UNIVERSITÀ DI TRENTO

well being at work LAB

UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Psicologia e Scienze Cognitive

Measurement Invariance with Structural Equation Modeling

Enrico Perinelli

Department of Psychology and Cognitive Science, University of Trento

PSICOSTAT 3.3 Meeting

June 17, 2022

Contents

Introduction

- Measurement Invariance (MI) is a statistical property of a measurement tool

- MI indicates whether the instrument measures the same construct…
  - …across different **groups** and/or…
  - …over **time**

- Measurement invariance should be (at)tested before **comparing groups** in a specific variable, as well as before running **longitudinal analyses** with the same construct(s) assessed over time

- Despite it is a long-standing topic (e.g., Meredith, 1964, *Psychometrika*), still several research assumes that instruments are invariant "by default"

Introduction

Two botanists see two plants. One plant is on the south side of a hill; it is tall, has broad leaves, and a slender stem. The other is on the north side of a hill; it is short, thin leaved, and has a thick stem. Although these outward characteristics and differences are easily observed, they wonder whether the two plants are the same species or not. **If the plants are the same species, they argue, then the different growth patterns have been shaped by the context of being on the north or south side of a hill. Otherwise, the differences occur because they are different species of plant**.

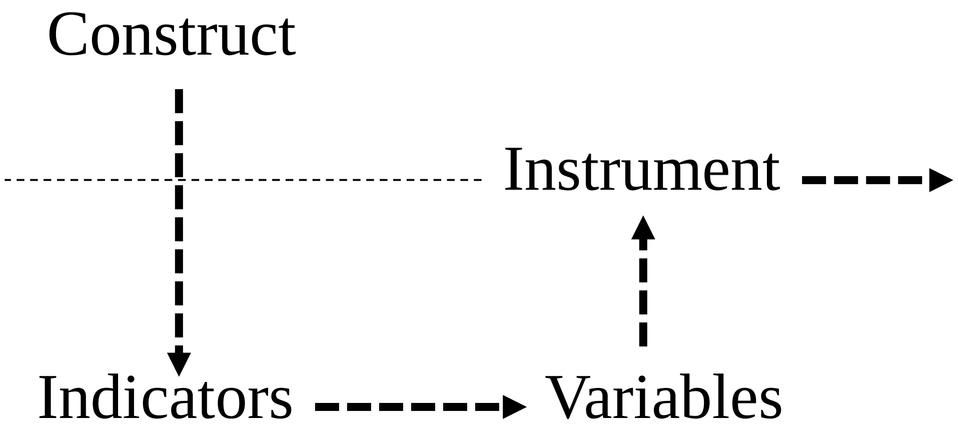
To test their conjectures, **they carefully dig down to the roots of the plants**. If each plant is the same species, it should have a signature root system. In this case, they see that the first plant has three roots, which follow a particular pattern of length and thickness: the first is medium long and medium thick, the second is longest and thickest, and the third is shortest and thinnest. They carefully dig up the other plant and see that it, too, has three roots that follow the same pattern of relative length and thickness, except that, for this plant, they see that, like the plant itself, all the roots are longer and thinner by about the same proportions.

