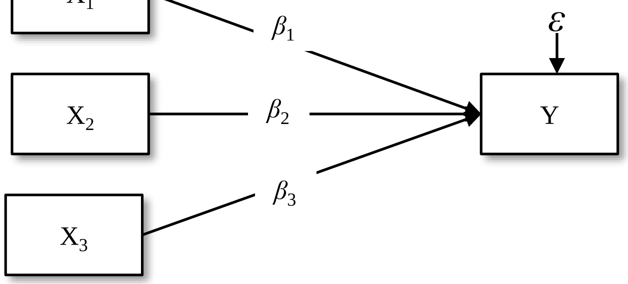
Connecting Constructs (Regression)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$



The need for Latent Variable Models

Connecting Constructs (Regression)



$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

Main Limitations

1) Don’t take into account measurement error
2) One Dependent Variable

Structural Equation Modeling (SEM)

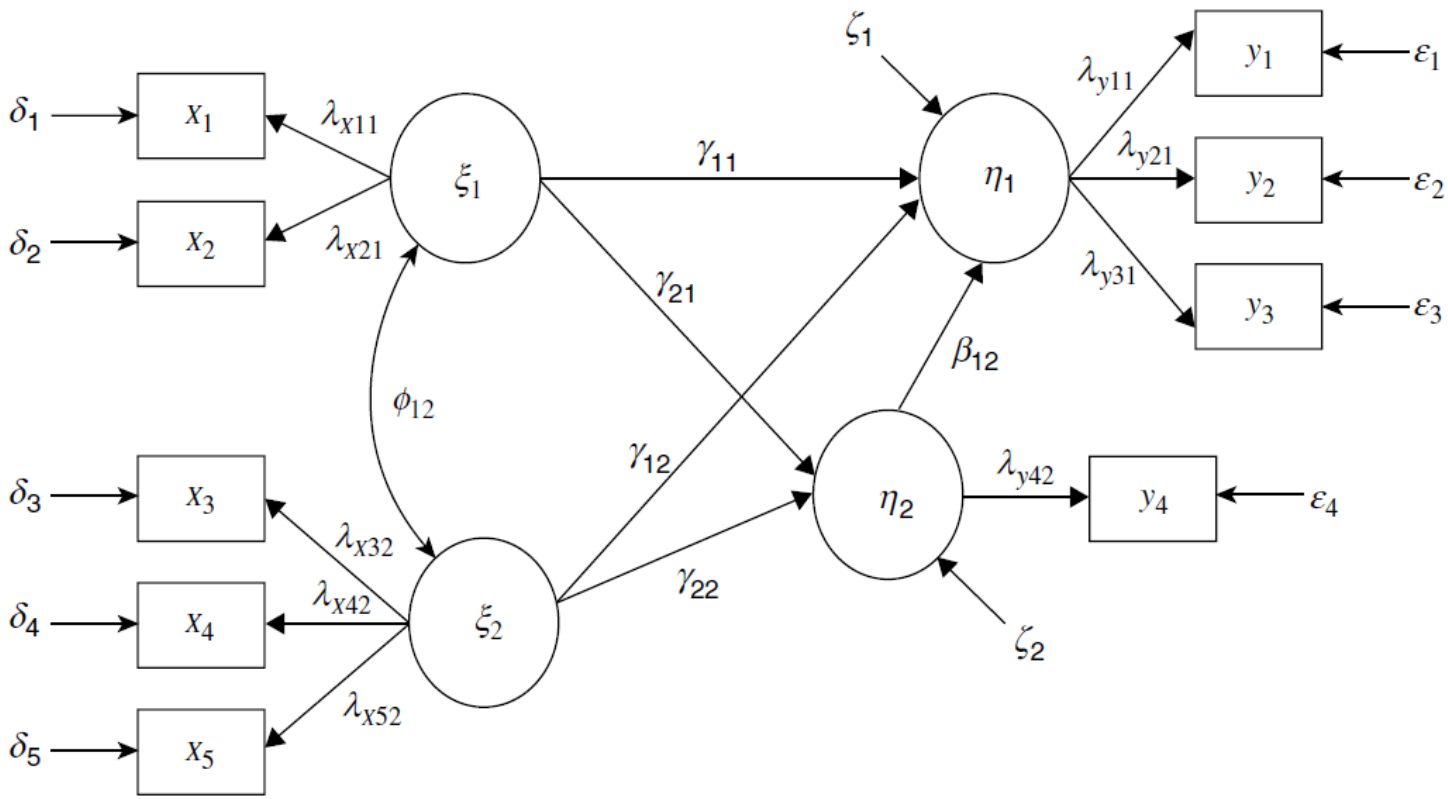


Figure 1.1 A hypothesized general structural equation model.

(Wang & Wang, 2012)

History of SEM

SEM (also called **covariance structure analysis**, **covariance structure modeling**, or **analysis of covariance structures**) does not designate a single statistical technique but instead refers to <u>a family of related procedures</u> (Kline, 2016, p. 9).

In SEM, unobservable latent variables (constructs or factors) are estimated from observed indicator variables, and the focus is on estimation of the relations among the latent variables free of the influence of measurement errors.

Measurement (Factor Analysis)
(Spearman, 1904; Tucker, 1955)

Simultaneous Equations (Path Analysis)
(Wright, 1918, 1921, 1934)

↓

SEM
(Jöreskog, 1967, 1969, 1973; Keesling, 1972; Wiley, 1973)

JKW Model

Matsueda, R. L. (2012). Key advances in the history of Structural Equation Modeling. In R. H. Hoyle (Eds.), *Handbook of Structural Equation Modeling* (pp. 17-42). New York, NY: Guilford

History of SEM

Matsueda (2012) argued that we can track **4 stages across SEM history**, each of which is characterized by a strong reliance on insights from statistics

(1) early disciplinary-specific developments of ***path analysis*** first from **genetics** and later **sociology**, ***factor analysis*** from **psychology**, and simultaneous equation models in **economics**;
(2) cross-disciplinary fertilization between economics, sociology, and psychology, leading to an **explosion of empirical applications** of SEM;
(3) a period of developing methods for handling **discrete, ordinal, and limited dependent variables**;
(4) a recent period of incorporating **statistical advances** into the SEM framework, including generalized linear models, mixed effects models, **mixture** regression models, **Bayesian** methods, graphical models, and methods for identifying **causal effects**.

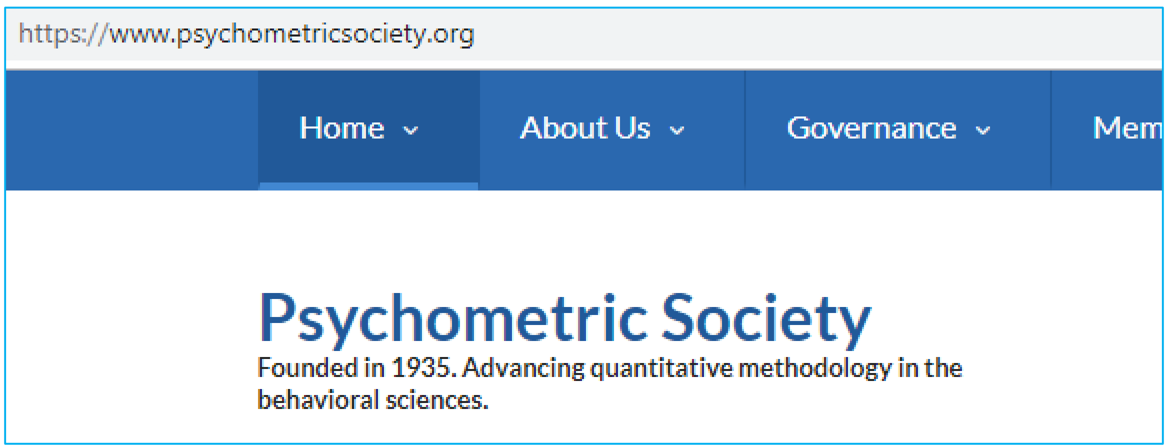
We are now approaching a fifth stage (e.g., Jacobucci & Grimm, 2020):
(5) Latent variables ↔ data science, artificial intelligence (machine learning)

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, *15*(3), 809-816. https://doi.org/10.1177/1745691620902467

Matsueda, R. L. (2012). Key advances in the history of Structural Equation Modeling. In R. H. Hoyle (Eds.), *Handbook of Structural Equation Modeling* (pp. 17-42). Guilford

SEM and Psychometrics

The study of quantitative measurement in psychology



Assessment

- ***INSTRUMENTS***
  Self-reports, Other-reports, Objective measures
- ***DESIGNS***
  Cross-sectional, Longitudinal, Hierarchical

Data Analytic Strategy

General Linear Models, Factor Analysis, Structural Equation Modeling, Cluster analysis

Simulation

Monte Carlo Methods, Performance of estimators, Optimal sample size

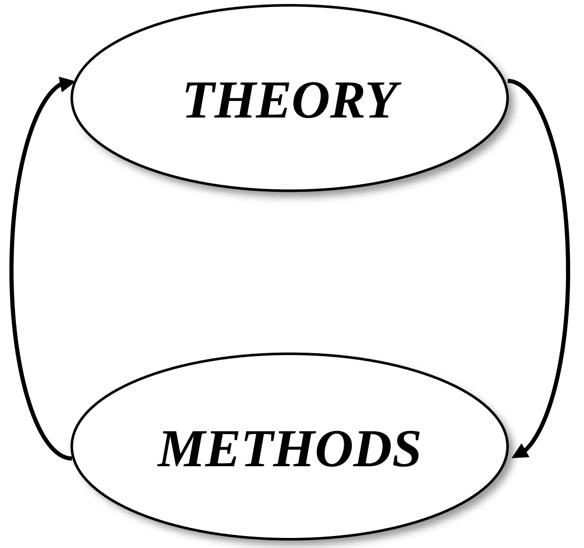
PSYCHOMETRIKA—VOL. 71, NO. 3, 425–440
SEPTEMBER 2006
DOI: 10.1007/s11336-006-1447-6

THE ATTACK OF THE PSYCHOMETRICIANS

DENNY BORSBOOM

UNIVERSITY OF AMSTERDAM

This paper analyzes the theoretical, pragmatic, and substantive factors that have hampered the integration between psychology and psychometrics. Theoretical factors include the operationalist mode of thinking which is common throughout psychology, the dominance of classical test theory, and the use of "construct validity" as a catch-all category for a range of challenging psychometric problems. Pragmatic factors include the lack of interest in mathematically precise thinking in psychology, inadequate representation of psychometric modeling in major statistics programs, and insufficient mathematical training in the psychological curriculum. Substantive factors relate to the absence of psychological theories that are sufficiently strong to motivate the structure of psychometric models. Following the identification of these problems, a number of promising recent developments are discussed, and suggestions are made to further the integration of psychology and psychometrics.



"The founding fathers of the Psychometric Society—scholars such as Thurstone, Thorndike, Guilford, and Kelley—**were substantive psychologists as much as they were psychometricians**.
Contemporary psychometricians do not always display a comparable interest with respect to the substantive field that lends them their credibility.
It is perhaps worthwhile to emphasize that, even though psychometrics has benefited greatly from the input of mathematicians, **psychometrics is not a pure mathematical discipline but an applied one**. If one strips the application from an applied science one is not left with very much that is interesting; and **psychometrics without the "psycho" is not, in my view, an overly exciting discipline**. It is therefore essential that a psychometrician keeps up to date with the developments in one or more subdisciplines of psychology." (Borsboom, 2006, p. 436)

Structural Equation Modeling

Steps

1. ***Model formulation*** Specifying the SEM model that the researcher wants to test. The model may be formulated on the basis of theory or empirical findings. A general SEM model is composed of two parts: the measurement model and the structural model.
2. ***Model identification*** It determines whether there is a unique solution for all the free parameters in the specified model.
3. ***Model estimation*** Estimate model parameters and generate fitting function. Various estimation methods are available for SEM (e.g., ML, WLS, Bayes, ecc.).
4. ***Model evaluation*** Assess whether the model fits the data.
5. ***Model modification*** If the model does not fit the data, re-specification or modification of the model is needed. In this instance, the researcher makes a decision regarding how to delete, add, or modify parameters in the model.

(Wang & Wang, 2012, p. 2)

Model Formulation

Steps

1. ***Model formulation*** Specifying the SEM model that the researcher wants to test. The model may be formulated on the basis of theory or empirical findings. A general SEM model is composed of two parts: the measurement model and the structural model.
2. ***Model identification*** It determines whether there is a unique solution for all the free parameters in the specified model.
3. ***Model estimation*** Estimate model parameters and generate fitting function. Various estimation methods are available for SEM (e.g., ML, WLS, Bayes, ecc.).
4. ***Model evaluation*** Assess whether the model fits the data.
5. ***Model modification*** If the model does not fit the data, re-specification or modification of the model is needed. In this instance, the researcher makes a decision regarding how to delete, add, or modify parameters in the model.

(Wang & Wang, 2012, p. 2)

Model Formulation

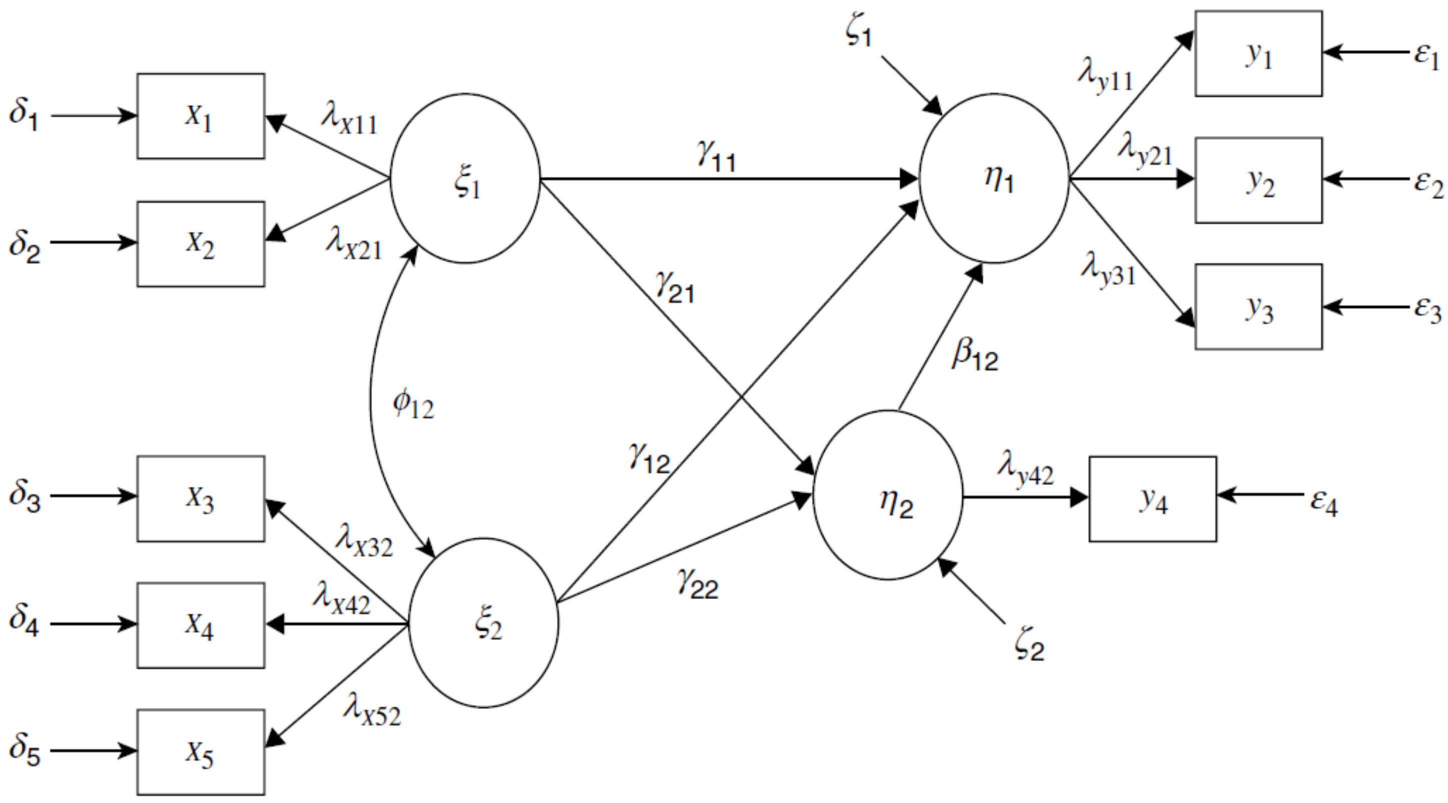


Figure 1.1 A hypothesized general structural equation model.

(Wang & Wang, 2012)

Model Formulation

Measurement Models

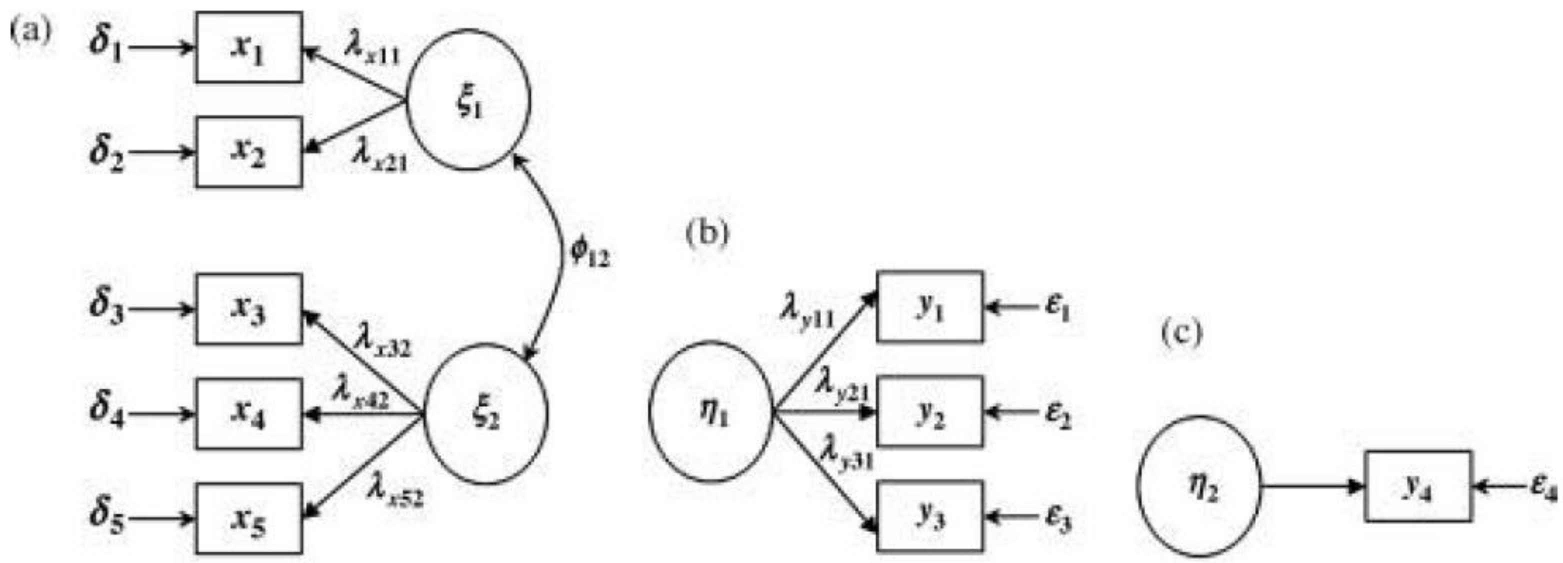
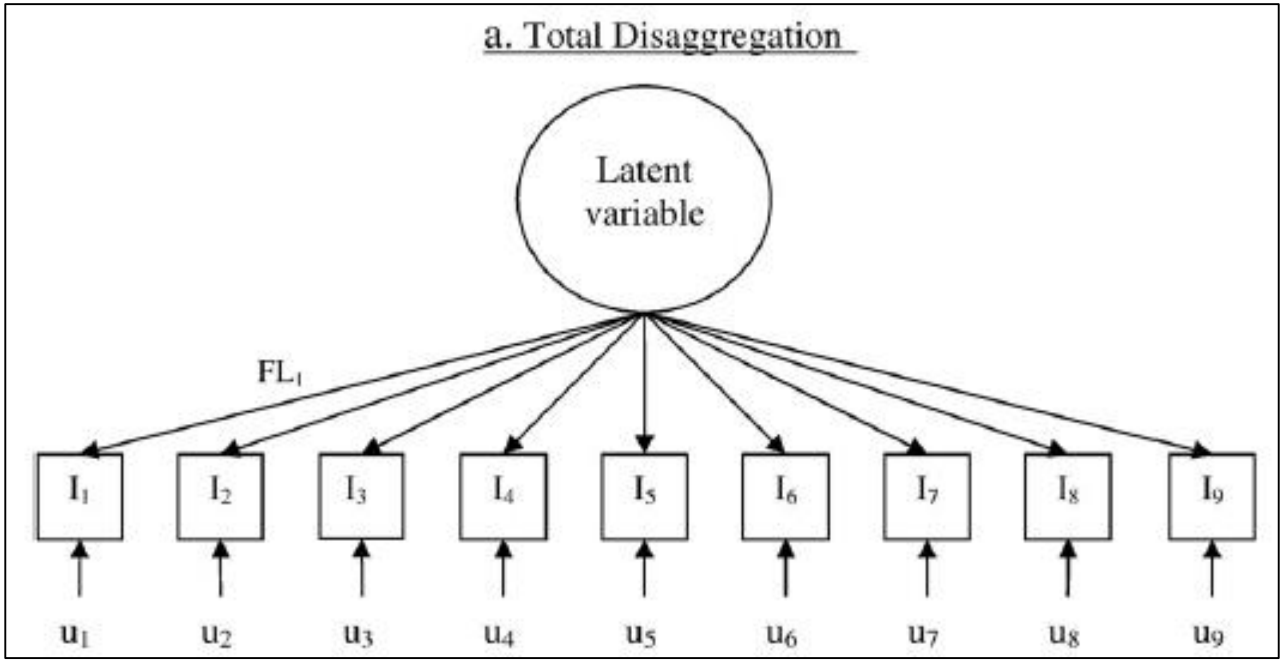


Figure 1.2 (a) Measurement model 1. (b) Measurement model 2. (c) Measurement model 3.

