Erasmus School of
Social and
Behavioural Sciences

Studying change: is it real or not?

On the importance of longitudinal invariance

Marleen de Moor
demoor@essb.eur.nl
February 22nd 2023

Section of Methods & Techniques
Department of Psychology, Education and Child Studies
Erasmus University Rotterdam

Erasmus University Rotterdam

Overview of this talk

- ❑ Studying change
- ❑ Testing for longitudinal measurement invariance
- ❑ Issues in testing for longitudinal measurement invariance
- ❑ Conclusions
- ❑ Further reading and practice

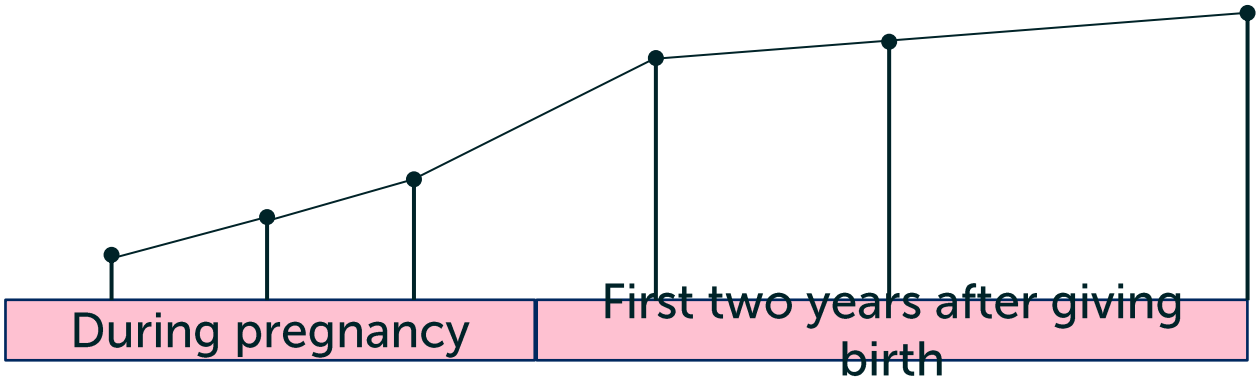
Overview of this talk

- ❑ Studying change
- ❑ Testing for longitudinal measurement invariance
- ❑ Issues in testing for longitudinal measurement invariance
- ❑ Conclusions
- ❑ Further reading and practice

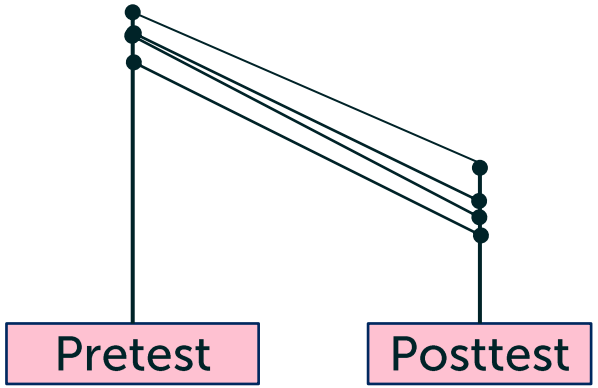
Research questions about change

- ❑ Longitudinal cohort studies:
  - How does child IQ develop over time?
  - How does parenting competence change across time?
- ❑ Randomized controlled trials:
  - What is the effect of psychotherapy on change in depressive symptoms?
  - What is the effect of nasal oxytocin administration on fathers' involvement?