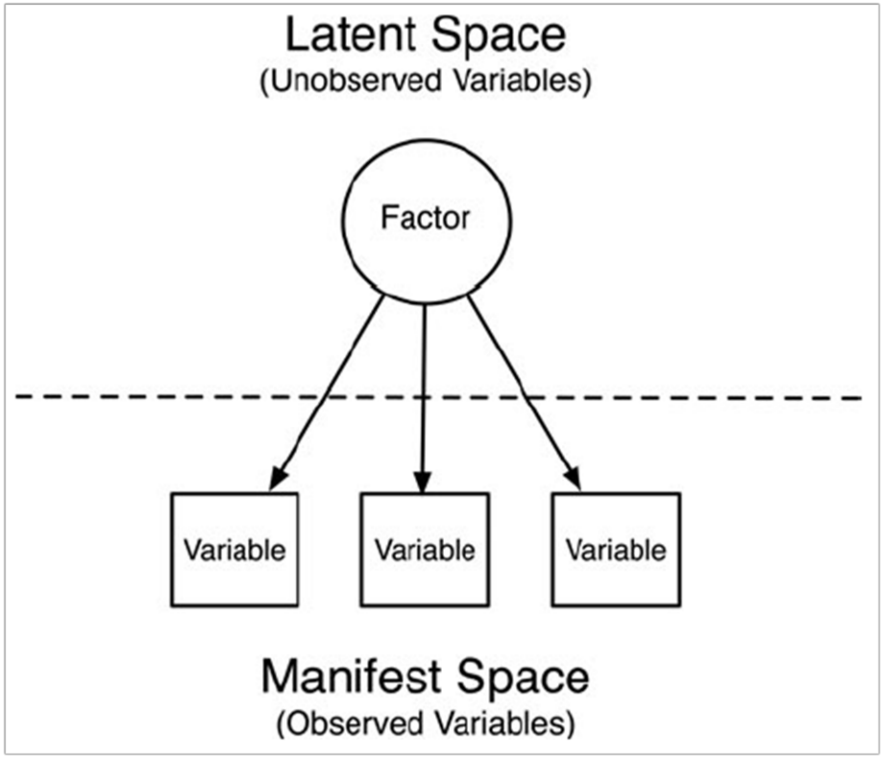
Because both plants have the same number of roots that follow the same pattern of length (**loadings**) and thickness (**intercepts**), they conclude that the plants (**constructs**) are fundamentally the same species (**factorially invariant**) and that the observed differences (**cross-time differences or group differences**) are due to the context. They also notice that the amounts of dirt and spindly bits still on the roots appear to be about the same, but