(Wang & Wang, 2012)

Model Formulation

Types of Measurement Models

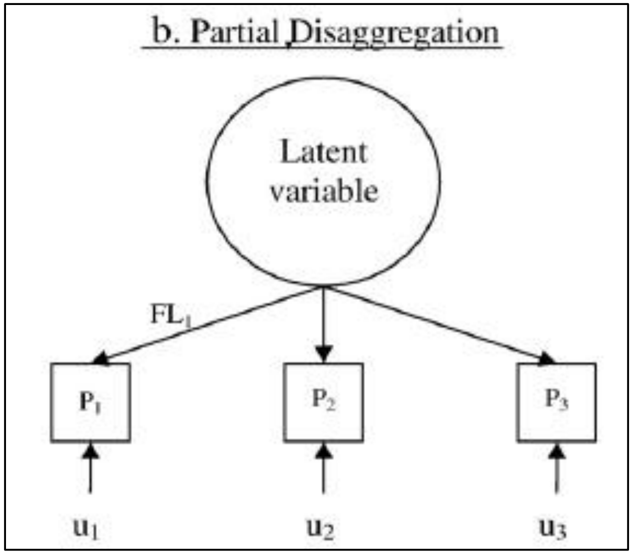


- *Single items*

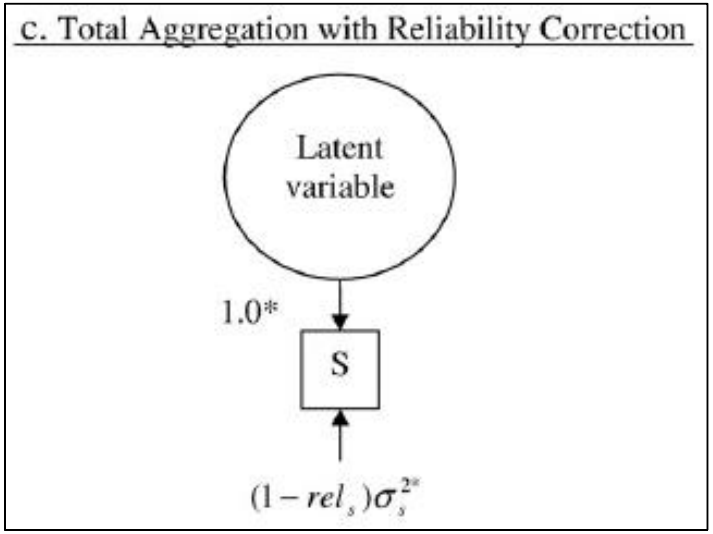
(Williams & O'Boyle, 2008, p. 234)

Model Formulation

Types of Measurement Models



- *Parcel approach*
- *Domain representative approach*



- *Composite score*

(Williams & O’Boyle, 2008, p. 234)

Model Formulation

Structural Model

The term "structural" stands for the assumption that the parameters are not just descriptive measures of association but rather that they reveal an invariant "causal" relation.
(Bollen, 1989, p. 4)

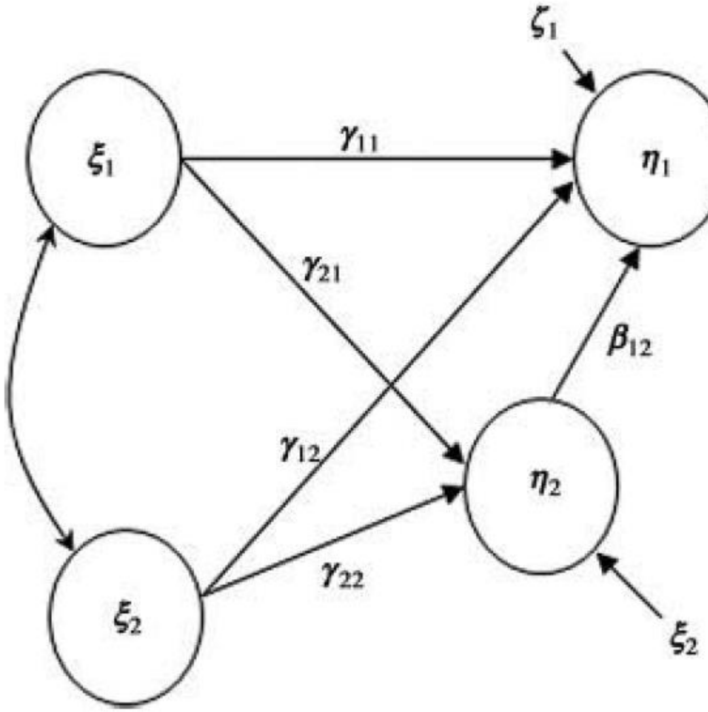


Figure 1.3 Structural model.

(Wang & Wang, 2012)

Model Formulation

Variables

Variable	Definition	Dimension
η (eta)	Latent endogenous variable	$m \times 1$
ξ (xi)	Latent exogenous variable	$n \times 1$
ζ (zeta)	Residual term in equations	$m \times 1$
y	Endogenous indicators	$p \times 1$
x	Exogenous indicators	$q \times 1$
ε (epsilon)	Measurement errors of y	$p \times 1$
δ (delta)	Measurement errors of x	$q \times 1$

exogenous = independent
endogenous = dependent (but can also be predictors!!)

n = latent exogenous variables (ξ)
m = latent endogenous variables (η)
q = exogenous indicators (x)
p = endogenous indicators (y)

(Wang & Wang, 2012)

Model Formulation

Matrices

Matrix	Definition	Dimension
Coefficient matrices		
Λ_y (lambda y)	Factor loadings relating y to η	$p \times m$
Λ_x (lambda x)	Factor loadings relating x to ξ	$q \times n$
$\boldsymbol{B}$ (beta)	Coefficient matrix relating η to η	$m \times m$
Γ (gamma)	Coefficient matrix relating ξ to η	$m \times n$
Variance/covariance matrices		
Φ (phi)	Variance/covariance matrices of ξ	$n \times n$
Ψ (psi)	Variance/covariance matrices of ζ	$m \times m$
Θ_ε (theta-epsilon)	Variance/covariance matrices of ε	$p \times p$
Θ_δ (theta-delta)	Variance/covariance matrices of δ	$q \times q$

(Wang & Wang, 2012)

Model Formulation

Equations

Structural Model $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$

Measurement Models $\mathbf{Y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}$

$\mathbf{X} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}$

(Wang & Wang, 2012)

Model Formulation

Equations

Structural Model $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$

Measurement Models $\mathbf{Y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}$

$\mathbf{X} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}$

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

$$\eta_{1i} = \beta_{12}\,\eta_{2i} + \gamma_{11}\,\xi_{1i} + \gamma_{12}\,\xi_{2i} + \zeta_{1i}$$

$$\eta_{2i} = \gamma_{21}\,\xi_{1i} + \gamma_{22}\,\xi_{2i} + \zeta_{2i}$$

(Wang & Wang, 2012)

Model Formulation

Equations

Structural Model $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$

Measurement Models $\mathbf{Y} = \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$

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

$$\boxed{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{y21} & 0 \\ \lambda_{y31} & 0 \\ 0 & \lambda_{y42} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}}$$

$$y_{1i} = \eta_{1i} + \varepsilon_{1i}$$
$$y_{2i} = \lambda_{y21}\,\eta_{1i} + \varepsilon_{2i}$$
$$y_{3i} = \lambda_{y31}\,\eta_{1i} + \varepsilon_{3i}$$
$$y_{4i} = \lambda_{y42}\,\eta_{2i} + \varepsilon_{4i}$$

(Wang & Wang, 2012)

Model Formulation

Equations

Structural Model $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$

Measurement Models $\mathbf{Y} = \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$

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

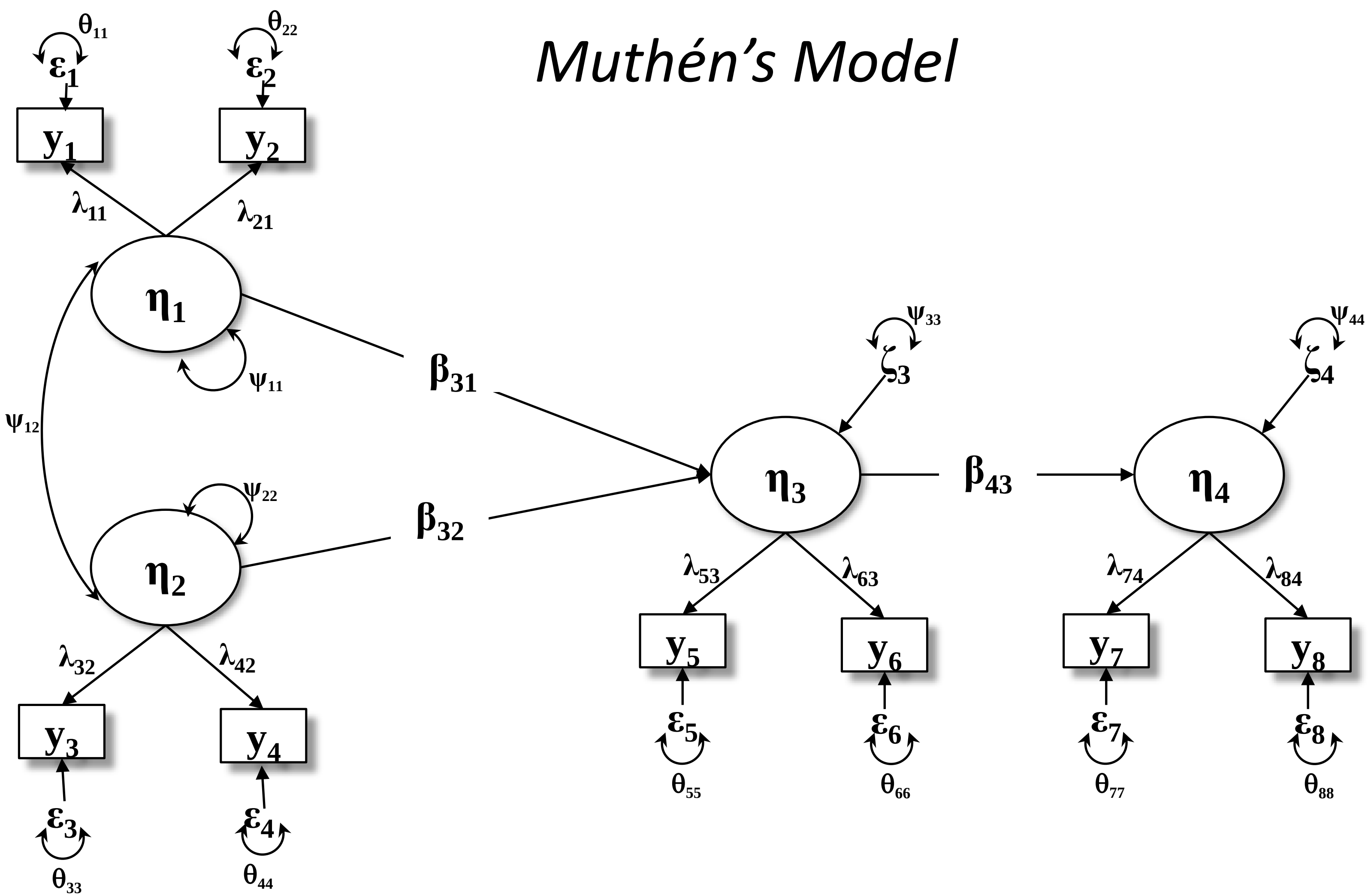
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{x21} & 0 \\ 0 & 1 \\ 0 & \lambda_{x42} \\ 0 & \lambda_{x52} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix}$$

$$x_{1i} = \xi_{1i} + \delta_{1i}$$
$$x_{2i} = \lambda_{x21}\,\xi_{1i} + \delta_{2i}$$
$$x_{3i} = \xi_{1i} + \delta_{3i}$$
$$x_{4i} = \lambda_{x42}\,\xi_{2i} + \delta_{4i}$$
$$x_{5i} = \lambda_{y52}\,\xi_{2i} + \delta_{5i}$$

(Wang & Wang, 2012)

Model Formulation

Muthén's Model



Model Formulation

Muthén's Model

Structural Model	$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} \cancel{+ \boldsymbol{\Gamma}\boldsymbol{\xi}} + \boldsymbol{\zeta}$
Measurement Models	$\mathbf{Y} = \boldsymbol{\Lambda}_{y}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$
	$\cancel{\mathbf{X} = \boldsymbol{\Lambda}_{x}\boldsymbol{\xi} + \boldsymbol{\delta}}$

Structural Model	$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta}$
Measurement Model	$\mathbf{Y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$

By adding intercepts/means

Structural Model	$\boldsymbol{\eta} = \boldsymbol{\nu} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta}$
Measurement Models	$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$

Model Formulation

Exercise

- Draw a SEM or find it in a publication
- Specify which are the measurement models
- Specify the “type/s” of measurement model (i.e., total disaggregation, partial disaggregation, total aggregation with reliability correction)
- Specify what is the structural model

Model Identification

Steps

1. ***Model formulation*** Specifying the SEM model that the researcher wants to test. The model may be formulated on the basis of theory or empirical findings. A general SEM model is composed of two parts: the measurement model and the structural model.
2. ***Model identification*** It determines whether there is a unique solution for all the free parameters in the specified model.
3. ***Model estimation*** Estimate model parameters and generate fitting function. Various estimation methods are available for SEM (e.g., ML, WLS, Bayes, ecc.).
4. ***Model evaluation*** Assess whether the model fits the data.
5. ***Model modification*** If the model does not fit the data, re-specification or modification of the model is needed. In this instance, the researcher makes a decision regarding how to delete, add, or modify parameters in the model.

(Wang & Wang, 2012, p. 2)

Model Identification

Parameters

Model identification concerns whether a unique value for each and every unknown parameter can be estimated from the observed data (Wang & Wang, 2012, p. 11).

Parameters can be:

1. FIXED: forced to assume a specific value, such as 0 or 1
2. CONSTRAINED: to be equal to another parameter
3. FREE: estimated by the model

Model Identification

Degrees of Freedom

$$df = [p(p + 1)]/2 - \text{NFP} > 0$$

p = number of observed variables, NFP = Number of free parameters

	1.	2.	3.	4.	5.	6.	7.	8.
1. y_1	$Var(y_1)$							
2. y_2	$Cov(y_1\ y_2)$	$Var(y_2)$						
3. y_3	$Cov(y_1\ y_3)$	$Cov(y_2\ y_3)$	$Var(y_3)$					
4. y_4	$Cov(y_1\ y_4)$	$Cov(y_2\ y_4)$	$Cov(y_3\ y_4)$	$Var(y_4)$				
5. y_5	$Cov(y_1\ y_5)$	$Cov(y_2\ y_5)$	$Cov(y_3\ y_5)$	$Cov(y_4\ y_5)$	$Var(y_5)$			
6. y_6	$Cov(y_1\ y_6)$	$Cov(y_2\ y_6)$	$Cov(y_3\ y_6)$	$Cov(y_4\ y_6)$	$Cov(y_5\ y_6)$	$Var(y_6)$		
7. y_7	$Cov(y_1\ y_7)$	$Cov(y_2\ y_7)$	$Cov(y_3\ y_7)$	$Cov(y_4\ y_7)$	$Cov(y_5\ y_7)$	$Cov(y_6\ y_7)$	$Var(y_7)$	
8. y_8	$Cov(y_1\ y_8)$	$Cov(y_2\ y_8)$	$Cov(y_3\ y_8)$	$Cov(y_4\ y_8)$	$Cov(y_5\ y_8)$	$Cov(y_6\ y_8)$	$Cov(y_7\ y_8)$	$Var(y_8)$

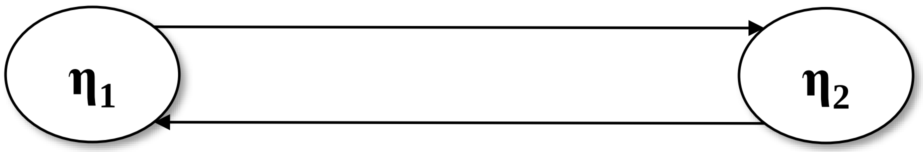
e.g. with 8 observed variables (p) we have (8*9)/2 = 36 non-redundant elements. Thus, NFP could be up to 35.

Model Identification

Other assumptions

There are a number of other assumptioms for ensuring identification (see Bollen, 1989), but the most important is that the model must be recursive.

An example of nonrecursive model is a model with loops (Rigdon, 1995):



Model Identification

Exercise

On your previous SEM (or another one)

- Compute the maximum number of possible free parameters
- Identify some fixed and free parameters (and constrained, if there are)

Model Estimation

Steps

1. ***Model formulation*** Specifying the SEM model that the researcher wants to test. The model may be formulated on the basis of theory or empirical findings. A general SEM model is composed of two parts: the measurement model and the structural model.
2. ***Model identification*** It determines whether there is a unique solution for all the free parameters in the specified model.
3. ***Model estimation*** Estimate model parameters and generate fitting function. Various estimation methods are available for SEM (e.g., ML, WLS, Bayes, ecc.).
4. ***Model evaluation*** Assess whether the model fits the data.
5. ***Model modification*** If the model does not fit the data, re-specification or modification of the model is needed. In this instance, the researcher makes a decision regarding how to delete, add, or modify parameters in the model.

(Wang & Wang, 2012, p. 2)

Model Estimation

Estimation of SEM models is different from that of multiple regressions.

Instead of minimizing the discrepancies between the fitted and observed values of the response variable [i.e., $\Sigma(y - \hat{y})$], SEM estimation procedures minimize the residuals that are differences between the **sample variances/covariances** and the **variances/covariances** estimated from the model. (Wang & Wang, 2012, p. 14).

$$\Sigma = \Sigma(\theta)$$

Given that the **<u>population</u> variances/covariances matrix (Σ)** is unkown, it is replaced by the **<u>sample</u> variances/covariances matrix (S)**, so that the null hypothesis is

$$S = \Sigma(\theta)$$

Model Estimation

There are several estimators, but the most common procedure is the use of **Maximum Likelihood function**

$$F_{\mathrm{ML}} = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \mathrm{tr}\left(\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right) - \log|\mathbf{S}| - (p + q) \qquad (4.67)$$

(Bollen, 1989, p. 107)

$(N - 1)F_{\mathrm{ML}}$ is chi-squared distributed with $df = \{[p(p + 1)]/2 - \mathrm{NFP}\}$, where N is the sample size and F_{ML} is the value of the fitting function evaluated at the final estimates (Bollen, 1989, p. 110).

Recall that the null hypothesis of this chi-square test is H_0: S = Σ(θ).
Thus, a good model should show a **non-significant chi-square test**, because we expect **non-significant differeces** between S and Σ(θ)

Model Estimation

$$\log L_0 = -\frac{N-1}{2}\left\{\log|\hat{\boldsymbol{\Sigma}}| + \operatorname{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S})\right\} \tag{7.43}$$

$$\begin{aligned}\log L_1 &= -\frac{(N-1)}{2}\left\{\log|\mathbf{S}| + \operatorname{tr}(\mathbf{S}^{-1}\mathbf{S})\right\} \\ &= -\frac{(N-1)}{2}\left\{\log|\mathbf{S}| + q\right\}\end{aligned} \tag{7.47}$$

$$-2\log\left(\frac{L_0}{L_1}\right) = -2\log L_0 + 2\log L_1$$

Bollen (1989, p. 263-265)

Model Estimation

Example from an Mplus Output

```
THE MODEL ESTIMATION TERMINATED NORMALLY



MODEL FIT INFORMATION

Number of Free Parameters                        116

Loglikelihood

          H0 Value                         -5249.015
          H1 Value                         -4935.988

Information Criteria

          Akaike (AIC)                     10730.030
          Bayesian (BIC)                   11167.155
          Sample-Size Adjusted BIC         10799.223
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                              626.054
          Degrees of Freedom                     397
          P-Value                             0.0000
```

Log-Likelihood for my specified model (log L_0)

Log-Likelihood for the saturated model (log L_1), namely for the model in which $S = \Sigma(\theta)$

```
> (-2*-5249.015)+(2*(-4935.988))
[1] 626.054
```

Model from Perinelli, Alessandri, Cepale, & Fraccaroli (2022)

Model Estimation

Exercise

Compute both chi-square statistics and degrees of freedom using the following information. **Pay attention!** Add the vector of means (for simplicity, again p) to the computation of degrees of freedom, that is $\{ \{[p*(p + 1)]/2\} + p \}$ – NFP

```
Number of observed variables                       9
Number of Free Parameters                         41

Loglikelihood

          H0 Value                         -1406.522
          H1 Value                         -1397.180
```

Source: Model 3 from Avanzi, Perinelli et al. (2021, *Anxiety, Stress, & Coping*)

Model Estimation

Exercise

Compute both chi-square statistics and degrees of freedom using the following information. **Pay attention!** Add the vector of means (for simplicity, again p) to the computation of degrees of freedom, that is $\{ \{[p*(p + 1)]/2\} + p \}$ – NFP

```
Number of observed variables                        9
Number of Free Parameters                          41

Loglikelihood

          H0 Value                          -1406.522
          H1 Value                          -1397.180
```

Solution

```
Chi-square = (-2*-1406.522)+(2*-1397.180) = 18.684
df = (((9*10)/2)+9)-41 = 13
```

Source: Model 3 from Avanzi, Perinelli et al. (2021, *Anxiety, Stress, & Coping*)

Model Evaluation

Steps

1. ***Model formulation*** Specifying the SEM model that the researcher wants to test. The model may be formulated on the basis of theory or empirical findings. A general SEM model is composed of two parts: the measurement model and the structural model.
2. ***Model identification*** It determines whether there is a unique solution for all the free parameters in the specified model.
3. ***Model estimation*** Estimate model parameters and generate fitting function. Various estimation methods are available for SEM (e.g., ML, WLS, Bayes, ecc.).
4. ***Model evaluation*** Assess whether the model fits the data.
5. ***Model modification*** If the model does not fit the data, re-specification or modification of the model is needed. In this instance, the researcher makes a decision regarding how to delete, add, or modify parameters in the model.