Findings with regard to change



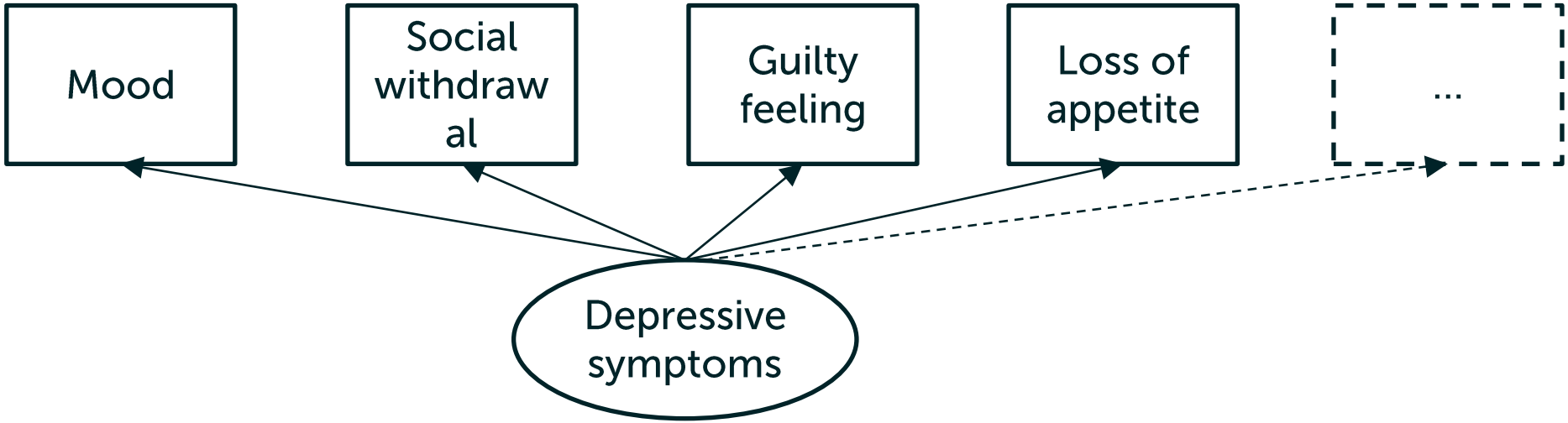
Parenting competence grows
(Verhage et al. 2013; Kunseler et al. 2014
Wernand et al. 2014)



Depressive symptoms decline in all groups,
No sign difference between groups
(Elkin et al. 1989)

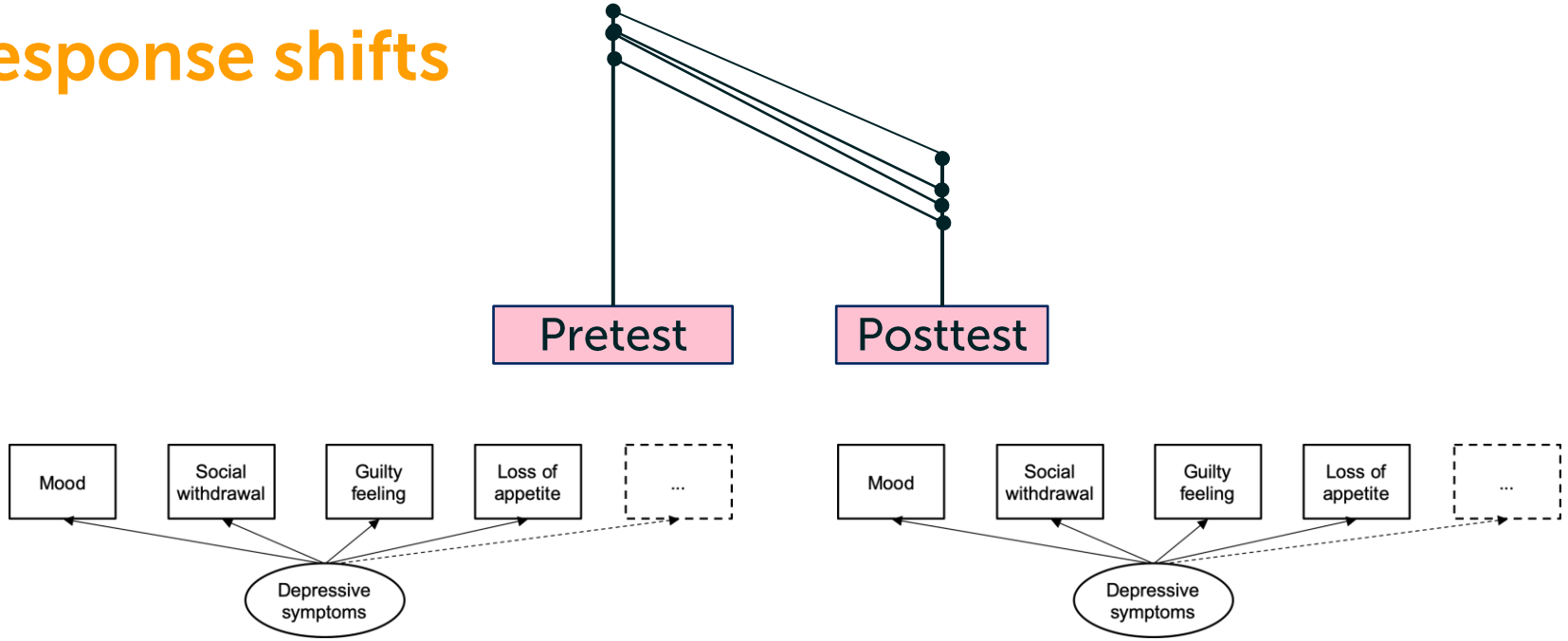
Measurement

- ❑ Direct versus indirect measurements
- ❑ Psychological constructs: difficult to observe and complex