they ignore this information because it's just dirt and spindly bits (**residuals**).

Little (2013, p. 137-138; *Longitudinal Structural Equation Modeling*, Guilford)

Introduction

Operationalization Process





Screenshot from Nesselroade & Molenaar (2016)

- Are you sure that the link between latent variable and its indicators is the same across groups and/or time??

Introduction

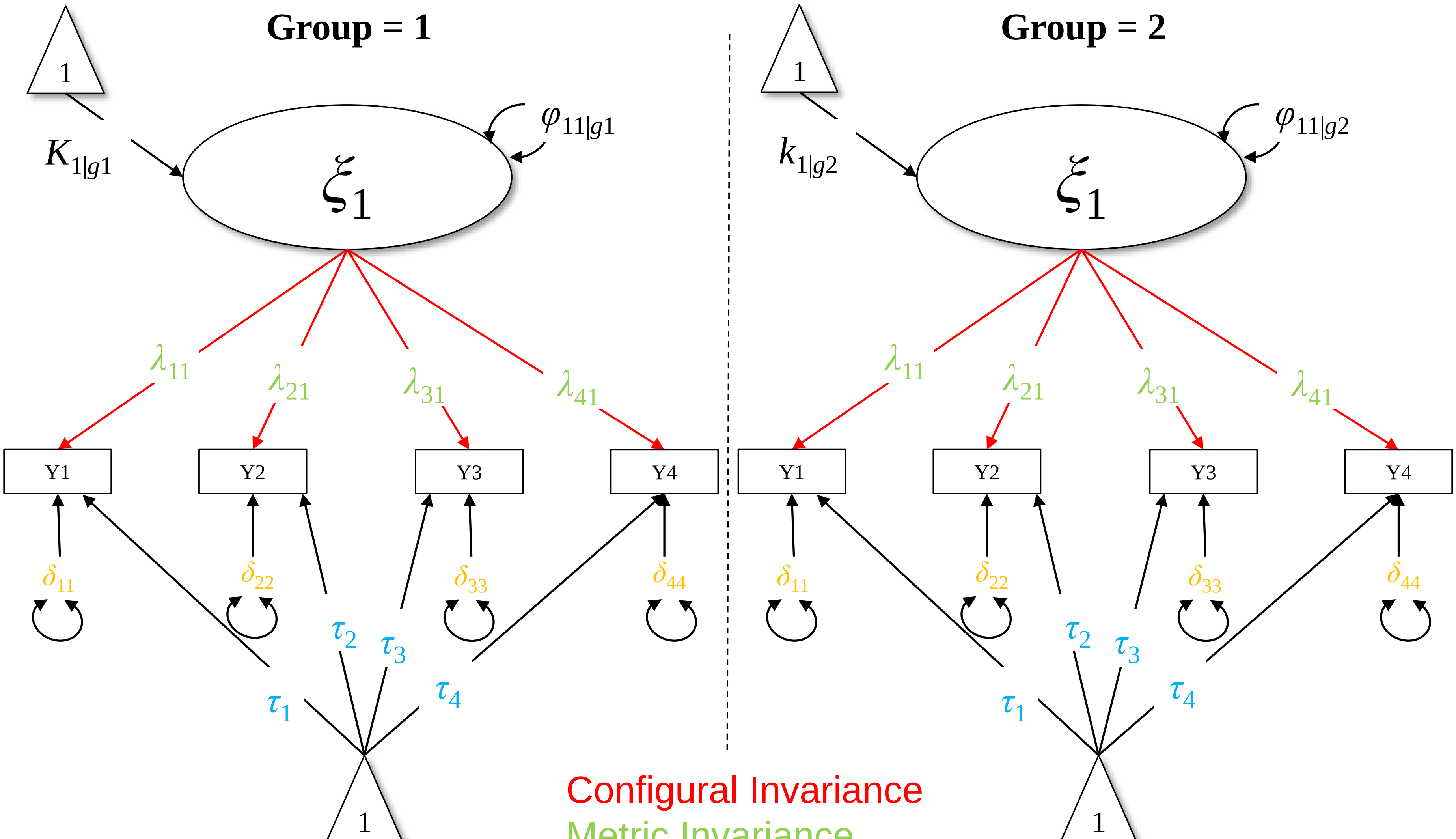
- Thus, MI is a fundamental aspect of Internal/Structural Validity of a measure