(Wang & Wang, 2012, p. 2)

Model Evaluation

χ^2 test is highly sensitive to sample size and to the number of variables included in the model.

"As such, the significance of the χ^2 test should not be a reason by itself to reject a model" (Wang & Wang, 2012, p. 18).

Model Evaluation

Other Common Indices of Fit

$$CFI = \frac{d_{null} - d_{specified}}{d_{null}}$$ > .90-.95

d = noncentrality parameter = $\chi^2 - df$
null = the worst model (all cov. are zero)

$$TLI = \frac{\left(\frac{\chi^2_{null}}{df_{null}} - \frac{\chi^2_{specified}}{df_{specified}}\right)}{\left(\frac{\chi^2_{null}}{df_{null}} - 1\right)}$$ > .90-.95

$$RMSEA = \sqrt{\frac{\chi^2 - df}{(n-1)df}}$$ < .08 - .05

$$SRMR = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}\left(\frac{s_{ij} - \hat{\sigma}_{ij}}{s_{ii}s_{jj}}\right)^2}{p(p+1)/2}} = \sqrt{\frac{sum\ of\ squared\ std\ residuals}{size\ var{-}covar\ matrix}}$$ < .08 - .05

Model Evaluation

Nested models

Definition:

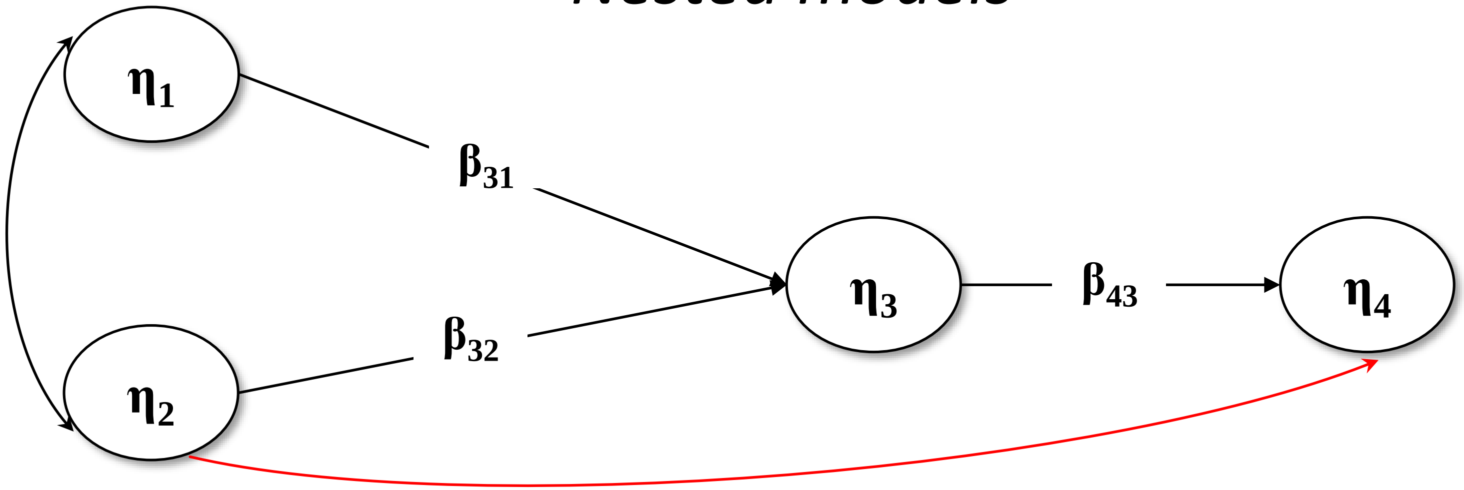
- Models "are *nested* [when] one model is produced by changing the status of one or more parameters in the other" (Hoyle, 2012, p. 11).
- "*Nested* models are those for which the parameter space associated with one model is a subset of the parameter space associated with the other model" (Merkle, You, & Preacher, 2016, p. 152-153)
- "Two models are nested if they are of the same form, but the free parameters of one model are a subset of the free parameters of the other model" (Hoyle, 2012, p. 141)

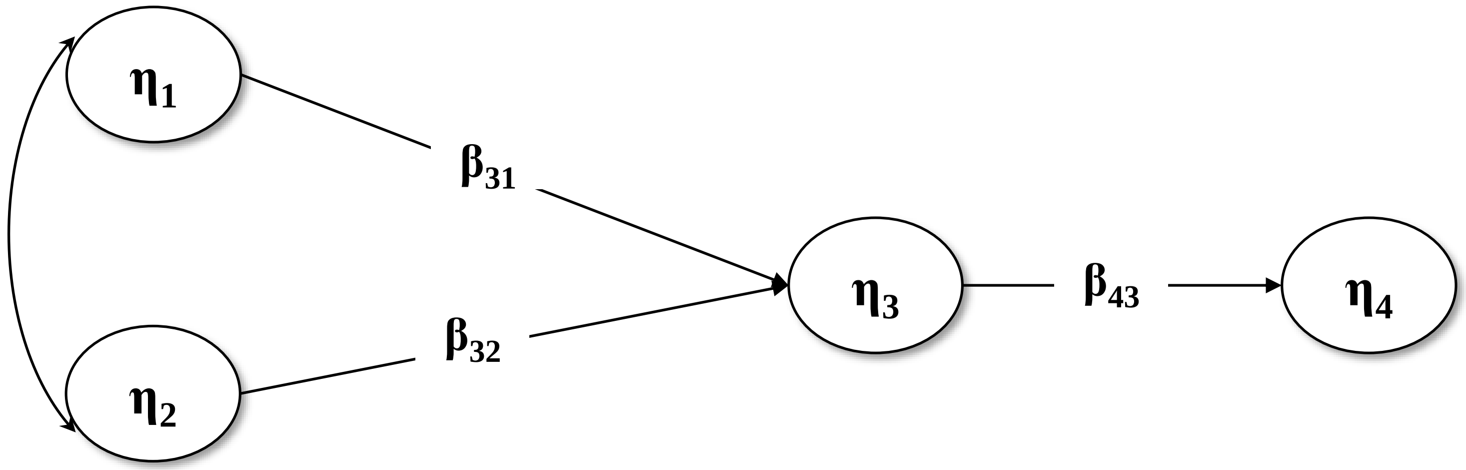
Nested models can be compared by means of $\Delta\chi^2$ statistics, distributed with Δdf.
The resulting *p*-value will attest

- whether removing 1 or more parameters significantly worsen (or not) the fit of the model
- whether adding 1 or more parameters significantly improve the fit of the model

Model Evaluation

Nested models





$\Delta df = 1$

Model Evaluation

Non-Nested models

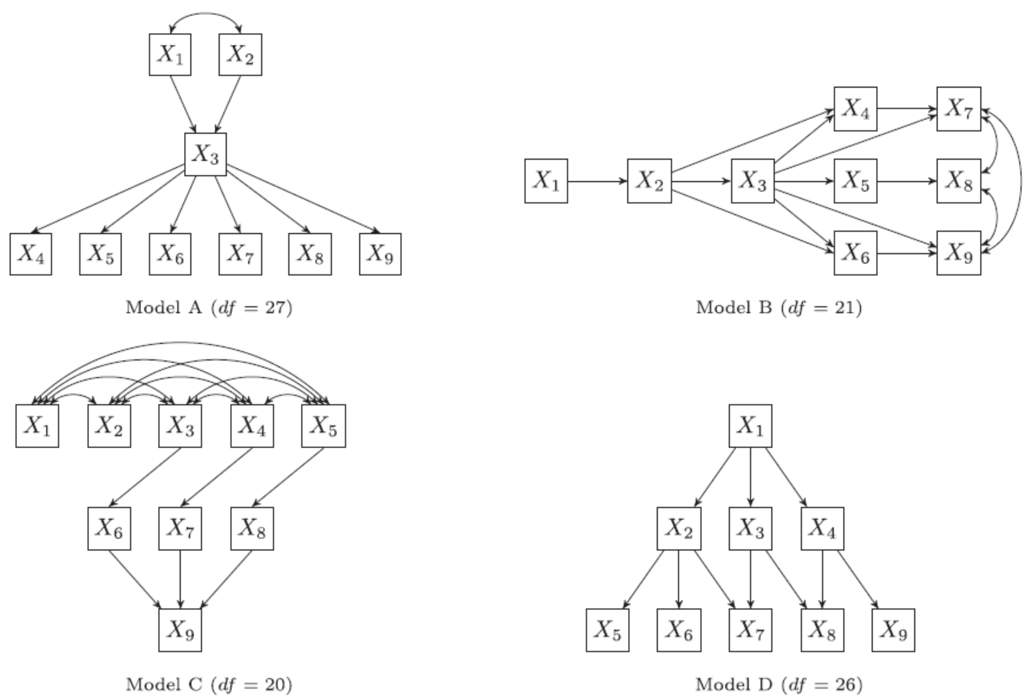


Figure 2. Path diagrams reflecting the models used in Simulation 2.

Model Evaluation

Information Criteria Indices

Useful for comparing **non-nested model**

$$\text{AIC} = -2\text{Ln}(L) + 2q$$

$$\text{BIC} = -2\text{Ln}(L) + q\text{Ln}(N)$$

$$\text{ABIC} = -2\text{Ln}(L) + q\text{Ln}[(N+2)/24)]$$

Ln(L) = log-likelihood value
q = number of free parameters
N = sample size

The model with the **smaller values** in these indices is preferred.

IMPORTANT: YOU CANNOT USE THESE INDICES WHEN THE log L_1 IS DIFFERENT ACROSS MODELS

Model Evaluation

Exercise: Fit and Nested Model Comparison

Table 2. Model fits of the predicted and alternative model tests.

	χ^2	df	p	CFI	TLI	RMSEA	SRMR	AIC	MC	$\Delta\chi^2$	Δdf	p
M1: No constraint	12.794	9	.172	0.991	0.967	0.066	0.022	2897.154				
M2: CS → OI fixed to be zero	13.819	11	.243	0.993	0.980	0.052	0.027	2894.179	M2 Vs M1	1.025	2	0.599
M3: PD → OI fixed to be zero	**18.684**	**13**	**.133**	**0.986**	**0.966**	**0.067**	**0.041**	**2895.044**	**M3 Vs M2**	**4.865**	**2**	**0.088**
M4 PD → CS fixed to be zero	29.094	15	.016	0.966	0.926	0.099	0.073	2901.454	M4 Vs M3	10.410	2	0.005

Note: CS = Colleague Support; OI = Organizational Identification, PD = Psychological Distress. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; MC = Model Comparison. M1 = no constraint were imposed; M2 = the paths from colleague support T1 to organizational identification T2 and from colleague support T2 to organizational identification T3 were fixed to be zero; M3 = the paths from psychological distress T1 to organizational identification T2 and from psychological distress T2 to organizational identification T3 were fixed to be zero; M4 = the paths from psychological distress T1 to colleague support T2 and from psychological distress T2 to colleague support T3 were fixed to be zero.

Avanzi, Perinelli et al. (2021)

Model Evaluation

Exercise: Fit and Non-Nested Model Comparison

688 J Psychopathol Behav Assess (2017) 39:680–692

Table 3 Indices of hypothesized models (see Fig. 2) and alternative models ($N = 109$)

Model	Free Parameters	χ^2(df)	CFI	TLI	RMSEA[90% CI]	SRMR	AIC
Hyp. mod.	21	10.10 (6)$^{n.s.}$	.974	.934	.080[.000, .163]	.037	3085.58
Alt. mod. 1	23	6.55 (4)$^{n.s.}$	.984	.939	.077[.000, .179]	.026	3086.03
Alt. mod. 2	22	8.98 (5)$^{n.s.}$	.975	.924	.086[.000, .176]	.038	3086.47
Alt. mod. 3	21	12.9 (6)*	.955	.888	.104[.017, .183]	.048	3088.45

See paragraph "Alternative Models" in Results section for the description of each alternative model

All the models had the same maximum log-likelihood value for the unrestricted (H1) model, thus they are comparable with AIC

Hyp. mod. hypothesized model (see Fig. 2), *Alt. mod.* alternative model, *df* degrees of freedom, *CFI* Comparative Fit Index, *TLI* Tucker-Lewis Index, *RMSEA* Root Mean Square Error of Approximation, *CI* Confidence interval, *SRMR* Standardized Root Mean Square Residual, *AIC* Akaike Information Criterion

$^{n.s.}$ $p > .10$, * $p < .05$

Caprara, Gerbino, Perinelli et al. (2017)

Model Modification

Steps

1. ***Model formulation*** Specifying the SEM model that the researcher wants to test. The model may be formulated on the basis of theory or empirical findings. A general SEM model is composed of two parts: the measurement model and the structural model.
2. ***Model identification*** It determines whether there is a unique solution for all the free parameters in the specified model.
3. ***Model estimation*** Estimate model parameters and generate fitting function. Various estimation methods are available for SEM (e.g., ML, WLS, Bayes, ecc.).
4. ***Model evaluation*** Assess whether the model fits the data.
5. ***Model modification*** If the model does not fit the data, re-specification or modification of the model is needed. In this instance, the researcher makes a decision regarding how to delete, add, or modify parameters in the model.

(Wang & Wang, 2012, p. 2)

Model Modification

“If the model does not fit the data, re-specification or modification of the model is needed. In this instance, the researcher makes a decision regarding how to delete, add, or modify parameters in the model” (Wang & Wang, 2012, p. 2).

Modification Indices (MI): Change in χ^2 (i.e., $\Delta\chi^2$) after freeing a parameter

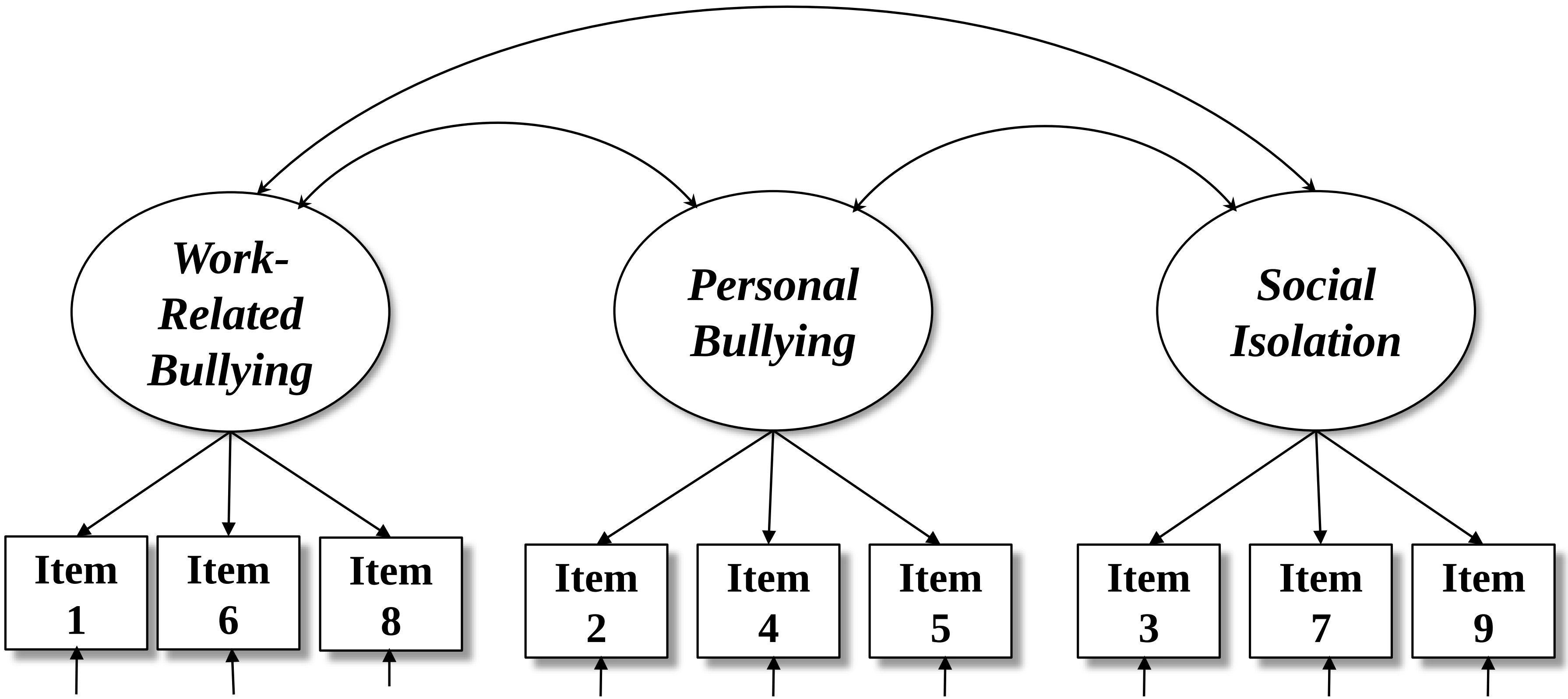
Usually, MI are provided in the output.

Use MI with caution: They must be **justified** from a theoretical and methodological point of view

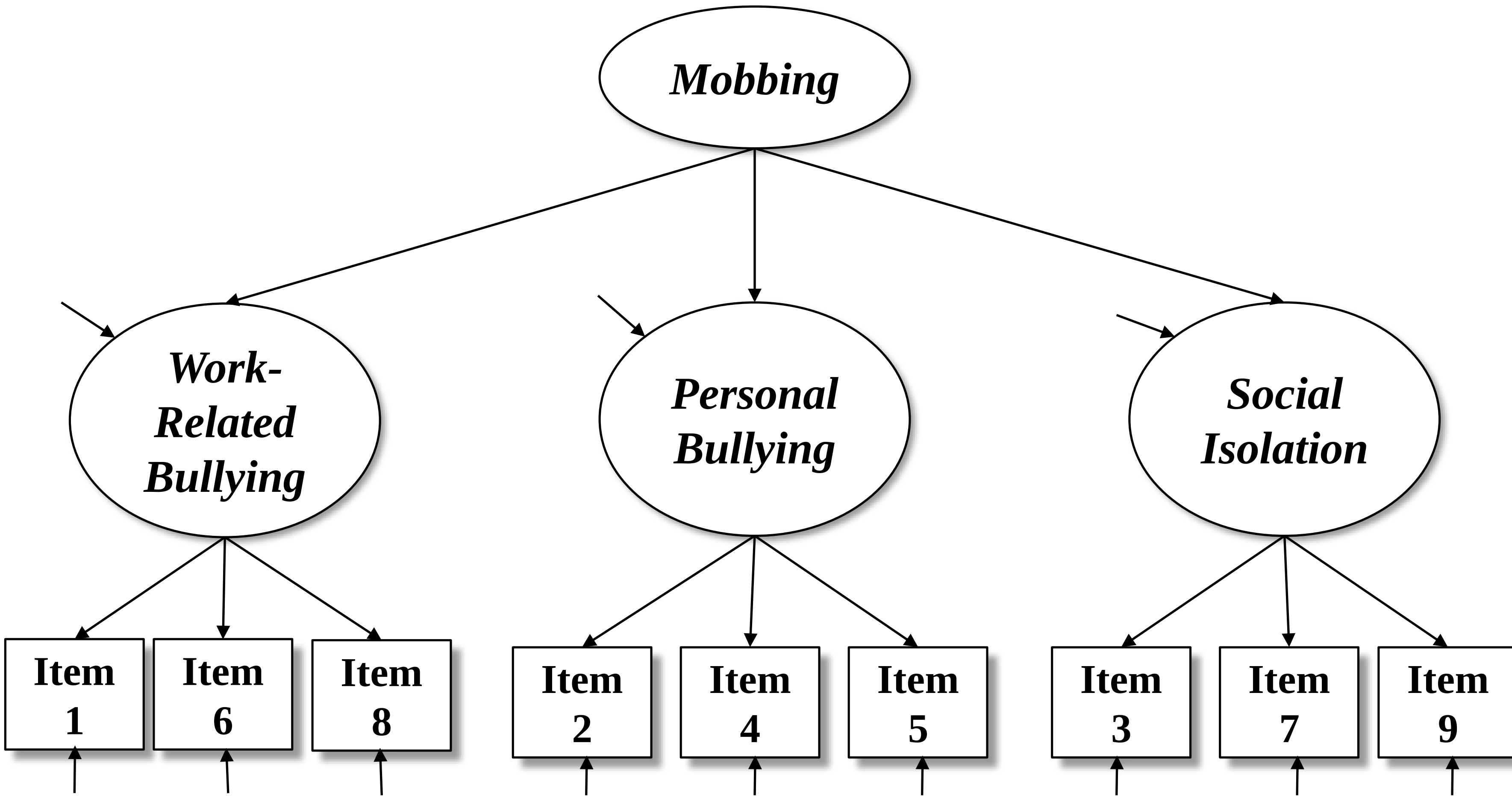
Structural Equation Modeling

Some Types and Applications

Confirmatory Factor Analysis (CFA)