  - **Qualitative Validity**
    - Face Validity
    - Content Validity
  - **Internal (or Structural) Validity**
    - Reliability
    - Measurement Invariance
  - **External Validity**
    - Construct
    - Criterion
    - Convergent
    - Divergent
    - Discriminant

MI with Continuous Indicators

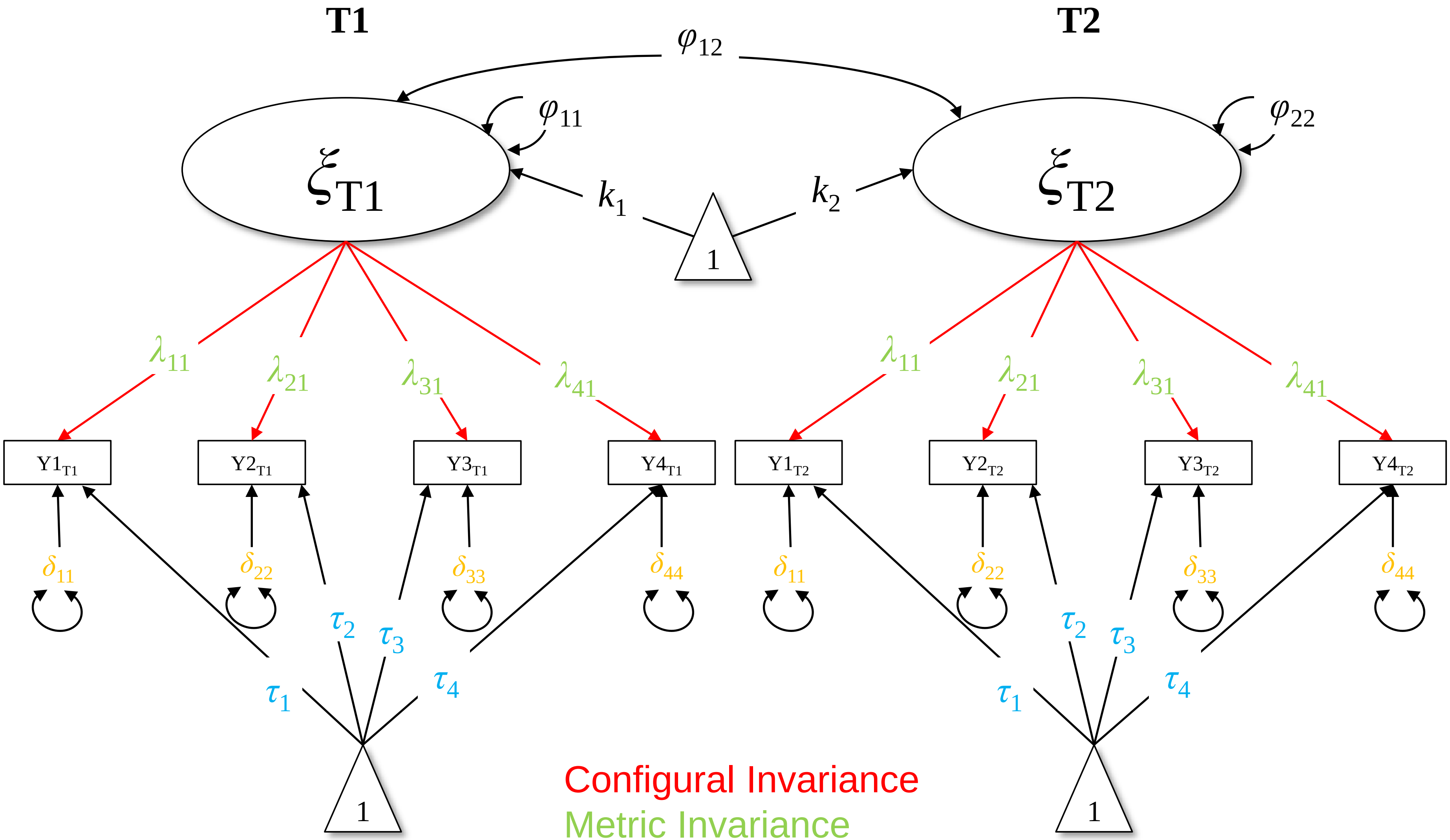
- MI is tested by comparing increasingly constrained models
  - **Configural Invariance** -> the structure of the latent variable(s) is the same across groups ($g \neq g'$) and/or over time ($\xi_g = \xi_{g'}$)
  - **Metric Invariance** (also called Weak Invariance) -> factor loadings are equivalent across groups and/or over time ($\boldsymbol{\Lambda}_g = \boldsymbol{\Lambda}_{g'}$)
  - **Scalar Invariance** (also called Strong Invariance) –> intercepts of observed variables are equivalent across groups and/or over time ($\boldsymbol{\tau}_g = \boldsymbol{\tau}_{g'}$)
  - **Strict Invariance** (also called Residual or Invariant Uniqueness Invariance) -> residual variances of observed exogenous variables are equivalent across groups and/or over time ($\boldsymbol{\Theta}_{\delta,g} = \boldsymbol{\Theta}_{\delta,g'}$)
- Scalar/Strong invariance is requested to compare means

MI with Continuous Indicators – Multiple-group



Configural Invariance
Metric Invariance
Scalar Invariance
Strict Invariance

MI with Continuous Indicators – Longitudinal



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

MI with Continuous Indicators – Reporting results

Table S2

Longitudinal Measurement Invariance Routine for Effortful Control (5 Time Points)

Informant	Model	$\chi^2(df)$	CFI	TLI	RMSEA	AIC	$\Delta\chi^2(\Delta df)$	ΔCFI
P	Configural	87.149*(60)	0.985	0.973	0.049	4,362.84	-	-
	Metric	100.847**(68)	0.981	0.971	0.051	4,360.54	13.698(8)$^{+}$	.004
	Scalar	**117.468**(76)**	**0.977**	**0.968**	**0.054**	**4,361.16**	**16.621(8)***	**.004**
T	Configural	85.014**(55)	0.975	0.953	0.055	4,105.863	-	-
	Metric	102.259**(63)	0.968	0.946	0.058	4,107.108	17.245*(8)	.007
	Scalar	122.357***(71)	0.958	0.937	0.063	4,111.207	20.098*(8)	.010
	Scalar partial	**114.139***(70)**	**0.963**	**0.945**	**0.059**	**4,104.988**	**11.880$^{n.s.}$(7)**	**.005**

Note. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; AIC = Akaike's Information Criterion; P = Parent; T = Teacher.
In bold, the level of invariance obtained.
$^{n.s.}p > .10$. $^{+}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$

Invariance step is passed if ΔCFI < .10; otherwise «partial measurement invariance»

MI with Continuous Indicators – Reporting results

Table S2

Longitudinal Measurement Invariance Routine for Effortful Control (5 Time Points)

Informant	Model	$\chi^2(df)$	CFI	TLI	RMSEA	AIC	$\Delta\chi^2(\Delta df)$	ΔCFI
P	Configural	$87.149^{*}(60)$	0.985	0.973	0.049	4,362.84	-	-
	Metric	$100.847^{**}(68)$	0.981	0.971	0.051	4,360.54	$13.698(8)^{+}$	.004
	Scalar	$\mathbf{117.468^{**}(76)}$	**0.977**	**0.968**	**0.054**	**4,361.16**	$\mathbf{16.621(8)^{*}}$	**.004**
T	Configural	$85.014^{**}(55)$	0.975	0.953	0.055	4,105.863	-	-
	Metric	$102.259^{**}(63)$	0.968	0.946	0.058	4,107.108	$17.245^{*}(8)$	.007
	Scalar	$122.357^{***}(71)$	0.958	0.937	0.063	4,111.207	$20.098^{*}(8)$	.010
	Scalar partial	$\mathbf{114.139^{***}(70)}$	**0.963**	**0.945**	**0.059**	**4,104.988**	$\mathbf{11.880^{n.s.}(7)}$	**.005**

Note. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; AIC = Akaike's Information Criterion; P = Parent; T = Teacher.
In bold, the level of invariance obtained.
$^{n.s.}p > .10$. $^{+}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$

Alessandri, G., Perinelli, E., Filosa, L., Eisenberg, N., & Valiente, C. (in press). The validity of the higher-order structure of effortful control as defined by inhibitory control, attention shifting, and focusing: A longitudinal and multi-informant study. *Journal of Personality*. https://doi.org/10.1111/jopy.12696 [**Table S2 is in Supplementary Material**]

OSF (M*plus* and R scripts)
https://osf.io/j8cfv/?view_only=94794d1aaf2e41c0990287833ec04594

MI with Continuous Indicators – Multiple-group X Time

Table A1
Gender*Longitudinal Measurement Invariance.

Self-Concept	Gen.Inv.	Long.Inv.	NFP	YBχ^2	*df*	*p*	SCF	CFI	TLI	RMSEA	SRMR	CD	SB$\Delta\chi^2$	Δdf	*p*	ΔCFI
Math	configural	configural	62	45.150	26	0.011	1.1880	0.997	0.993	0.030	0.016					
	metric	configural	56	55.236	32	0.007	1.1009	0.996	0.993	0.030	0.025	0.723	9.912	6	0.128	0.001
	scalar	configural	50	74.157	38	< 0.001	1.1071	0.994	0.991	0.034	0.029	1.140	18.873	6	0.005	0.002
	strict	configural	42	87.701	46	< 0.001	1.0729	0.993	0.992	0.033	0.030	0.910	13.175	8	0.106	0.001
	strict	metric	39	93.600	49	< 0.001	1.0630	0.993	0.992	0.033	0.033	0.911	5.929	3	0.115	0
	strict	scalar	36	120.583	52	< 0.001	1.0654	0.989	0.988	0.040	0.038	1.105	26.229	3	< 0.001	0.004
	strict	**strict**	**32**	**125.644**	**56**	**< 0.001**	**1.0836**	**0.988**	**0.988**	**0.039**	**0.038**	**1.320**	**5.816**	**4**	**0.213**	**0.001**
Verbal	configural	configural	62	56.422	26	< 0.001	1.0674	0.992	0.983	0.037	0.024					
	metric	configural	56	58.367	32	0.003	1.0665	0.993	0.988	0.031	0.026	1.063	1.904	6	0.928	-0.001
	scalar	configural	50	71.194	38	< 0.001	1.0718	0.992	0.988	0.032	0.030	1.100	12.779	6	0.047	0.001
	strict	configural	42	105.586	46	< 0.001	1.0866	0.985	0.982	0.039	0.047	1.157	33.213	8	< 0.001	0.007
	strict	metric	39	108.375	49	< 0.001	1.0791	0.985	0.983	0.038	0.048	0.964	2.300	3	0.512	0
	strict	scalar	36	111.746	52	< 0.001	1.0823	0.985	0.984	0.037	0.047	1.135	3.521	3	0.318	0
	strict	**strict**	**32**	**127.500**	**56**	**< 0.001**	**1.0773**	**0.982**	**0.982**	**0.039**	**0.053**	**1.012**	**16.214**	**4**	**0.003**	**0.003**