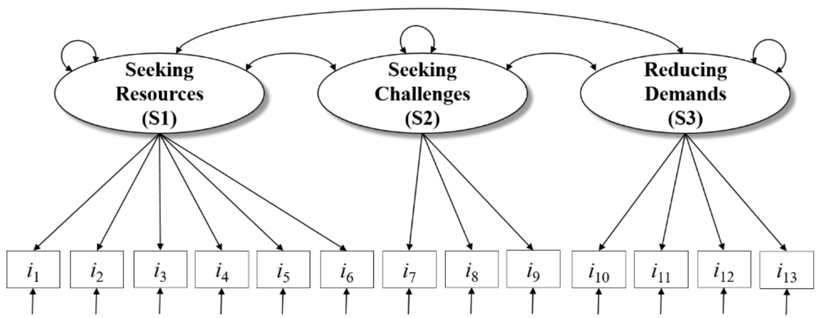


Higher-order (or Second-order) CFA

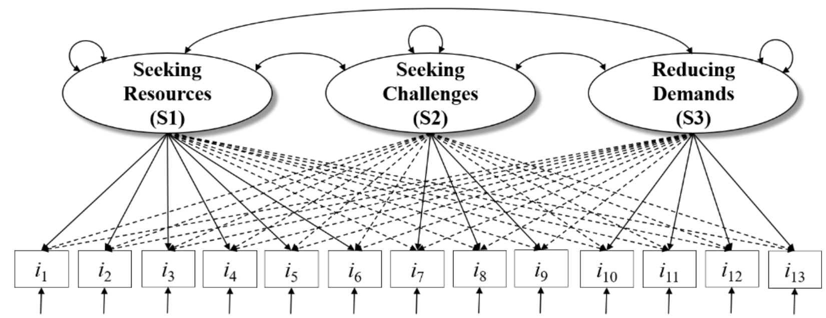


Multidimensional Structure

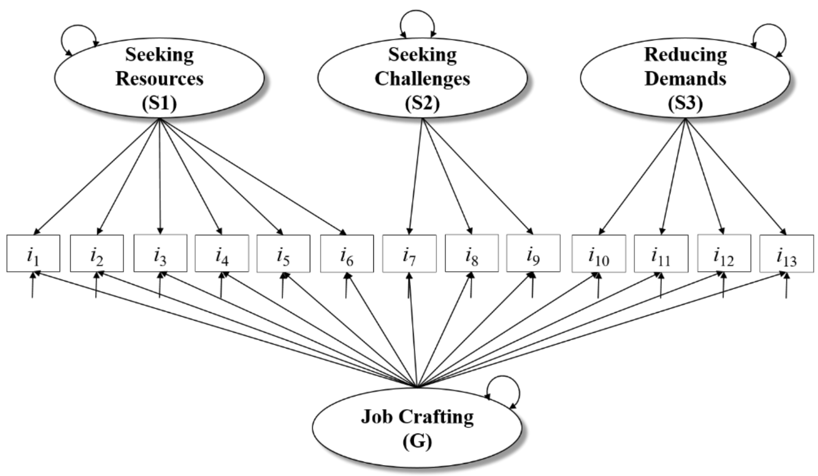
ICM-CFA



ESEM



B-CFA



B-ESEM

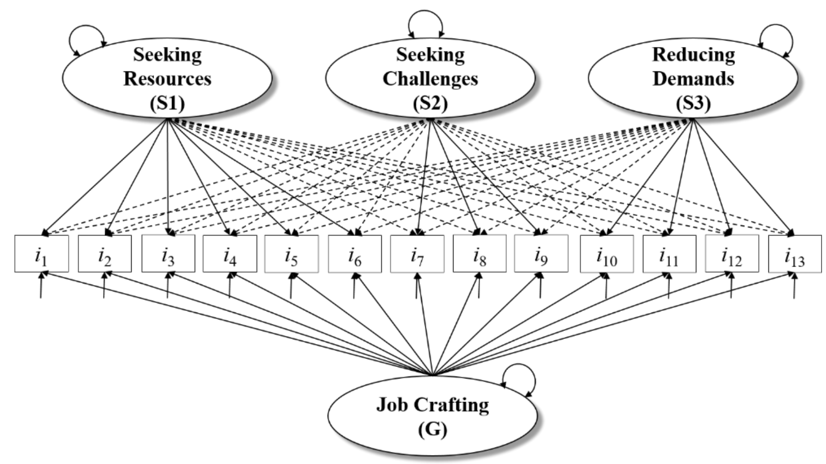
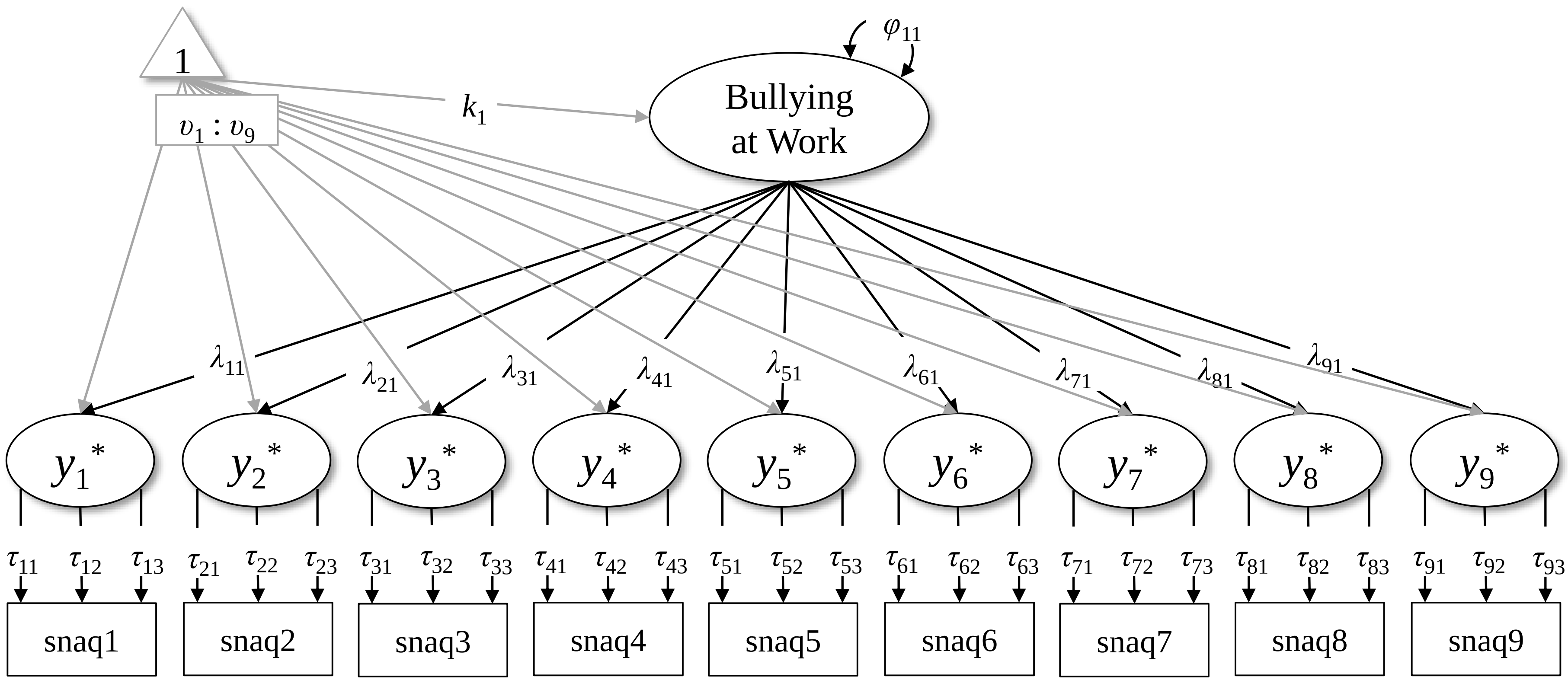


Figure from Vignoli, Perinelli, Demerouti, & Truxillo (2023, *Work, Aging and Retirement*)

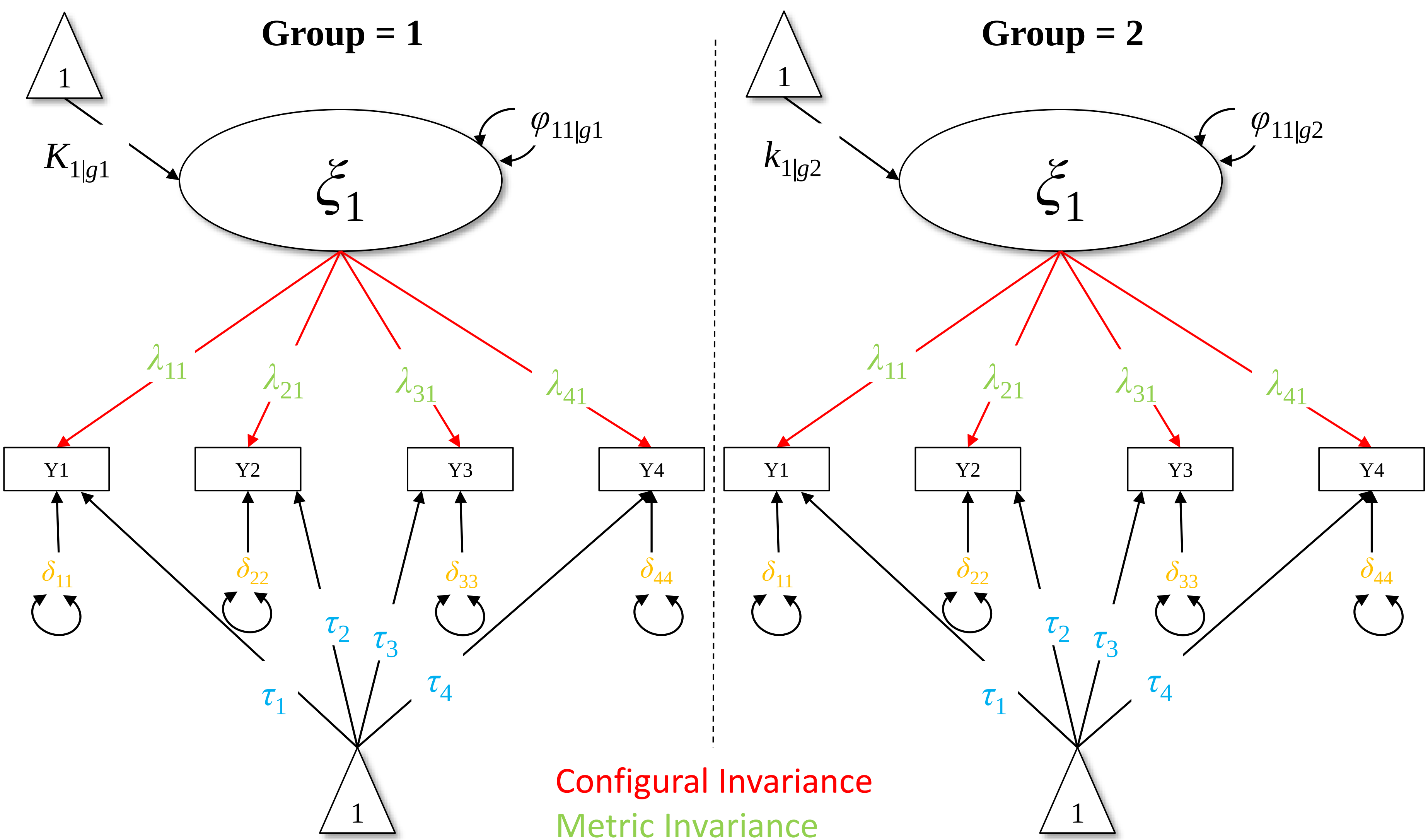
Morin, A. J., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, Exploratory Structural Equation Modeling (ESEM), and bifactor-ESEM. In G. Tenenbaum and R. C. Eklund (Eds.), *Handbook of sport psychology* (Vol. 1, 4th ed., pp. 1044-1073). John Wiley & Sons, Inc.

CFA with Ordinal Data



From Perinelli, Balducci, & Fraccaroli (2023, *Current Psychology*)
R codes are available here https://psicostat.dpss.psy.unipd.it/files/2022-06-17_perinelli.html , while R codes and *Mplus* syntaxes are available here https://doi.org/10.1007/s12144-022-03741-4 under “Supplementary Information”

Measurement Invariance (Multiple-group)



Configural Invariance

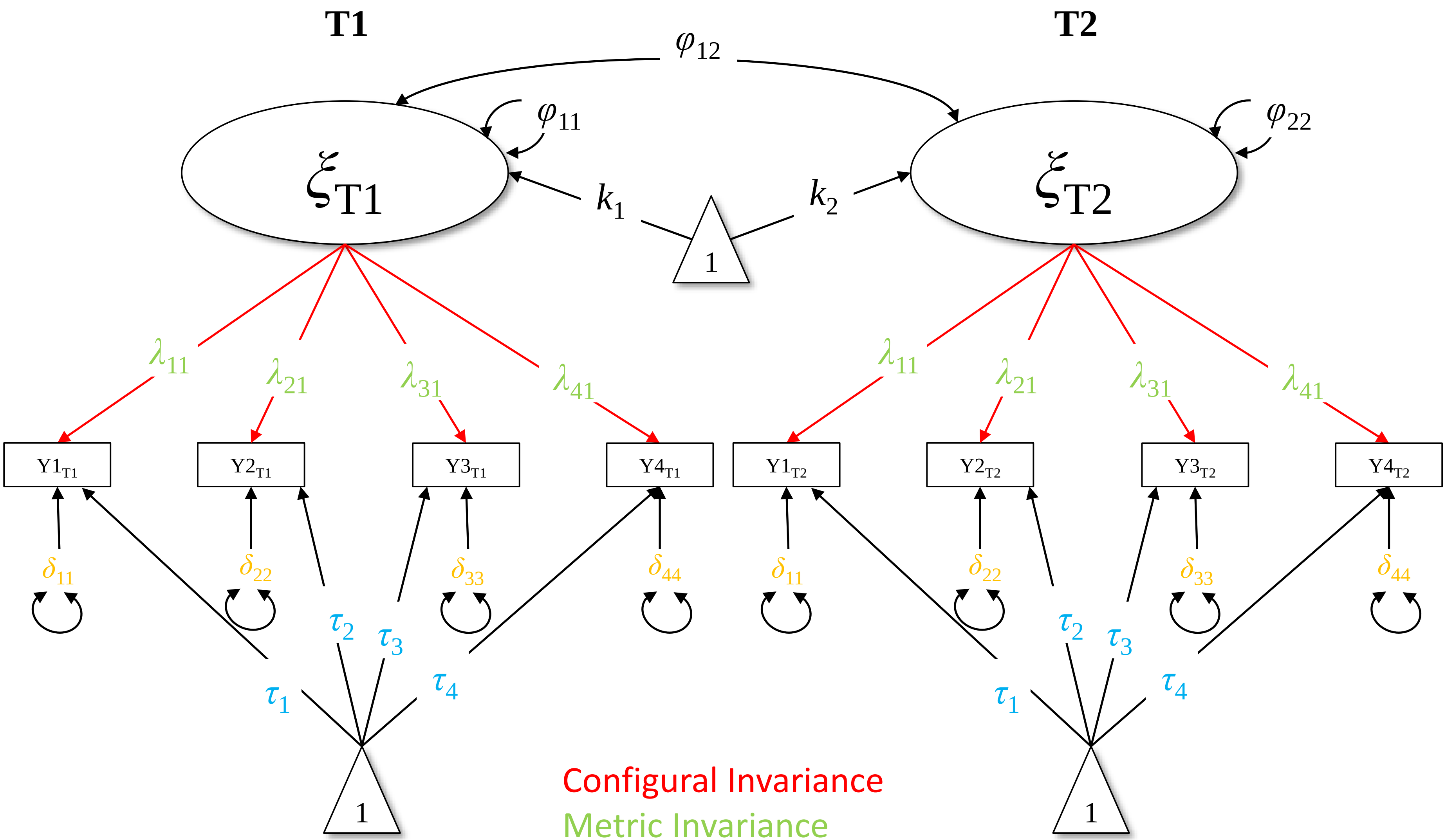
Metric Invariance

Scalar Invariance

Strict Invariance

Some material available at
https://github.com/EnricoPerinelli/Measurement-Invariance-PSICOSTAT

Measurement Invariance (Longitudinal)

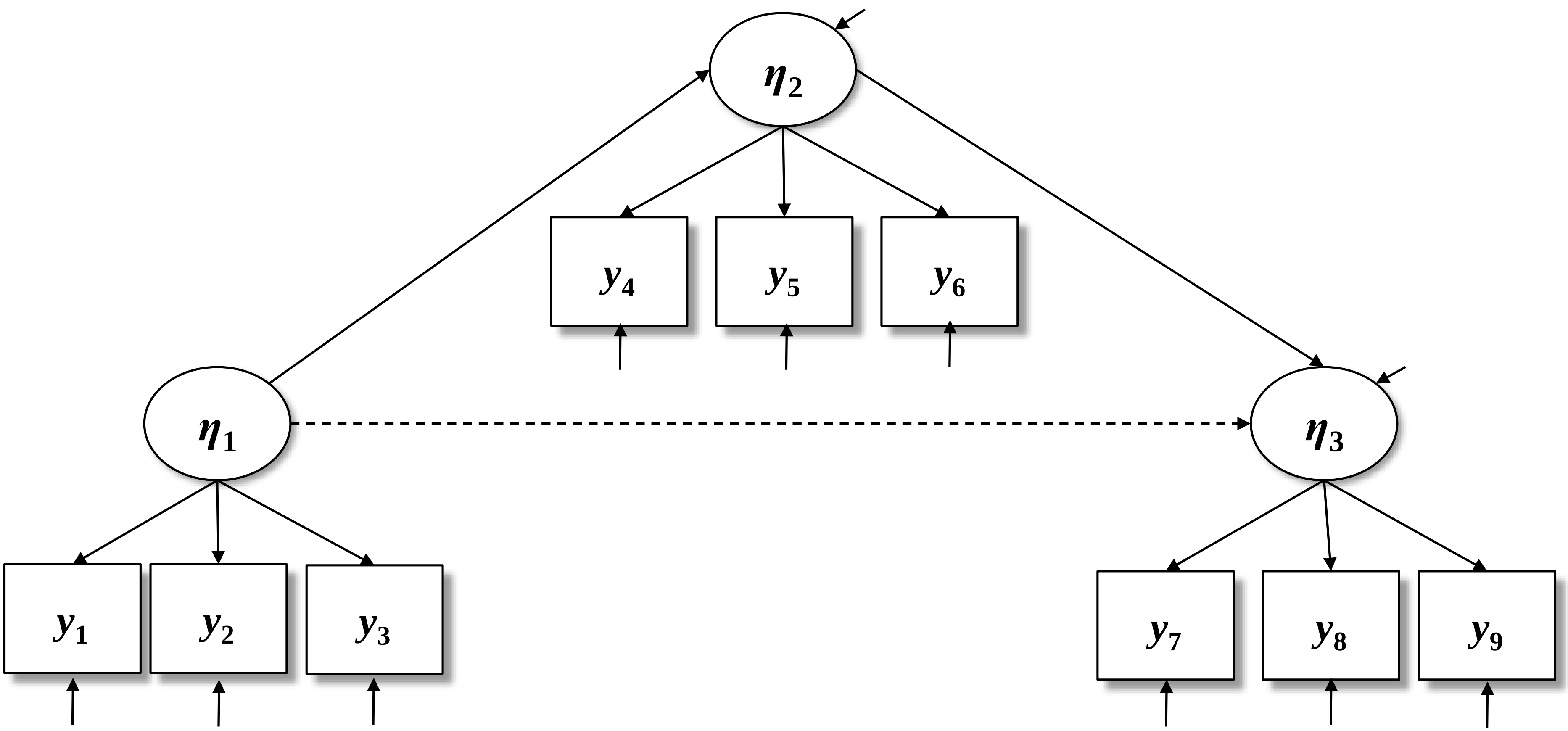


Note. Covariances between uniqueness are omitted for the sake of clarity

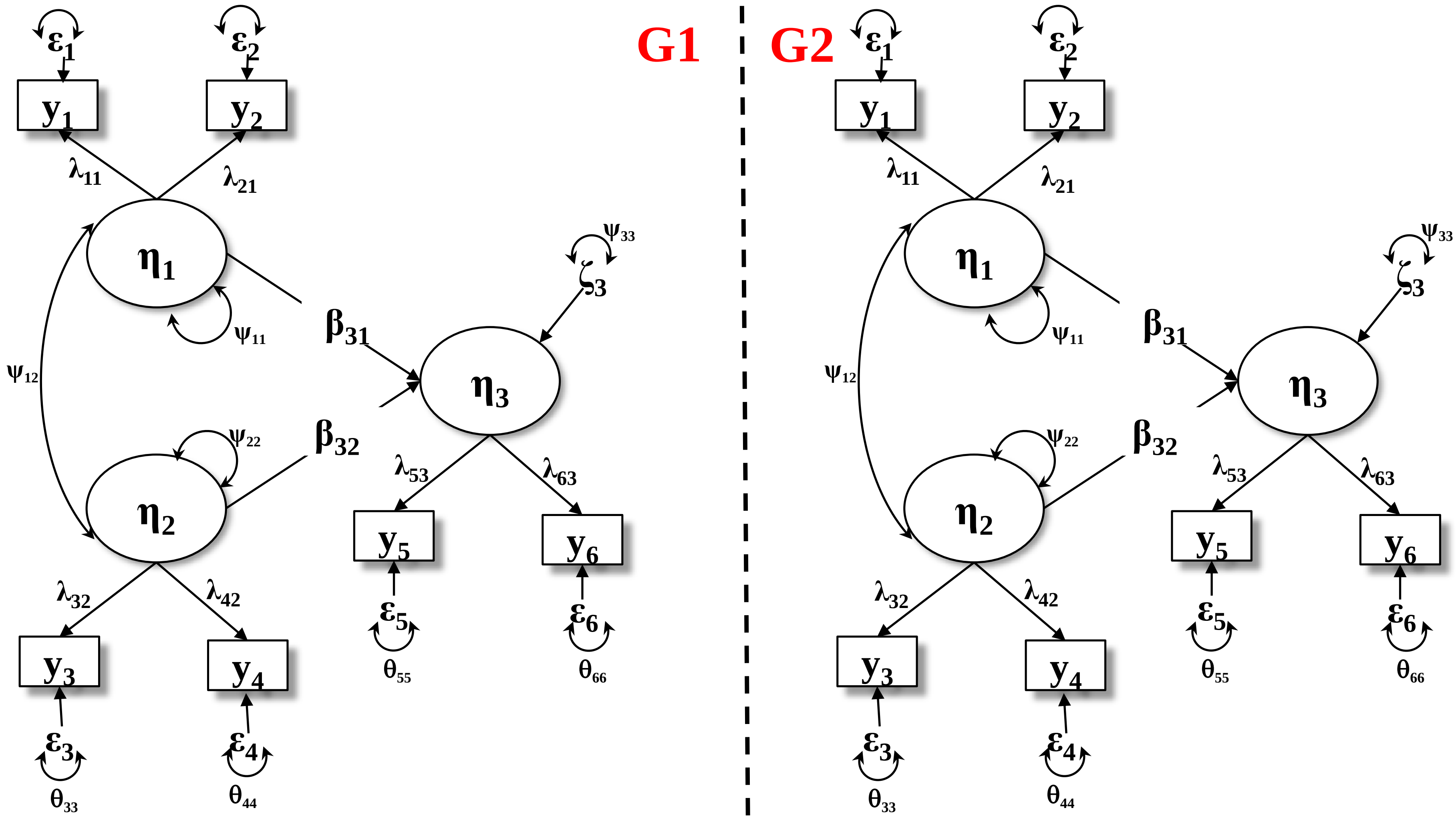
Configural Invariance
Metric Invariance
Scalar Invariance
Strict Invariance