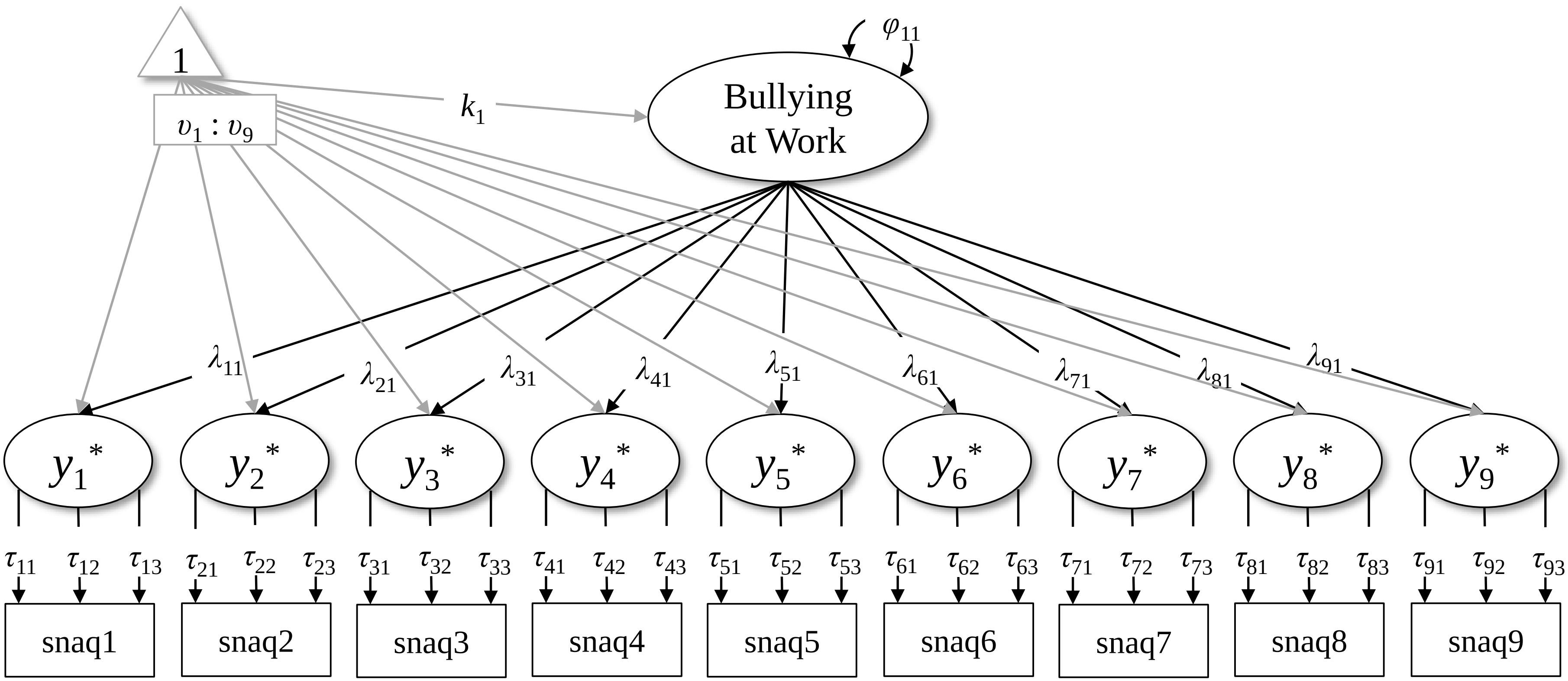
Note. Estimation method: MLR + Cluster robust-standard errors. Gen.Inv. = Gender Invariance Step; Long.Inv. = Longitudinal Invariance Step; NFP = Number of Free Parameters; YBχ^2 = Yuan-Bentler scaled chi-square; *df* = degrees of freedom; SCF = Scaling Correction Factor; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; CD = Difference Test Scaling Correction; SB$\Delta\chi^2$ = Satorra-Bentler scaled chi-square difference; Δdf = difference in degrees of freedom; ΔCFI = difference in CFI.
The level of measurement invariance obtained by the instrument is reported in bold.