Some material available at
https://github.com/EnricoPerinelli/Measurement-Invariance-PSICOSTAT

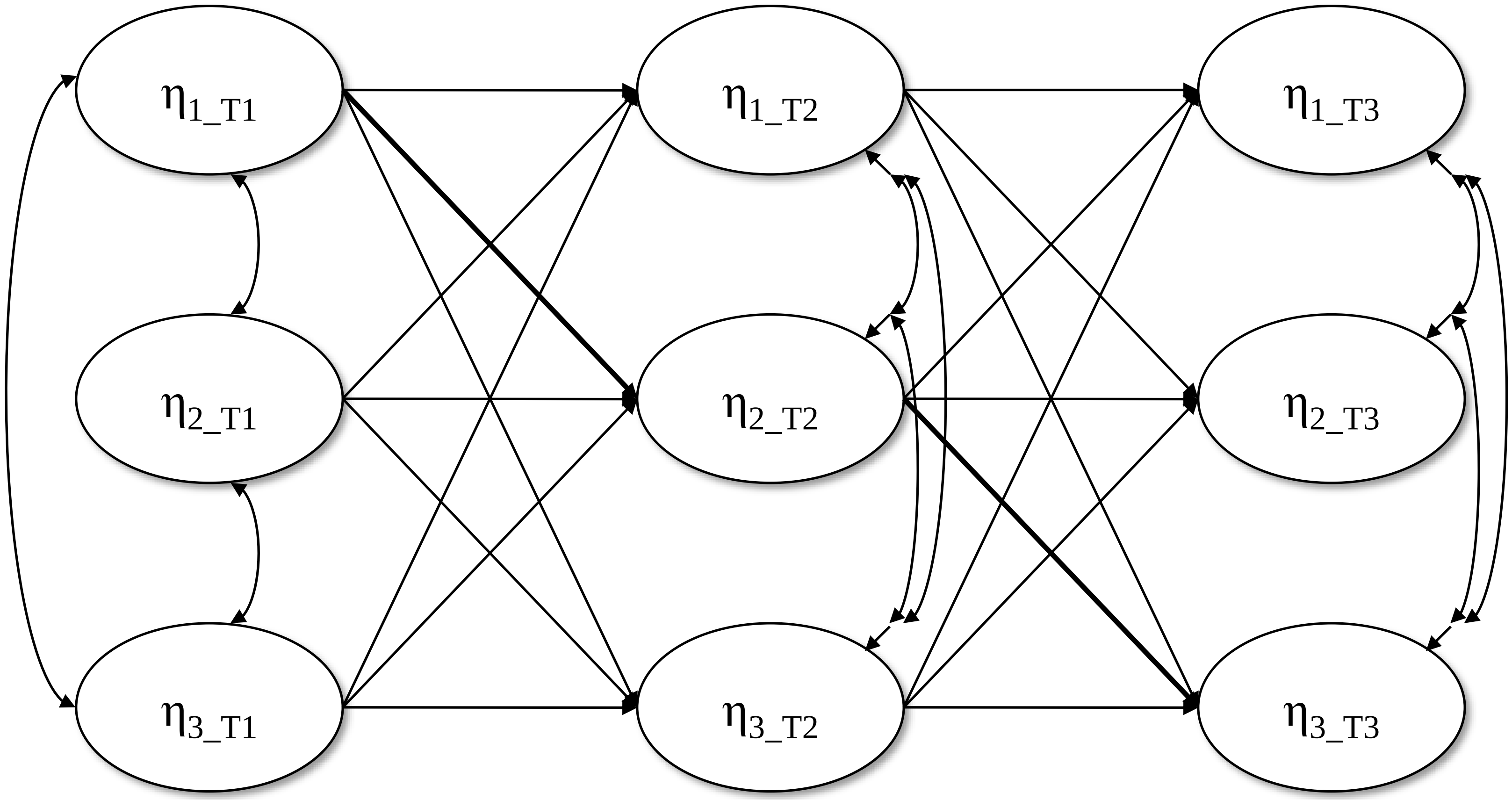
Cross-sectional Mediation with Latent Variables



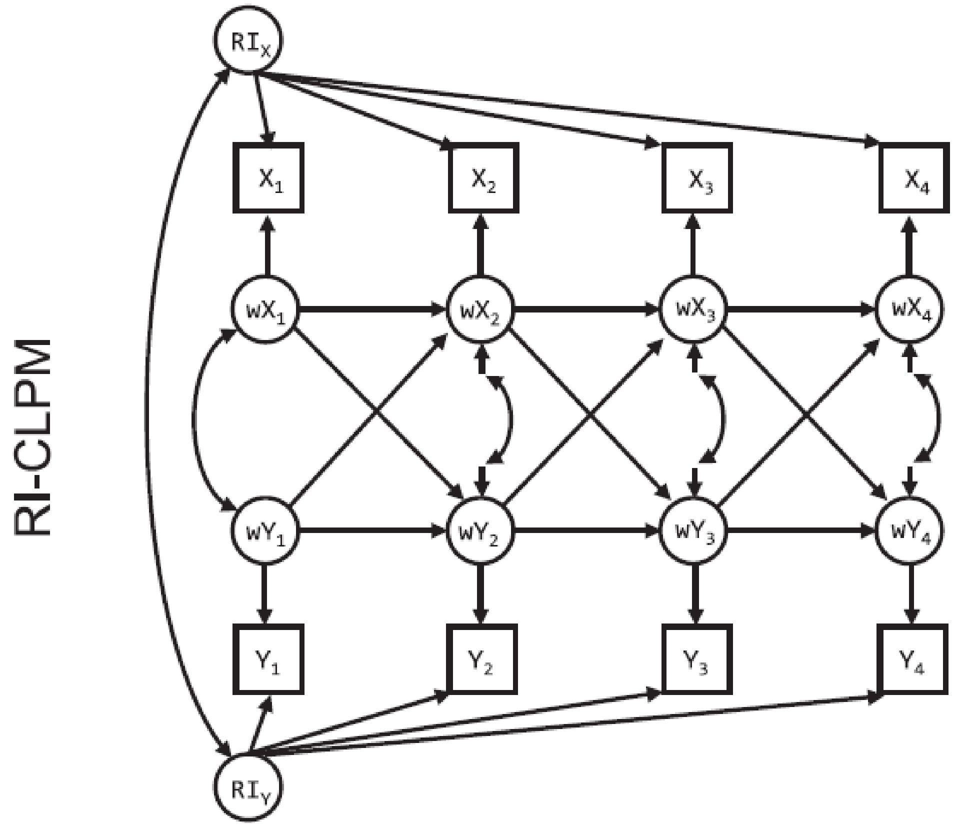
Multiple-group SEM (Jöreskog, *Psychometrika*, 1971)



Autoregressive Cross-Lagged Panel Model

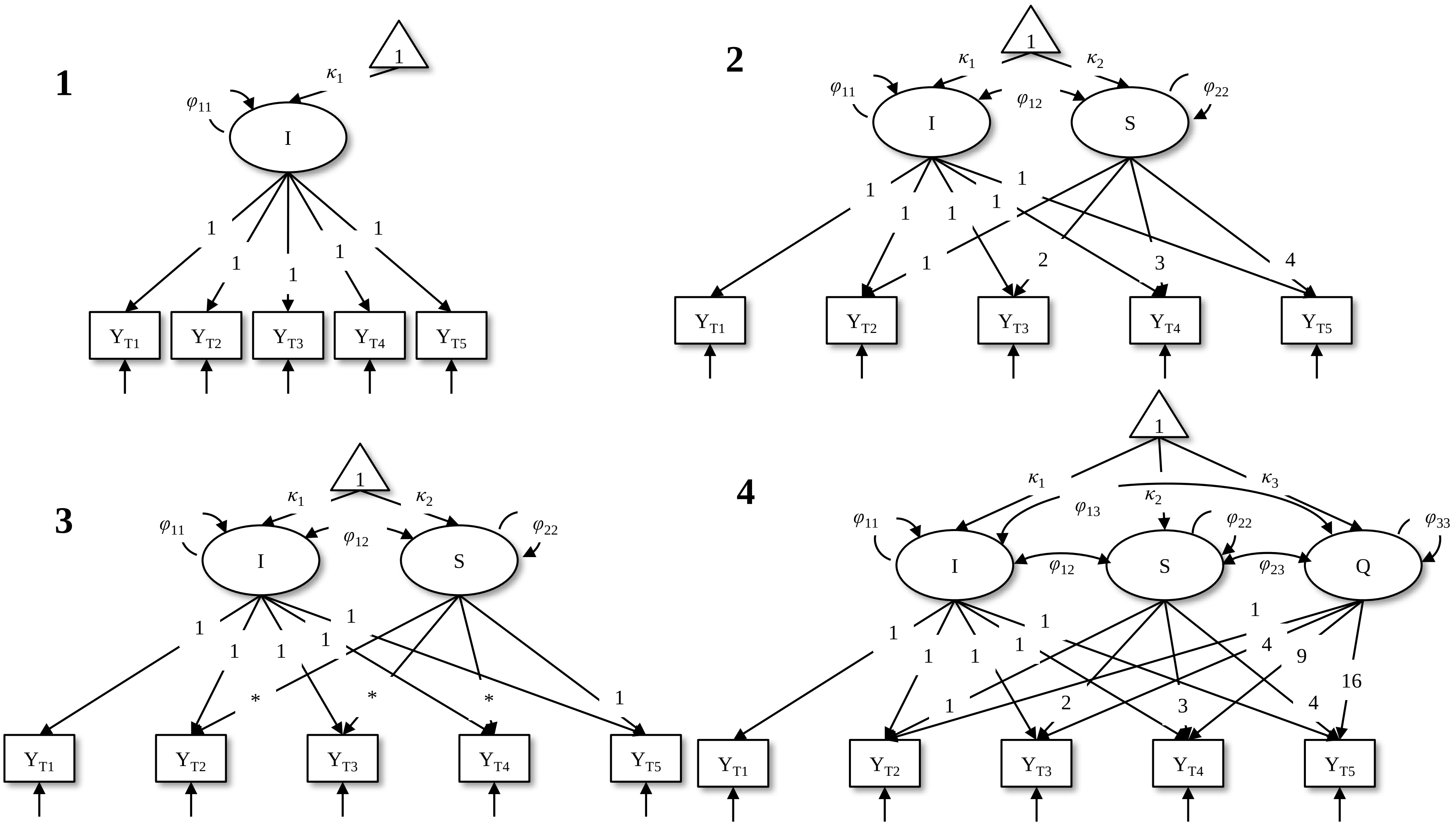


Random-Intercept CLPM (RI-CLPM)



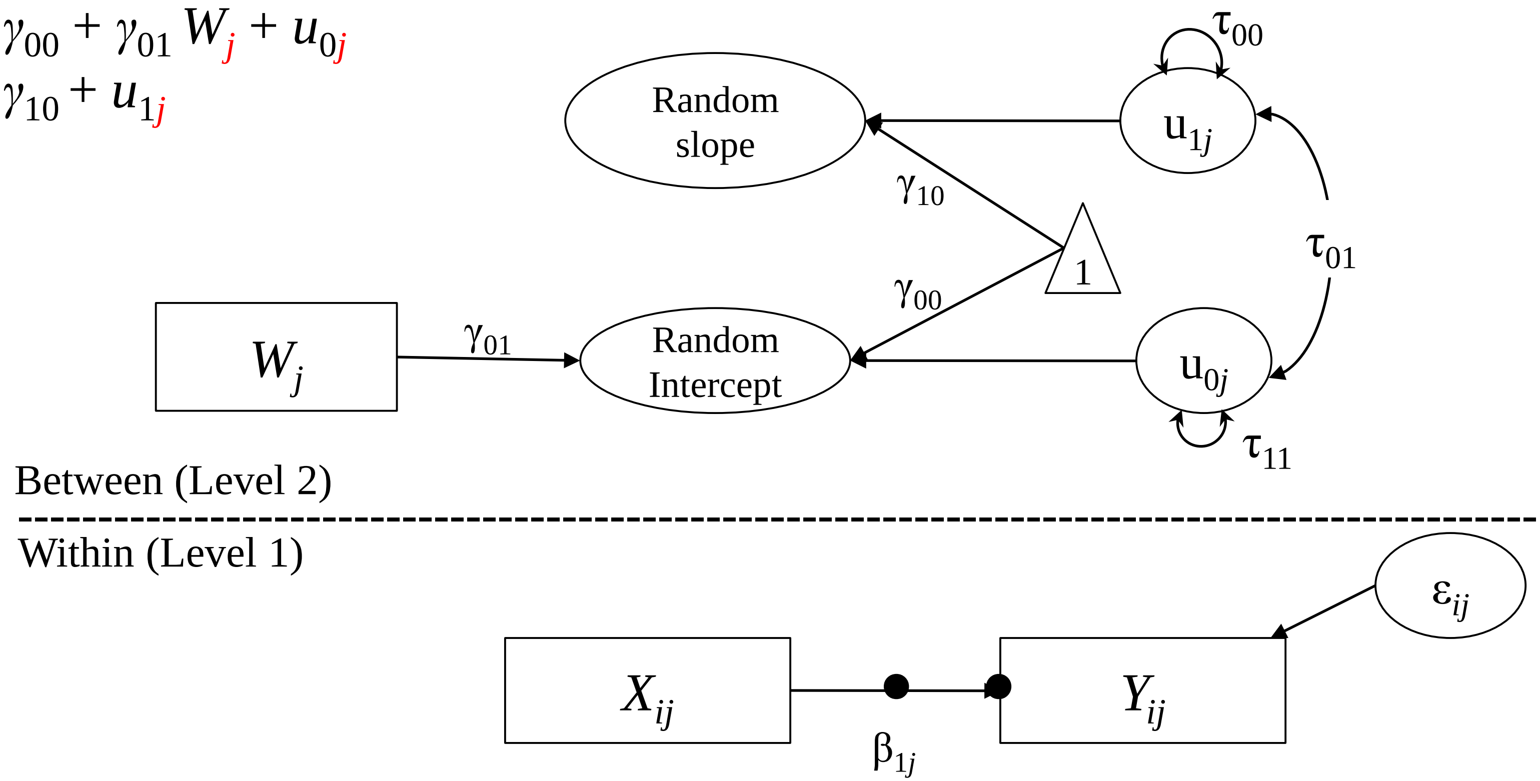
Note. Figure from Hamaker (in press). For more details on RI-CLPM and its dispute with classical ACLPM, see also Hamaker et al. (2015) and Lucas (2023). For a more theoretical contribution, see Hamaker (2012) and Borsboom & Haslbeck (2024)

Latent Growth Models (Meredith & Tisak, *Psychometrika*, 1990)



Multilevel SEM

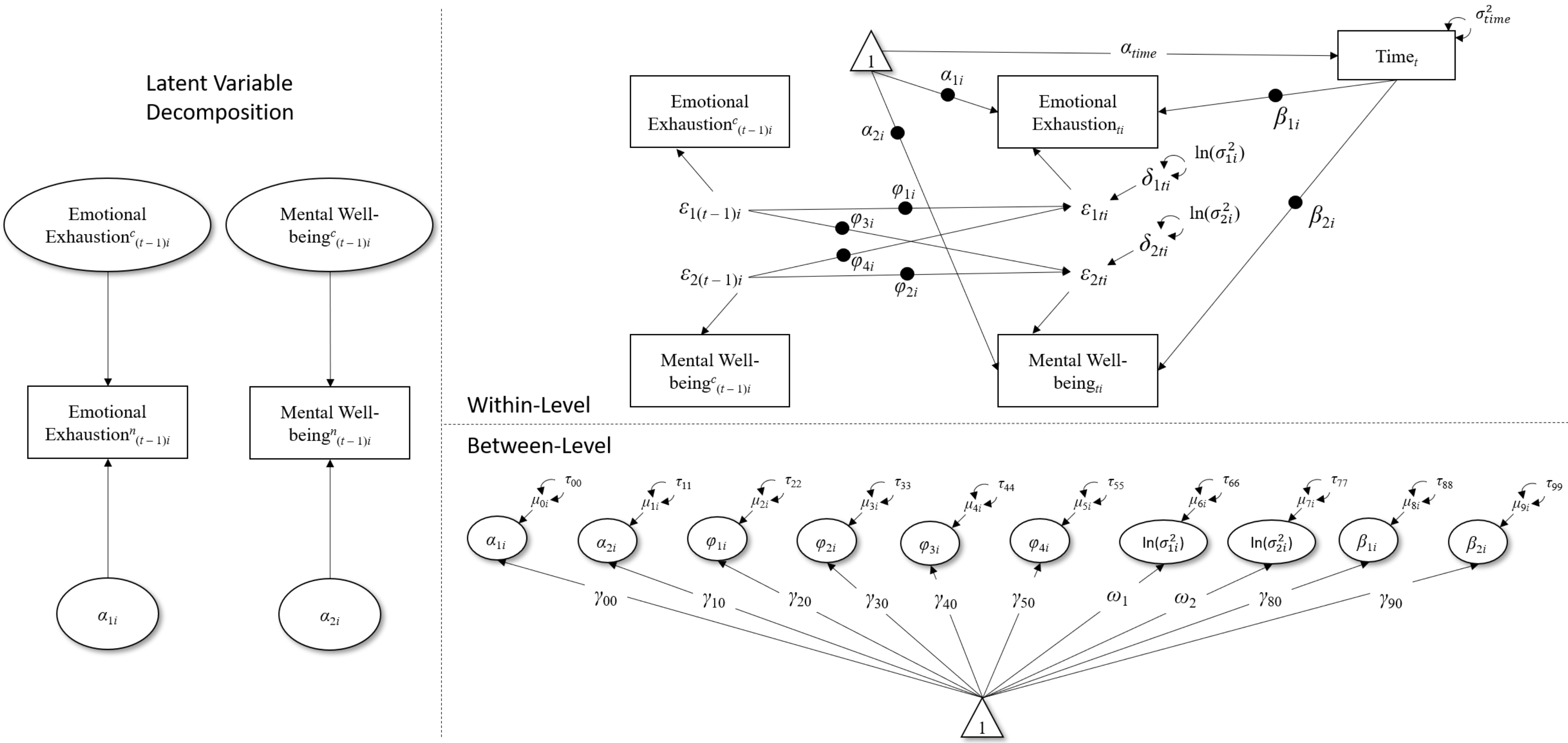
Level 1: $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij}$
Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$
Level 2: $\beta_{1j} = \gamma_{10} + u_{1j}$



For sake of clarity, Formula and Figure refer to a simple multilevel regression.
The use of latent variables, or adding more than one dependent variable, would make this model a Multilevel SEM.

Dynamic SEM

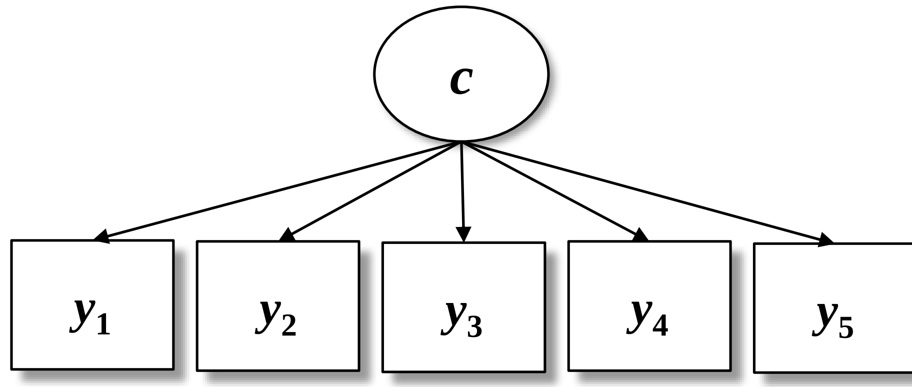
(Combination of Time Series Modeling, Multilevel Modeling, and SEM)



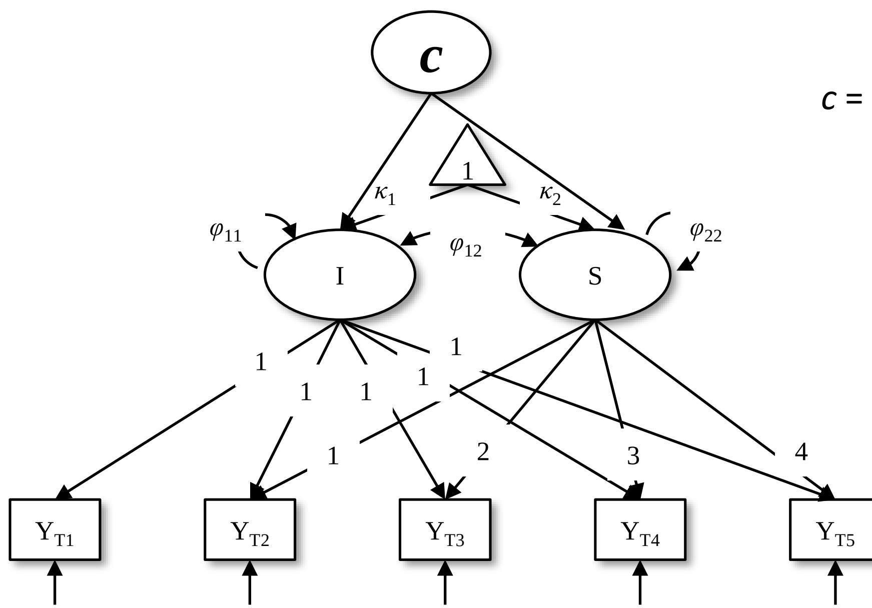
From Perinelli, Vignoli, Kröner, Müller, Genrich, & Fraccaroli (2023). Syntaxes available in Suppl Material. See «Suggested Reading» section for literature on DSEM

Mixture Models (Muthén, *Psychometrika*, 1989)

Latent Class Analysis (LCA; if *y* is dichotomous)
Latent Profile Anlaysis (LPA; if *y* is continuous)



Growth Mixture Model (GMM)



c = latent classes

Advancing General Linear Models

frontiers in Psychology

METHODS
published: 02 March 2017
doi: 10.3389/fpsyg.2017.00223

Check for updates

Evaluating Intervention Programs with a Pretest-Posttest Design: A Structural Equation Modeling Approach

Guido Alessandri[1], Antonio Zuffianò[2] and Enrico Perinelli[1]*

[1] Department of Psychology, Sapienza University of Rome, Rome, Italy, [2] Department of Psychology, Liverpool Hope University, Liverpool, UK

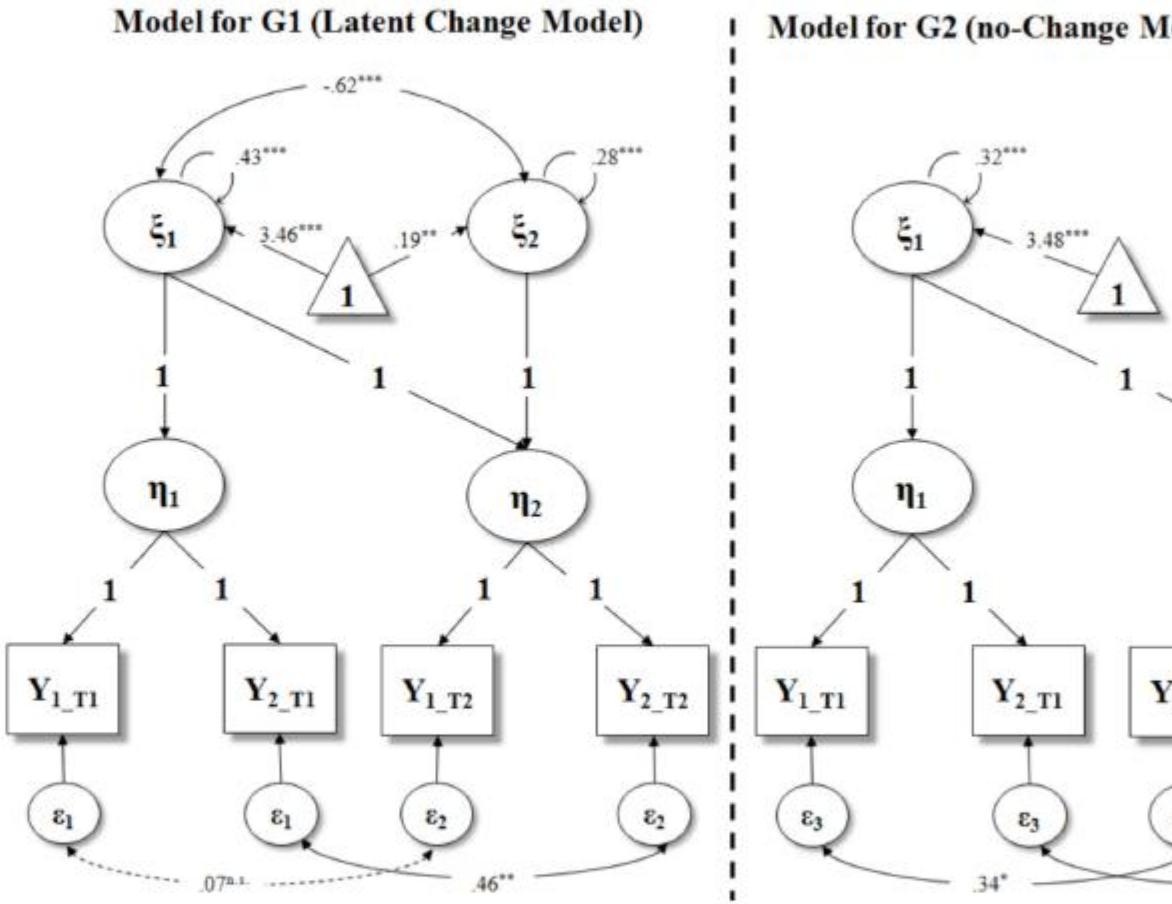
A common situation in the evaluation of intervention programs is the researcher's possibility to rely on two waves of data only (i.e., pretest and posttest), which profoundly impacts on his/her choice about the possible statistical analyses to be conducted. Indeed, the evaluation of intervention programs based on a pretest-posttest design has been usually carried out by using classic statistical tests, such as family-wise ANOVA analyses, which are strongly limited by exclusively analyzing the intervention effects at

Alessandri, G., Zuffianò, A., & Perinelli, E. (2017). Evaluating intervention programs with a pretest-posttest design: A Structural Equation Modeling approach. *Frontiers in Psychology: Quantitative Psychology and Measurement*, *8*, Article 223. https://doi.org/10.3389/fpsyg.2017.00223

Advancing General Linear Models

Second-Order Multiple-Group Latent Curve Model

G1 = Intervention group
G2 = Control group



Alessandri, Zuffianò, & Perinelli (2017)

Advancing Mixed Models

Second-Order Multiple-Group Latent Curve Model

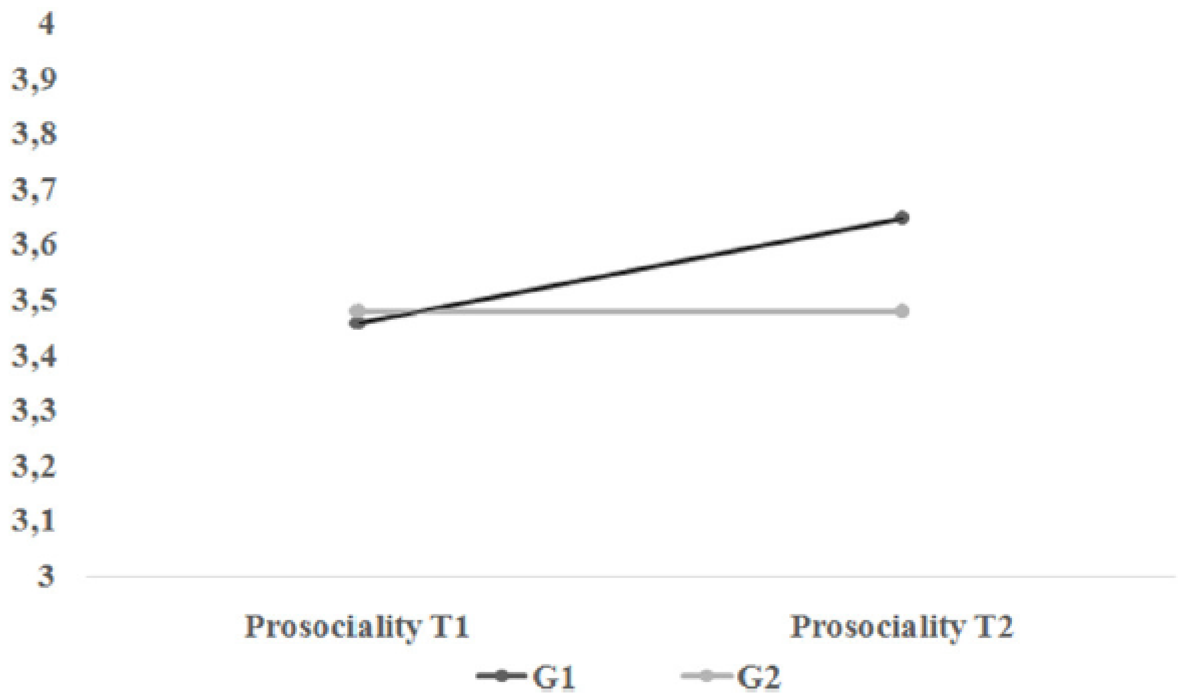


FIGURE 3 | Trajectories of prosocial behavior for intervention group (G1) and control group (G2) in the best fitting model (Model 2 in Table 2).

G1 = Intervention group
G2 = Control group

Alessandri, Zuffianò, & Perinelli (2017)

Other Types

1) Latent Transition Analysis
2) Latent Change (Difference) Score
3) Latent State-Trait (LST) Models

PhD Thesis

On the Validity of the State-Like Component of Global Self-Esteem:

Relationships with Implicit Self-Esteem and Work-Related Variables

Using Different Latent State-Trait Models

Supervisor
Professor Guido Alessandri

PhD Candidate
Enrico Perinelli

February 2018

and much else….

Introduction to M*plus*

M*plus*

M*plus* is currently one of the most used software for latent variables analyses.

Copyright © 1998-2017 Muthén & Muthén
Program Copyright © 1998-2017 Muthén & Muthén
Version 8
April 2017

Muthén & Muthén
3463 Stoner Avenue
Los Angeles, CA 90066
Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support@StatModel.com

M*plus*

Its popularity is mainly due to:

a user-friendly syntax

Syntax	*Action*
`ON`	Regressed on (defines structural paths)
`BY`	Measured by (defines factor loadings)
`WITH`	Covaried with
`IND`	Indirect effect
`[]`	Mean structure and intercept
`()`	Assigns a label
`@`	Fixes a parameter to a specific value
`*`	Parameter freed/Specifies a starting value for a parameter
`MODEL INDIRECT`	Specifies the indirect effect to be computed
`MODEL CONSTRAINT`	Specifies linear and nonlinear constraints

M*plus*

Its popularity is mainly due to:

The availability of a web-site (https://www.statmodel.com/) that provides:

1. information on the use of the software through a full-detailed user's guide (https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf)

2. A forum in which users can deal (by questioning, by answering or simply by reading) with a broad variety of topics (https://www.statmodel.com/cgi-bin/discus/discus.cgi)

3. A pletora of real data examples (https://www.statmodel.com/ugexcerpts.shtml)

4. Videos and Handouts for Short Courses (https://www.statmodel.com/course_materials.shtml)

M*plus*

Its popularity is mainly due to:

The availability of several estimators.

Syntax	*Estimator*
`ML`	Maximum Likelihood
`MLM`	Satorra-Bentler chi-square
`MLMV`	ML with mean- and variance-adjusted
`MLR`	Yuan-Bentler chi-square
`MLF`	ML with standard errors approximated by first-order derivatives
`MUML`	Muthén's limited information parameter estimates
`WLS`	Weighted least square
`WLSM`	Weighted least square mean-adjusted
`WLSMV`	Weighted least square mean- and variance-adjusted
`ULS`	Unweighted least squares
`ULSMV`	ULS mean- and variance-adjusted
`GLS`	Generalized Least Square
`BAYES`	Bayesian posterior parameter estimates with credibility intervals and posterior predictive checking

M*plus*

Its popularity is mainly due to:

Constantly updated software

*Current version (Feb 2023): **8.8***

Version 8	April 2017
Version 7	September 2012
Version 6	April 2010
Version 5	November 2007
Version 4	February 2006
Version 3	March 2004
Version 2	February 2001
Version 1	November 1998

M*plus*

Its popularity is mainly due to:
The availability of several books

- Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York and London: Routledge.
- Geiser, C. (2013). *Data analysis with Mplus*. New York: Guilford Press.
- Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth Modeling: Structural equation and multilevel modeling approaches*. New York: Guilford Press.
- Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*. New York and London: Routledge.
- Muthén, B., Muthén, L. K., Asparouhov , T. (2016). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén.
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. Chichester, UK: John Wiley & Sons.
- Wickrama, K. A. S., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). *Higher order growth curves and mixture modeling with Mplus: A practical guide*. New York, NY: Routledge.

M*plus*

Important steps for preparing dataset to be used in M*plus*:

1) Coding missing values (I suggest -99 if you work in SPSS; for R, see the subsequent slide)
2) Create an ASCII data file and eventually convert commas with dots (in Notepad; for R this is not necessary)
3) Commands and Basic analysis (in M*plus*)
4) Output for Basic analysis (in M*plus*)

M*plus*

Important steps for preparing dataset to be used in M*plus*:

In order to create a dataset for M*plus*, for R use the subsequent syntax

```
# Initial setup -----------------------------------------------------------

rm(list = ls())
setwd("C:/Users/Enrico Perinelli/Desktop") # set your working directory or use a .Rproj file)


# Load or exctract dataset ------------------------------------------------

library(psych)
my_bfi <- bfi


# Load `MplusAutomation` package ------------------------------------------

library(MplusAutomation)


# Use the `prepareMplusData` function -------------------------------------

prepareMplusData(
  my_bfi,
  filename = "bfi.dat"
)
```

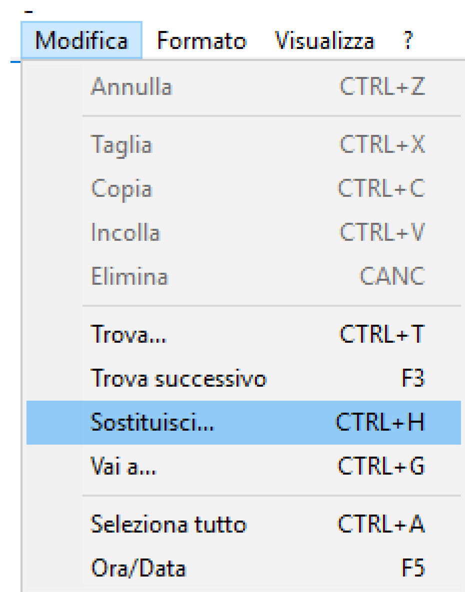
M*plus*

big_five - Blocco note

File Modifica Formato Visualizza ?

404	3,50	3,63	4,38	4,50	1,13
311	4,00	4,75	4,75	4,13	2,38
379	2,88	3,50	3,75	3,88	1,38
338	4,38	3,75	4,00	3,00	1,13
209	3,75	4,38	4,13	4,25	1,63
141	-99,00	5,00	5,00	5,00	4,50
5	4,00	4,00	4,00	3,88	1,50
83	3,13	3,13	4,25	3,75	1,25
106	3,88	2,63	3,88	4,00	1,38
409	3,75	4,13	-99,00	-99,00	1,00
414	3,63	4,00	4,13	4,00	1,25
312	3,38	3,13	4,00	3,13	1,75
54	4,00	3,38	4,88	4,75	-99,00
296	3,00	3,00	3,00	3,00	3,00
411	4,38	4,00	4,88	4,50	1,00
343	4,25	4,13	5,00	4,00	1,00
287	2,88	3,00	3,00	3,00	2,63
369	4,00	-99,00	4,63	4,63	1,25
65	-99,00	4,63	4,50	4,50	1,00
89	3,50	4,63	4,13	3,38	1,25
347	3,13	3,50	-99,00	4,25	2,50
96	3,63	2,00	4,25	4,13	1,00
206	4,75	3,50	5,00	4,88	-99,00
142	1,00	1,00	1,00	1,00	1,50
144	4,00	5,00	5,00	5,00	3,50
143	4,00	4,00	4,00	4,00	3,75
324	3,63	3,38	4,50	4,00	1,13

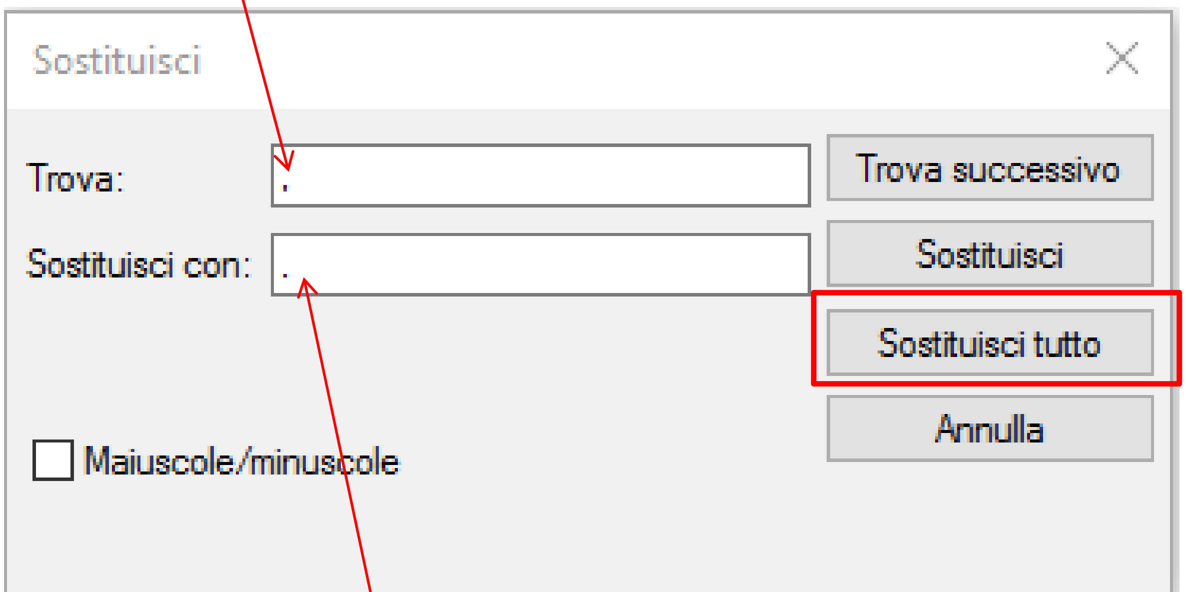
For SPSS, instead, you must be sure that commas are replaced with dots (1)



M*plus*

For SPSS, instead, you must be sure that commas are replaced with dots (2)

, (comma)



. (dot)

Mplus

For SPSS, instead, you must be sure that commas are replaced with dots (3)

big_five - Blocco note

File Modifica Formato Visualizza ?