Perinelli, E., Pisanu, F., Checchi, D., Scalas, L. F., & Fraccaroli, F. (2022). Academic self-concept change in junior high school students and relationships with academic achievement. *Contemporary Educational Psychology, 69*, Article 102071. https://doi.org/10.1016/j.cedpsych.2022.102071

MI with Ordinal Variables

- With ordinal variables the routine change
- First of all, pay attention that all the observed variables across groups hold the same categories; otherwise, a solution is to collapse two categories in one category (DiStefano, Shi, & Morgan, 2021; but see Rutkowski, Svetina, & Liaw, 2019).
- Second, the steps of invariance are different than the classical one
  - **Baseline Model** -> Similar to configural invariance for continuous variables
  - **Equal Threshold Model** -> Fix the thresholds to be equal across groups and/or over time
  - **Equal Thresholds and Loadings Model** –> Only now you can fix loadings to equality
- Suggested reading for M*plus* and R script: Svetina, Rutkowski, & Rutkowski (2020, *SEM*)

MI with Ordinal Variables



Note. Asterisks indicate latent continuous variables assumed to underlie the observed categorical indicators. Given that we collapsed category 5 within category 4 (due to cells imbalance across gender), each item has 3 thresholds. Residual variance of each *y** is fixed to be zero. Paths for mean-level structure are reported in grey for sake of clarity.

Perinelli, Balducci, & Fraccaroli (under review); a poster version is available at http://dx.doi.org/10.13140/RG.2.2.26701.10720

MI with Ordinal Variables

Table 3

Measurement Invariance Results

Software	Model	WLSMV-based χ^2	*df*	*p*	CFI	TLI	RMSEA	Scaled-$\Delta\chi^2$	Δdf	*p*	ΔCFI	ΔRMSEA
R (lavaan)	Baseline	150.014	54	< .001	.970	.960	.100	-			-	-
	Equal Thresholds	165.795	63	< .001	.968	.963	.096	13.514	9	.141	.002	-.004
	Equal Thresholds and Loadings	158.085	71	< .001	.973	.972	.083	5.140	8	.743	-.005	-.013
Mplus	Baseline	150.791	54	< .001	.970	.960	.100	-	-	-	-	-
	Equal Thresholds	165.590	63	< .001	.968	.964	.096	13.933	9	.125	.002	-.004
	Equal Thresholds and Loadings	152.627	71	< .001	.975	.974	.080	5.049	8	.752	-.007	-.016

Note. WLSMV = Weighted Least Squares Mean- and Variance-adjusted; *df* = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; ΔCFI = decrease in CFI; ΔRMSEA = increase in RMSEA.

Perinelli, Balducci, & Fraccaroli (under review); a poster version is available at http://dx.doi.org/10.13140/RG.2.2.26701.10720

MI with Ordinal Variables

Table 4

Parameters of interest from the 'Equal Thresholds and Loadings Model' estimated in lavaan

Unconstrained Parameters				Constrained Parameters		
Parameter		Male Model	Female Model	Parameter		Male/Female Model
Intercepts	υ_1	0	0.042	Factor Loadings	λ_{11}	0.794
	υ_2	0	-0.114		λ_{21}	0.875
	υ_3	0	0.216		λ_{31}	0.789
	υ_4	0	-0.099		λ_{41}	0.924
	υ_5	0	0.183		λ_{51}	0.671
	υ_6	0	-0.106		λ_{61}	0.530
	υ_7	0	0.156		λ_{71}	0.765
	υ_8	0	0.078		λ_{81}	0.805
	υ_9	0	-0.026		λ_{91}	0.830
Explained Variance	R^2_{snaq1}	0.630	0.538	Thresholds	$\tau_{11}\ \tau_{12}\ \tau_{13}$	-0.092, 0.822, 1.139
	R^2_{snaq2}	0.765	0.597		$\tau_{21}\ \tau_{22}\ \tau_{23}$	0.083, 1.107, 1.403
	R^2_{snaq3}	0.622	0.619		$\tau_{31}\ \tau_{32}\ \tau_{33}$	0.254, 1.166, 1.444
	R^2_{snaq4}	0.854	0.738		$\tau_{41}\ \tau_{42}\ \tau_{43}$	0.552, 1.309, 1.756
	R^2_{snaq5}	0.450	0.489		$\tau_{51}\ \tau_{52}\ \tau_{53}$	0.249, 1.242, 1.675
	R^2_{snaq6}	0.281	0.366		$\tau_{61}\ \tau_{62}\ \tau_{63}$	-0.749, 0.700, 1.145
	R^2_{snaq7}	0.586	0.728		$\tau_{71}\ \tau_{72}\ \tau_{73}$	0.093, 1.002, 1.388
	R^2_{snaq8}	0.649	0.623		$\tau_{81}\ \tau_{82}\ \tau_{83}$	0.231, 1.204, 1.591
	R^2_{snaq9}	0.689	0.549		$\tau_{91}\ \tau_{92}\ \tau_{93}$	1.245, 1.900, 2.082
Latent Variance	φ_{11}	1	0.648	Latent Mean	κ_1	0