404	3.50	3.63	4.38	4.50	1.13
311	4.00	4.75	4.75	4.13	2.38
379	2.88	3.50	3.75	3.88	1.38
338	4.38	3.75	4.00	3.00	1.13
209	3.75	4.38	4.13	4.25	1.63
141	-99.00	5.00	5.00	5.00	4.50
5	4.00	4.00	4.00	3.88	1.50
83	3.13	3.13	4.25	3.75	1.25
106	3.88	2.63	3.88	4.00	1.38
409	3.75	4.13	-99.00	-99.00	1.00
414	3.63	4.00	4.13	4.00	1.25
312	3.38	3.13	4.00	3.13	1.75
54	4.00	3.38	4.88	4.75	-99.00
296	3.00	3.00	3.00	3.00	3.00
411	4.38	4.00	4.88	4.50	1.00
343	4.25	4.13	5.00	4.00	1.00
287	2.88	3.00	3.00	3.00	2.63
369	4.00	-99.00	4.63	4.63	1.25
65	-99.00	4.63	4.50	4.50	1.00
89	3.50	4.63	4.13	3.38	1.25
347	3.13	3.50	-99.00	4.25	2.50
96	3.63	2.00	4.25	4.13	1.00
206	4.75	3.50	5.00	4.88	-99.00
142	1.00	1.00	1.00	1.00	1.50
144	4.00	5.00	5.00	5.00	3.50
143	4.00	4.00	4.00	4.00	3.75

NOW, ALL THE COMMAS ARE DOTS

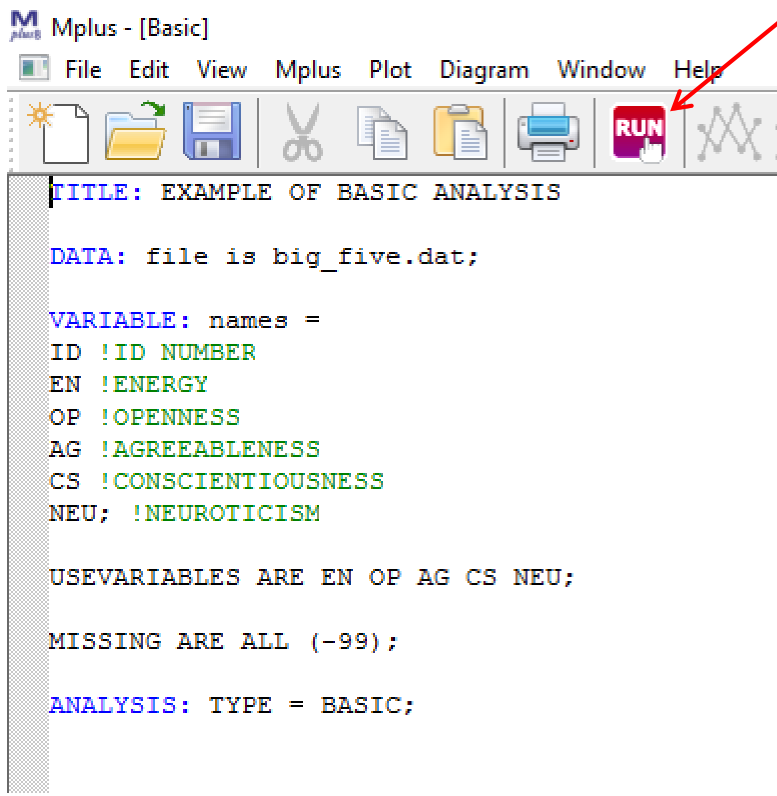
M*plus*

Syntax	*Function*
`TITLE`	Provide a title for the analysis
`DATA`	Provide information about the data set to be analyzed
`VARIABLE`	Provide information about the variables in the data set
`DEFINE`	Transform existing variables and create new variables
`ANALYSIS`	Describe the technical details of the analysis (e.g., Estimator, type)
`MODEL`	Describe the model to be estimated
`OUTPUT`	Request additional output not included as the default
`SAVEDATA`	Save the analysis data (e.g., Auxiliary data)
`PLOT`	Request graphical displays of analysis results
`MONTECARLO`	Specify the details of a monte carlo simulation study

!	This symbol is used to take notes
;	This symbol must be placed at the end of each command line

Mplus

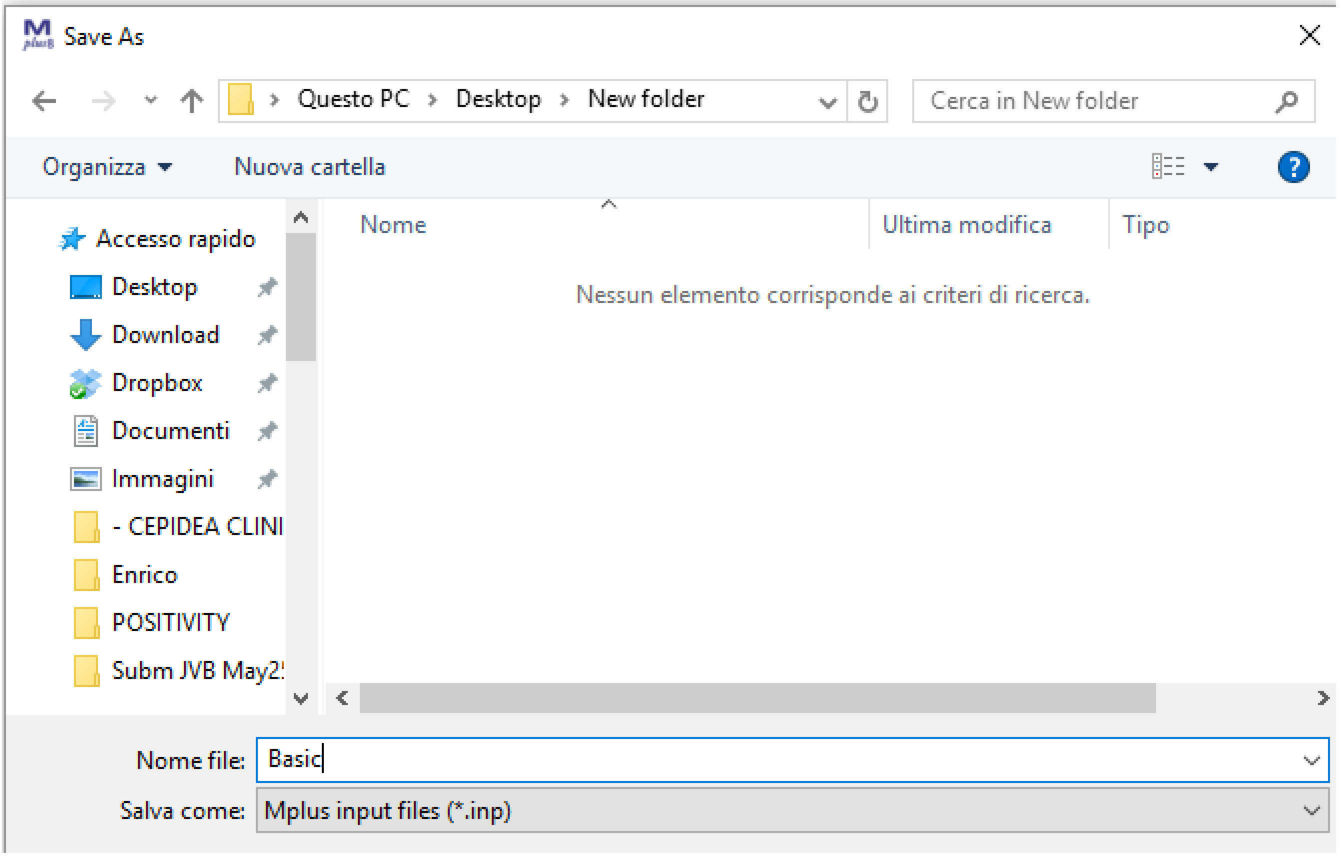
CLICK RUN



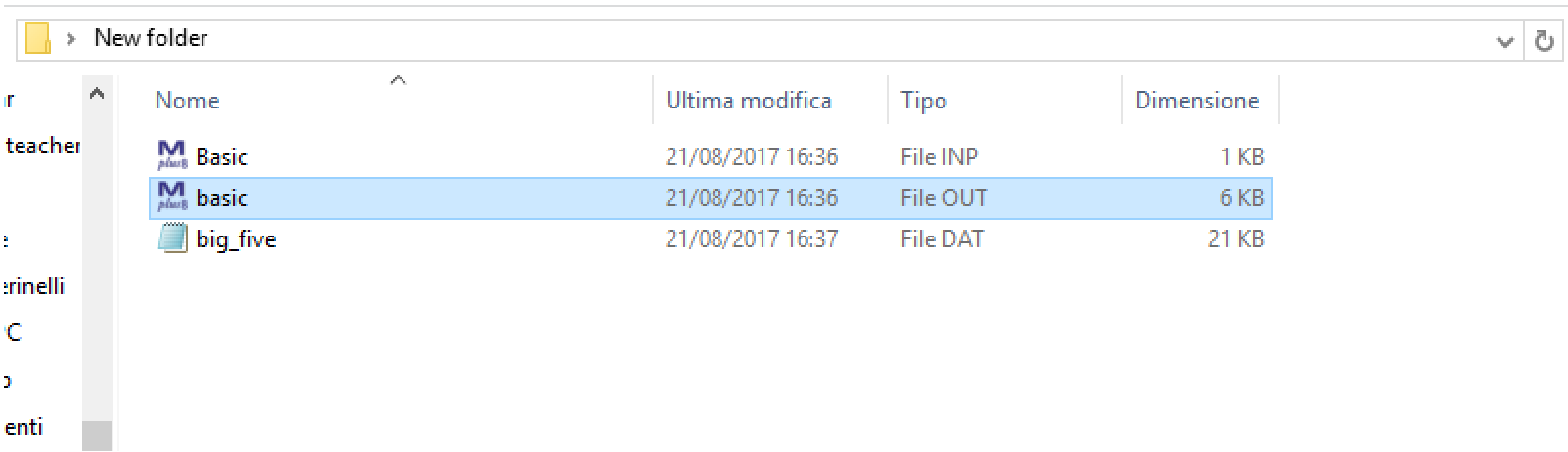
Remember that variable names cannot be longer than 8 characters

Mplus

Save the input file in the same folder of the .dat file



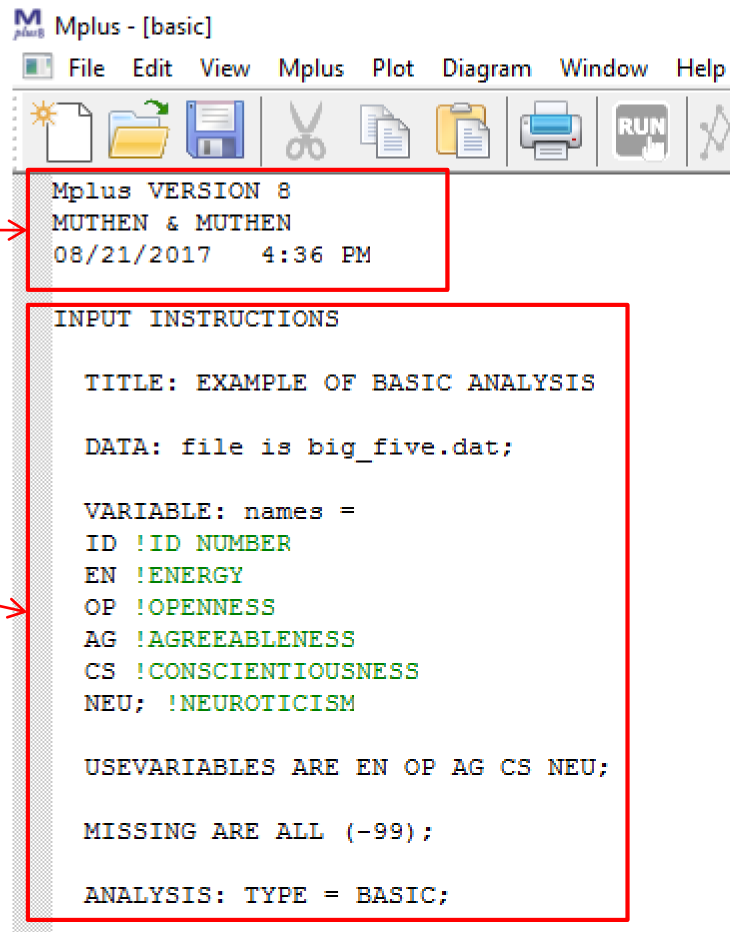
M*plus*


New folder
Nome
Ultima modifica
Tipo
Dimensione
Basic
21/08/2017 16:36
File INP
1 KB
basic
21/08/2017 16:36
File OUT
6 KB
big_five
21/08/2017 16:37
File DAT
21 KB
teacher
erinelli
enti

M*plus*

M*plus* version and date

Input instruction



Mplus

```
INPUT READING TERMINATED NORMALLY
```

The analysis has gone well!!

```
EXAMPLE OF BASIC ANALYSIS

SUMMARY OF ANALYSIS

Number of groups                                      1
Number of observations                              416

Number of dependent variables                         5
Number of independent variables                       0
Number of continuous latent variables                 0

Observed dependent variables

  Continuous
   EN          OP          AG          CS          NEU
```

Title and information on the number of variables. Here we have only continuous variables, but this part of the output may displays also the number of latent variables

```
Estimator                                                    ML
Information matrix                                     OBSERVED
Maximum number of iterations                               1000
Convergence criterion                                 0.500D-04
Maximum number of steepest descent iterations                20
Maximum number of iterations for H1                        2000
Convergence criterion for H1                          0.100D-03

Input data file(s)
  big_five.dat

Input data format  FREE
```

By default in most of analyses the estimator is ML (in Multilevel modeling the default estimator is MLR)

Mplus

```
COVARIANCE COVERAGE OF DATA

Minimum covariance coverage value   0.100


     PROPORTION OF DATA PRESENT


           Covariance Coverage
              EN            OP            AG            CS            NEU
              ________      ________      ________      ________      ________
 EN             0.988
 OP             0.954         0.966
 AG             0.957         0.935         0.969
 CS             0.966         0.950         0.952         0.978
 NEU            0.962         0.945         0.945         0.952         0.974
```

This matrix shows the percentage of covariance coverage for each pair of observed variables. When dataset does not have missing data, each pair of observed variables has a covariance coverage of 100%.

M*plus*

```
RESULTS FOR BASIC ANALYSIS


     ESTIMATED SAMPLE STATISTICS


           Means
              EN            OP            AG            CS            NEU
              ________      ________      ________      ________      ________
                3.812         3.811         4.463         4.097         1.382


           Covariances
              EN            OP            AG            CS            NEU
              ________      ________      ________      ________      ________
 EN             0.279
 OP             0.129         0.386
 AG             0.112         0.131         0.204
 CS             0.089         0.132         0.123         0.264
 NEU           -0.024        -0.004        -0.068        -0.012         0.264


           Correlations
              EN            OP            AG            CS            NEU
              ________      ________      ________      ________      ________
 EN             1.000
 OP             0.392         1.000
 AG             0.470         0.469         1.000
 CS             0.330         0.412         0.531         1.000
 NEU           -0.087        -0.012        -0.294        -0.046         1.000
```

Here, information concerning Means, Covariances and Correlations are reported.

M*plus*

UNIVERSITATE SAMPLE STATISTICS

UNIVARIATE HIGHER-ORDER MOMENT DESCRIPTIVE STATISTICS

Variable/ Sample Size		Mean/ Variance	Skewness/ Kurtosis	Minimum/ Maximum	% with Min/Max	Percentiles 20%/60%	40%/80%	Median
EN		3.810	-0.390	1.000	0.24%	3.380	3.630	3.880
	411.000	0.279	1.442	5.000	1.70%	4.000	4.250	
OP		3.809	-0.407	1.000	0.25%	3.250	3.750	3.880
	402.000	0.386	0.654	5.000	2.49%	4.000	4.380	
AG		4.465	-1.496	1.000	0.25%	4.000	4.380	4.500
	403.000	0.205	7.433	5.000	14.89%	4.630	4.880	
CS		4.098	-0.806	1.000	0.25%	3.630	4.000	4.130
	407.000	0.264	2.712	5.000	4.91%	4.250	4.500	
NEU		1.382	2.439	1.000	29.63%	1.000	1.130	1.130
	405.000	0.264	8.499	4.500	0.49%	1.250	1.750	

Here, information concerning univariate sample statistics are reported.

A VERY IMPORTANT STEP IS TO CHECK THE TRUSTWORTHINESS OF THESE VALUES!!!!
WHEN THE DATASET IS LARGE, IT IS EASY TO WRONG SPECIFY VARIABLE NAMES, OR FORGET TO SPECIFY THE MASSING DATA VALUE.

M*plus* User's guide

TABLE OF CONTENTS

Intro

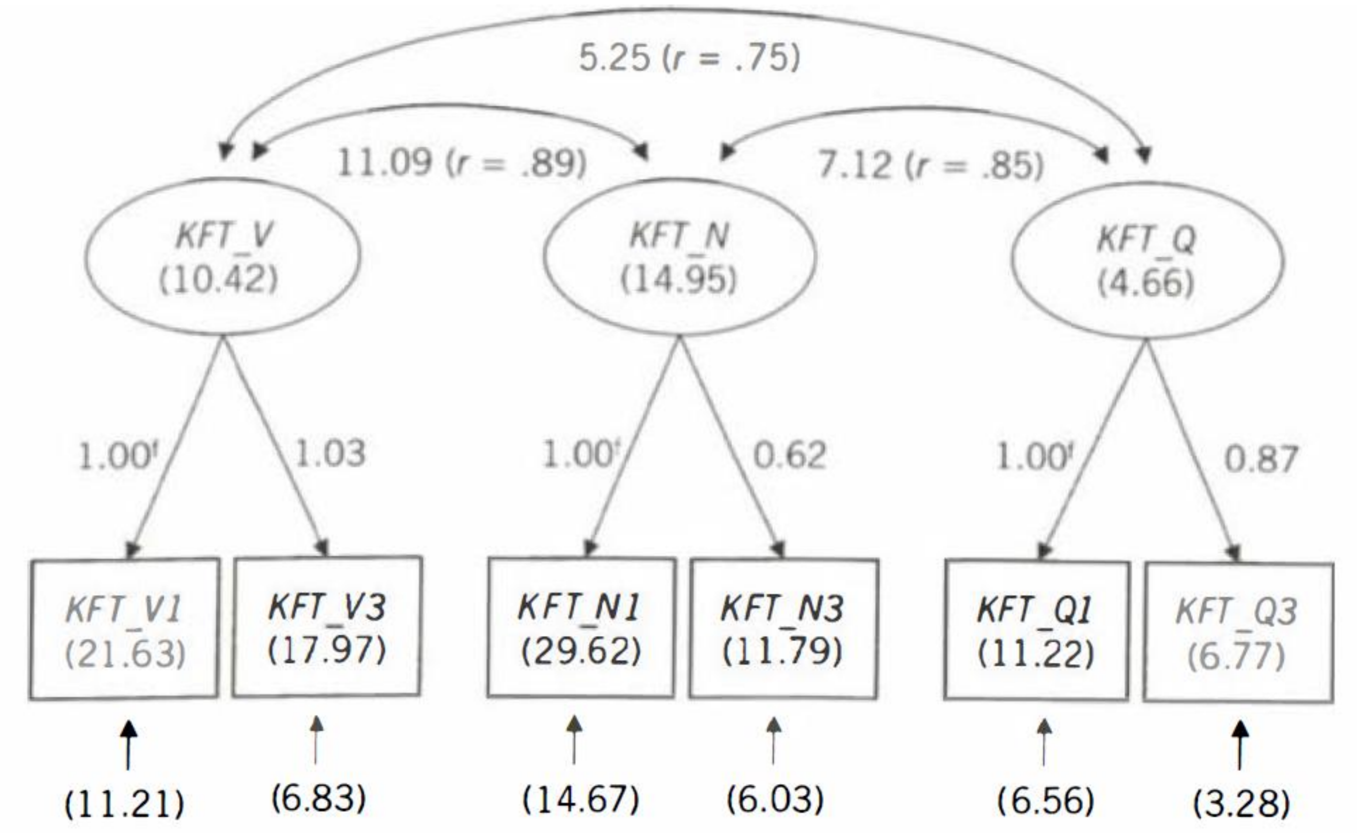
Specific Models (types of SEM)

Information regarding commands and syntaxes

Link:
https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

M*plus* CFA example

(Geiser, 2013, p. 53)



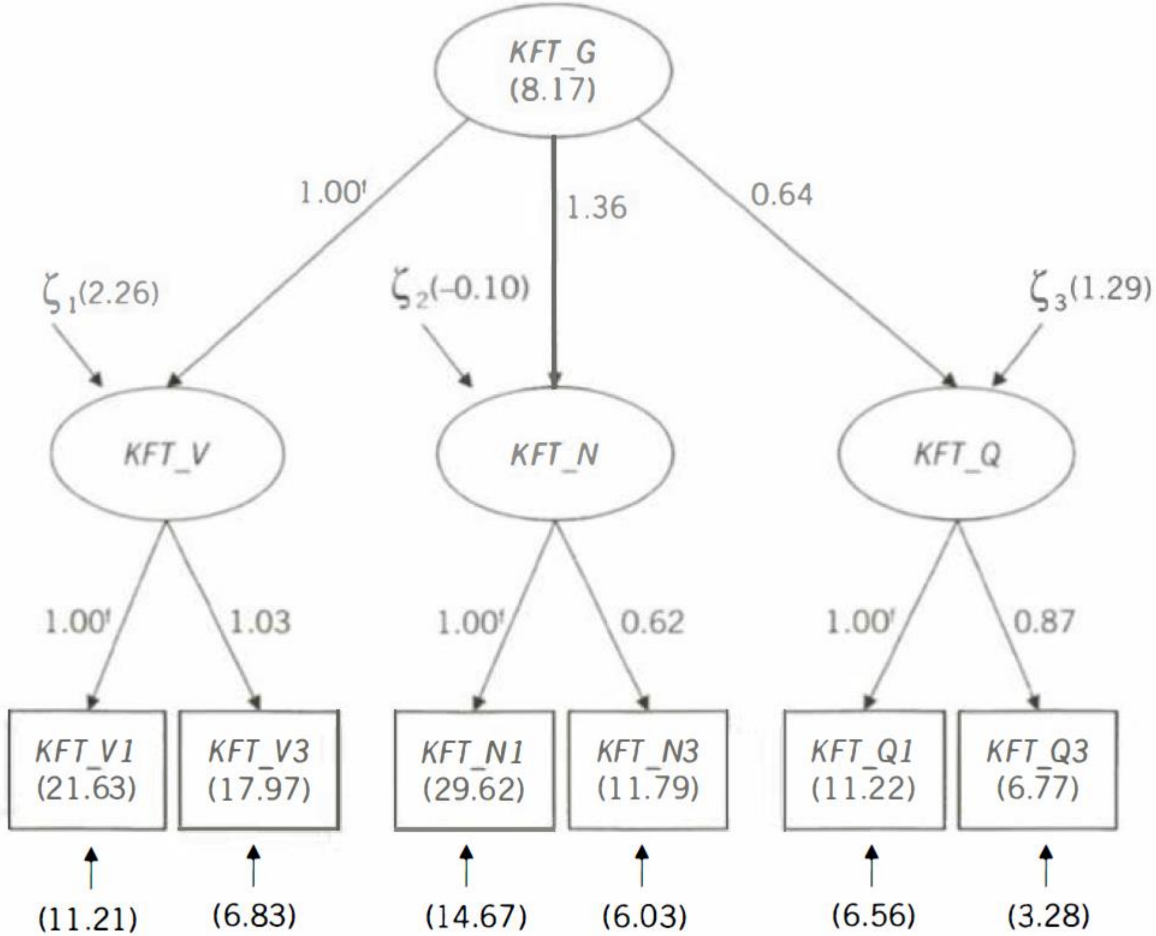
M*plus* CFA example

(Geiser, 2013, p. 53)

```
title: CFA for the six KFT variables
       3-Factor model (see Figure 3.8)

data: file = KFT.dat;
      listwise = on;

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
          missing = all(-99);

analysis: type = general;

model: KFT_V by kft_v1 kft_v3; ! verbal ability factor
       KFT_Q by kft_q1 kft_q3; ! numeric ability factor
       KFT_N by kft_n1 kft_n3; ! figural (non-verbal) ability factor

output: sampstat stdyx tech4;
```

M*plus* Second-Order CFA example

(Geiser, 2013, p. 59)



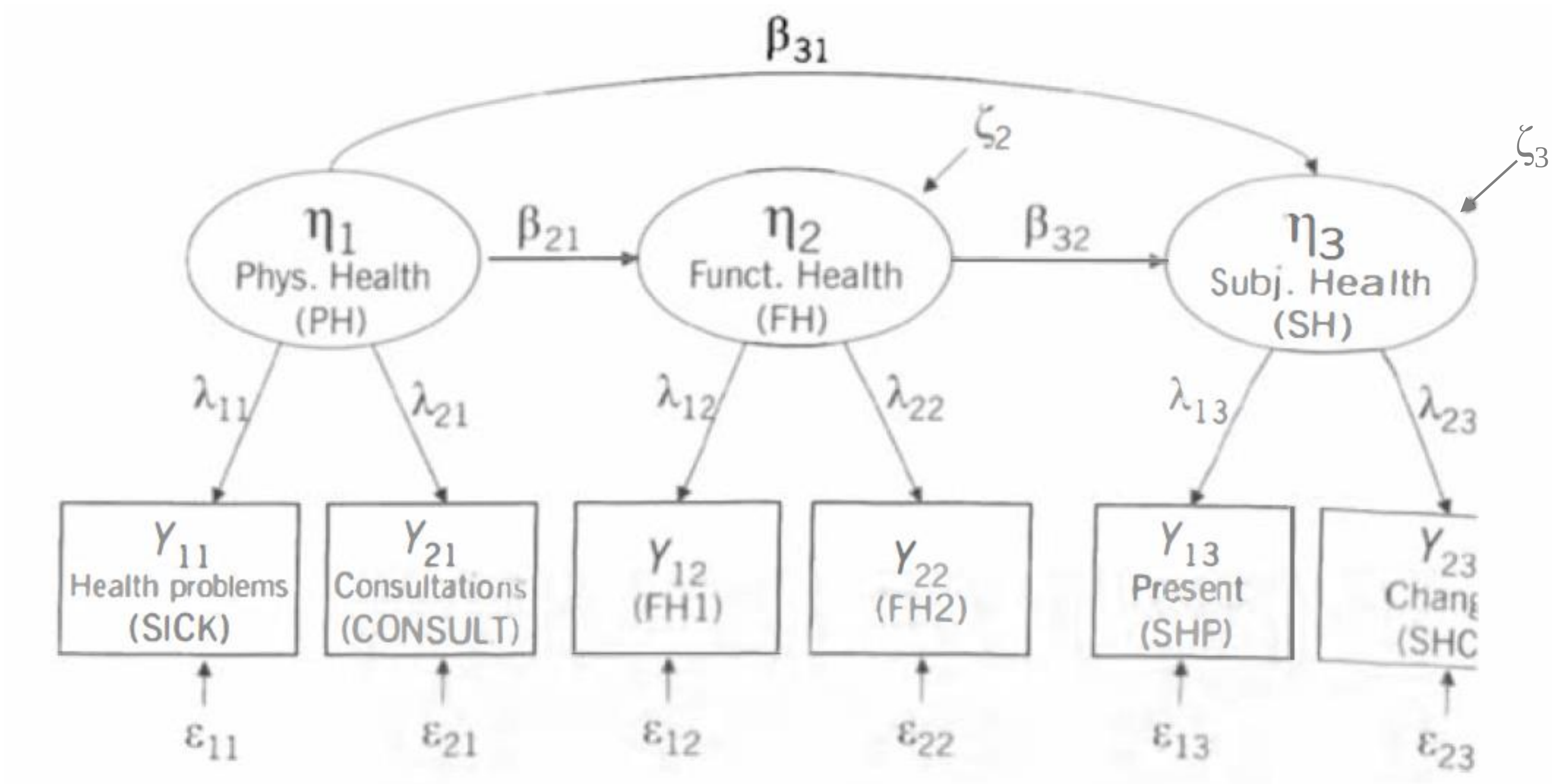
M*plus* Second-Order CFA example

(Geiser, 2013, p. 59)

```
title: CFA for the six KFT variables
       Second-order factor model (see Figure 3.11)

data: file = KFT.dat;
      listwise = on;

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
          missing = all(-99);

analysis: type = general;

model: ! First-order factors
       KFT_V by kft_v1 kft_v3; ! verbal ability factor
       KFT_Q by kft_q1 kft_q3; ! numeric ability factor
       KFT_N by kft_n1 kft_n3; ! figural (non-verbal) ability factor

       ! Second-order factor
       KFT_G by KFT_V KFT_Q KFT_N;

output: sampstat stdyx;
```

M*plus* SEM example

(Geiser, 2013, p. 74)



M*plus* SEM example

(Geiser, 2013, p. 74)

```
title: Latent path model shown in Figure 3.15
       Relationship between physical health, functional health, and subjective health

data: file = health.dat;

variable: names = SHP ! Subjektive health in the present
                  SHC ! Changes in subjective health across the past 6 years
                  SICK ! Number of health issues in the past 12 months
                  CONSULT ! Number of physician consultations in the past 12 months
                  FH ! Total score SF-36 scale on functional health
                  FH1 FH2; ! SF-36 scale split into 2 test halves/item parcels

          usevar = SHP SHC SICK CONSULT FH1 FH2;

analysis: type = general;

model: SH by SHP SHC; ! Factor/Measurement model subjective health
       PH by SICK CONSULT; ! Factor/Measurement model physical health
       FH by FH1 FH2; ! Factor/Measurement model functional health

! Path analysis at the level of the latent variables
       FH on PH;
       SH on PH FH;

       model indirect: SH ind PH; ! Request output of direct, indirect, and total effects

output: sampstat stdyx;
```

M*plus* SEM example (bootstrap)

(Geiser, 2013, p. 74)

```
title: Latent path model shown in Figure 3.15
       Relationship between physical health, functional health, and subjective health
       With bias-corrected bootstrap confidence intervals

data: file = health.dat;

variable: names = SHP ! Subjektive health in the present
                  SHC ! Changes in subjective health across the past 6 years
                  SICK ! Number of health issues in the past 12 months
                  CONSULT ! Number of physician consultations in the past 12 months
                  FH ! Total score SF-36 scale on functional health
                  FH1 FH2; ! SF-36 scale split into 2 test halves/item parcels

          usevar = SHP SHC SICK CONSULT FH1 FH2;

analysis: type = general;
          bootstrap = 10000; ! Request 10000 bootstrap samples

model: SH by SHP SHC; ! Factor/Measurement model subjective health
       PH by SICK CONSULT; ! Factor/Measurement model physical health
       FH by FH1 FH2; ! Factor/Measurement model functional health

! Path analysis at the level of the latent variables
       FH on PH;
       SH on PH FH;

       model indirect: SH ind PH; ! Request output of direct, indirect, and total effects

output: sampstat stdyx
        cinterval(bcbootstrap); ! Request confidence intervals based on bias-corrected bootstrap
```

Suggested Reading

Suggested 'Starters' for SEM

- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.
- Hoyle, R. K. (2023). *Handbook of Structural Equation Modeling* (2nd ed.). The Guilford Press.
- Hancock, G. R., & Mueller, R. O. (2013). *Structural Equation Modeling: A second course*. IAP.

Old but gold - Still the best book on SEM mathematical bases

- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.

Longitudinal SEM

- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Wiley.
- Little, T. D. (2024). *Longitudinal structural equation modeling* (2nd ed.). The Guilford Press.
- McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. American Psychological Association. doi: 10.1037/14440-000
- Newsom, J. T. (2024). *Longitudinal structural equation modeling: A comprehensive introduction* (2nd ed.). Routledge.
- Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth Modeling: Structural equation and multilevel modeling approaches*. The Guilford Press.
- Ferrer, E., Boker, S. M., & Grimm, K. J. (2019). *Longitudinal multivariate psychology*. Routledge.

Suggested Reading

CFA

- Brown, T. A. (2015). *Confirmatory Factor Analysis for applied research* (2nd ed.). The Guilford Press.

Mixture

- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 535-569. doi:10.1080/10705510701575396
- Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440–461. https://doi.org/10.1037/tps0000176
- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*, 302-317. doi:10.1111/j.1751-9004.2007.00054.x
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis*: *With applications in the social, behavioral, and health sciences.* John Wiley & Sons, Inc.
- van De Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-Checklist: Guidelines for Reporting on Latent Trajectory Studies. *Structural Equation Modeling: A Multidisciplinary Journal, 24*, 451-467. doi: 10.1080/10705511.2016.1247646

Suggested Reading

Dynamic Structural Equation Modeling (DSEM)

- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803
- Asparouhov, T., & Muthén, B. (2020). Comparison of models for the analysis of intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(2), 275–297. https://doi.org/10.1080/10705511.2019.1626733
- Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic Structural equation models for the affective measurements from the COGITO study. *Multivariate Behavioral Research*, *53*(6), 820–841. https://doi.org/10.1080/00273171.2018.1446819
- Hamaker, E. L., Asparouhov, T., & Muthén, B. (2023). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., pp. 576-596). The Guilford Press.
- McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in M*plus*. *Psychological Methods*, *25*(5), 610–635. https://doi.org/10.1037/met0000250
- Wang, J., & Wang, X. (2020). *Structural Equation Modeling: Applications Using Mplus* (2nd ed.). John Wiley & Sons. **[Section 4.8, pp. 241-252]**

Suggested Reading

Maximum Likelihood

- Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.)(Chapter 2). The Guilford Press.
- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*, 90-100. doi: 10.1016/S0022-2496(02)00028-7
- Stigler, S. M. (2007). The epic story of maximum likelihood. *Statistical Science, 22*(4), 598-620.

M*plus*

- Geiser, C. (2013). *Data analysis with Mplus*. The Guilford Press.
- Geiser, C. (2021). *Longitudinal structural equation modeling with Mplus: A Latent State-Trait perspective*. The Guilford Press.
- Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.
- Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using Mplus* (2nd ed.). John Wiley & Sons.
- Wickrama, K. A. S., Lee, T. K., O'Neal, C. W., & Lorenz, F. O. (2016). *Higher order growth curves and mixture modeling with Mplus: A practical guide*. Routledge.
- Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus* (3rd ed.). Routledge.
- Muthén, B., Muthén, L. K., Asparouhov , T. (2016). *Regression and mediation analysis using Mplus*. Muthén & Muthén.

Suggested Reading

lavaan (R package)

- https://lavaan.ugent.be/tutorial/
- https://lavaan.ugent.be/tutorial/tutorial.pdf
- https://stats.oarc.ucla.edu/r/seminars/rsem/
- https://users.ugent.be/~yrosseel/lavaan/gent2020/lavaan_twodays_gent2020.pdf
- Geiser, C. (2023). Structural equation modeling with the M*plus* and lavaan programs. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., pp. 241-258). Guilford Press.

SEM in Python with `semopy` package

- https://pypi.org/project/semopy/ and https://semopy.com/
- Igolkina, A. A., & Meshcheryakov, G. (2020). semopy: A Python package for Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(6), 952-963. https://doi.org/10.1080/10705511.2019.1704289

Bayesian SEM (BSEM)

Books and Chapters

- Depaoli, S. (2021). *Bayesian Structural Equation Modeling*. The Guilford Press.
- Kaplan, D. (2014). *Bayesian statistics for the social sciences*. The Guilford Press.
- Kaplan, D., & Depaoli, S. (2012). Bayesian Structural Equation Modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling* (pp. 650-673). The Guilford Press.
- Lee, S. Y., & Song, X. Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences*. John Wiley & Sons.

Suggested Reading

Bayesian SEM (BSEM)

Articles:

- Lee, S. Y. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika*, *46*(2), 153-160. **[First BSEM ever]**
- Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, *22*(2), 240–261.
- McNeish, D. (2016) On using bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750-773.
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335.
- Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2020). Bayesian versus frequentist estimation for Structural Equation Models in small sample contexts: A systematic review. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 131-161.
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*(3), 842-860.
- Zondervan-Zwijnenburg, M., Peeters, M., Depaoli, S., & van de Schoot, R. (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development*, *14*(4), 305-320.

Suggested Reading

Machine Learning and Latent Variables/Psychometrics (*Part 1*)

- Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review, 23*(2), 190-203. https://doi.org/10.1177/1088868318772990
- Brandmaier, A. M., & Jacobucci, R. C. (2023). Machine Learning approaches to structural Equation Modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., pp. 722-739). The Guilford Press.
- Brandmaier, A. M., von Oertzen, T., McArdle, J. J., & Lindenberger, U. (2013). Structural equation model trees. *Psychological Methods, 18*(1), 71–86. https://doi.org/10.1037/a0030001
- Grimm, K. J., Mazza, G. L., & Davoudzadeh, P. (2017). Model selection in finite mixture models: A *k*-fold cross-validation approach. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(2), 246-256. https://doi.org/10.1080/10705511.2016.1250638
- Hong, M., Jacobucci, R., & Lubke, G. (2020). Deductive data mining. *Psychological Methods, 25*(6), 691–707. https://doi.org/10.1037/met0000252
- Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A practical guide to variable selection in Structural Equation Modeling by using regularized multiple-indicators, multiple-causes models. *Advances in Methods and Practices in Psychological Science, 2*(1), 55-76. https://doi.org/10.1177/2515245919826527
- Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science, 15*(3), 809-816. https://doi.org/10.1177/1745691620902467
- Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(4), 555-566. https://doi.org/10.1080/10705511.2016.1154793
- Jacobucci, R., Grimm, K. J., & Zhang, Z. (2023). *Machine learning for social and behavioral research*. Guilford.
- Landers, R. N. (n.d.). *Data science for social scientists*. https://datascience.tntlab.org/

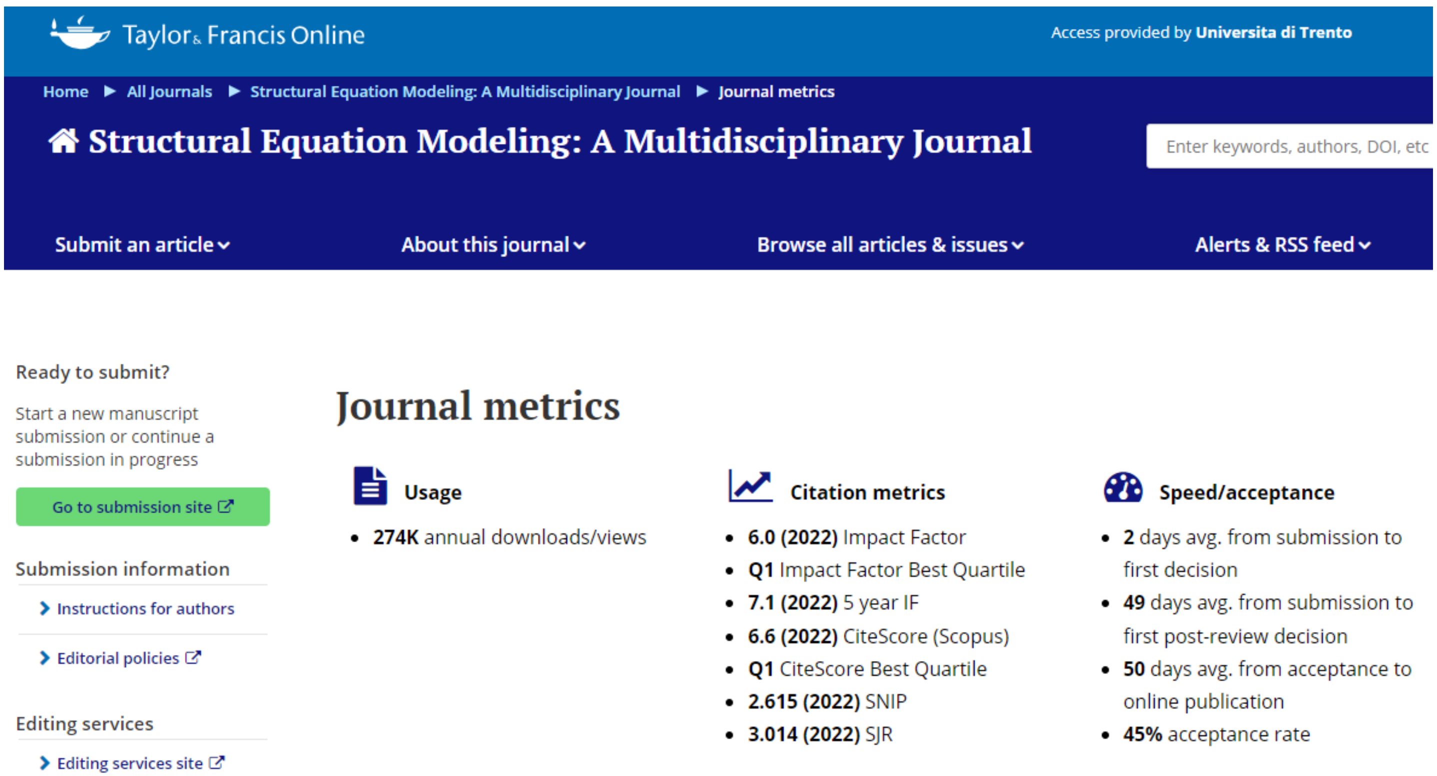
Suggested Reading

Machine Learning and Latent Variables/Psychometrics (*Part 2*)

- Liang, X., & Jacobucci, R. (2020). Regularized structural equation modeling to detect measurement bias: Evaluation of lasso, adaptive lasso, and elastic net. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(5), 722-734. https://doi.org/10.1080/10705511.2019.1693273
- Liem, C. C. S., Langer, M., Demetriou, A., Hiemstra, A. M. F., Wicaksana, A. S., Born, M. P., & König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In H. J. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlütürk, U. Güçlü, & M. van Gerven (Eds.), *Explainable and interpretable models in computer vision and machine learning* (pp. 197-253). Springer. https://doi.org/10.1007/978-3-319-98131-4_9
- McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate Behavioral Research, 50*(5), 471-484. https://doi.org/10.1080/00273171.2015.1036965
- Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine learning in psychometrics and psychological research. *Frontiers in Psychology, 10*, Article 2970. https://doi.org/10.3389/fpsyg.2019.02970
- Serang, S., Jacobucci, R., Brimhall, K. C., & Grimm, K. J. (2017). Exploratory mediation analysis via regularization. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(5), 733-744. https://doi.org/10.1080/10705511.2017.1311775
- Urban, C. J., & Gates, K. M. (2021). Deep learning: A primer for psychologists. *Psychological Methods, 26*(6), 743–773. https://doi.org/10.1037/met0000374
- Woo, S. E., Tay, L., & Proctor, R. W. (2020). *Big data in psychological research*. American Psychological Association. https://doi.org/10.1037/0000193-000
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

Suggested Reading

A whole journal devoted to the advancement of SEM (since 1994)



Taylor & Francis Online

Access provided by **Universita di Trento**

Home ▶ All Journals ▶ Structural Equation Modeling: A Multidisciplinary Journal ▶ Journal metrics

Structural Equation Modeling: A Multidisciplinary Journal

Enter keywords, authors, DOI, etc

Submit an article | About this journal | Browse all articles & issues | Alerts & RSS feed

Ready to submit?

Start a new manuscript submission or continue a submission in progress

Go to submission site

Submission information

- Instructions for authors
- Editorial policies

Editing services

- Editing services site

Journal metrics

Usage

- **274K** annual downloads/views

Citation metrics

- **6.0 (2022)** Impact Factor
- **Q1** Impact Factor Best Quartile
- **7.1 (2022)** 5 year IF
- **6.6 (2022)** CiteScore (Scopus)
- **Q1** CiteScore Best Quartile
- **2.615 (2022)** SNIP
- **3.014 (2022)** SJR

Speed/acceptance

- **2** days avg. from submission to first decision
- **49** days avg. from submission to first post-review decision
- **50** days avg. from acceptance to online publication
- **45%** acceptance rate

Sanjay Srivastava @hardsci · 8 ott

And those “latent variables,” are they in the room with us right now?

71 1.272 9.060

Thanks for your attention

enrico.perinelli@unitn.it

UNIVERSITÀ DI TRENTO
Dipartimento di Psicologia e Scienze Cognitive