Note. Parameters are reported in unstandardized form

Perinelli, Balducci, & Fraccaroli (under review); a poster version is available at http://dx.doi.org/10.13140/RG.2.2.26701.10720

MI with Ordinal Variables

- In the Appendix file (`MI_Ordinal.html`) you can find more information, such as
  - Kendall rank correlations (in APA style)
  - Reliability for ordinal data
  - R (`lavaan`) scripts for the Baseline, Equal Thresholds, Equal Thresholds and Loadings models
  - Summary of Results

Measurement Invariance with Ordinal Data

PSICOSTAT 3.3 Meeting

Enrico Perinelli
Department of Psychology and Cognitive Science, University of Trento

June 17, 2022

Introduction

This document is an Appendix of the main presentation **Measurement Invariance with Structural Equation Modeling**. In particular, this Appendix may be useful for the interpretation of the section ***MI with Ordinal Variables***

The items refer to the Italian version of the *Short Negative Acts Questionnaire* (SNAQ; Balducci et al., 2010; Notelaers et al., 2019)

It is important to specify that (due to the imbalance of the cells) two categories (4 and 5) were collapsed to one category (i.e., 4). Below, you can find the syntax with the relative item content (note that the value of 5 is collapsed in the value of 4)[1].

```
# Proseguiamo l'intervista chiedendole di indicare con che frequenza, negli ultimi sei mesi, ha subito ciascuno d
ei seguenti comportamenti nel suo luogo di lavoro
# 1 = mai, 2 = Una volta in tutto o di tanto in tanto, 3 = mensilmente, 4 = Settimanalmente, 5 = Quotidianamente

dat <- mobbing_data2 %>%  mutate(
  snaq1=replace(snaq1, snaq1==5, 4),
  # Le sono state nascoste informazioni che influenzano il suo lavoro
  snaq2=replace(snaq2, snaq2==5, 4),
```

Other topics

- Measurement invariance with second-order factors (Chen, Sousa, & West, 2005)
- Measurement invariance with Exploratory SEM (ESEM; Marsh, Morin, Parker, & Kaur, 2014)
- Approximate Measurement Invariance (this is a Bayesian approach; Muthén & Asparouhov, 2012, *Psychological Methods*; Van De Schoot, Kluytmans, Tummers, Lugtig, Hox, & Muthén, 2013)
- Alignment-within-CFA (AwC) approach (Marsh, Guo, Parker, Nagengast, Asparouhov, Muthén, & Dicke, 2018)
- What if groups are unbalanced? See Yoon and Lai (2018)
- `lavaan` tutorial on MI https://lavaan.ugent.be/tutorial/groups.html

Conclusion and food for thought

- Measurement "quality" is a pivotal requirement for attesting the goodness of the research, in particular in **non-experimental research**
- The increasing interest in Big Data, Machine Learning, and Data-driven approaches should not bring researchers to neglect the **internal validity of the instruments** we use (see the works by Ross Jacobucci, in particular Jacobucci & Grimm, 2020, Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*)
- Measurement invariance is a structural validity analysis that may give to psychometrics an important place even in fields of applications outside "classical" statistics (e.g., in **statistical learning**)
- See Bleidorn, Hopwood, & Wright (2017) for some examples about the importance of validity (and implicitly of MI) of digital footprints

UNIVERSITÀ DI TRENTO
well being at work LAB

UNIVERSITÀ DEGLI STUDI DI TRENTO
Dipartimento di Psicologia e Scienze Cognitive

Thanks for your attention

enrico.perinelli@unitn.it