- ❑ Measurement model
- ❑ Psychometric properties of scale

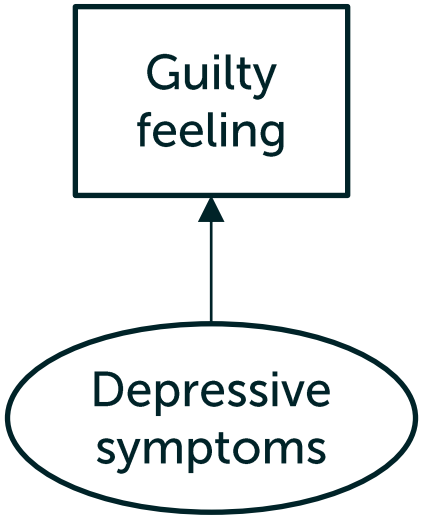
Response shifts



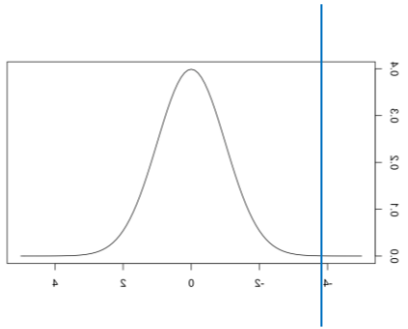
- Change in 'frame of reference'
  - As a result of the intervention
  - Or as a result of changes across development

Longitudinal measurement invariance

- Change in 'frame of reference'
  - As a result of the intervention
  - Or as a result of changes across development



0	I don't feel particularly guilty
1	I feel guilty a good part of the time.
2	I feel quite guilty most of the time.
3	I feel guilty all of the time.



Overview of this talk

- ☐ Studying change
- ☐ Testing for longitudinal measurement invariance
- ☐ Issues in testing for longitudinal measurement invariance
- ☐ Conclusions
- ☐ Further reading and practice

Confirmatory Factor Analysis

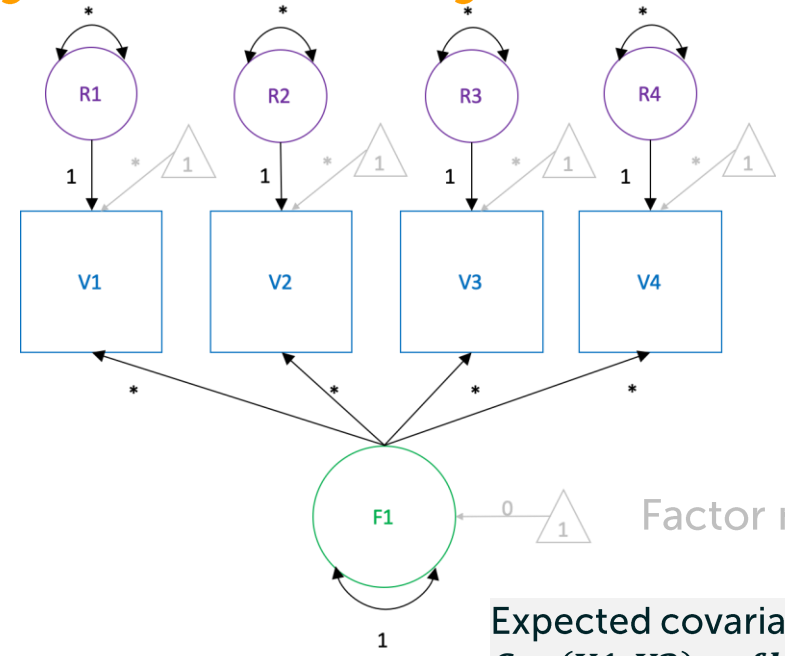
* Residual variances

Residual factors

Observed variables

* Factor loadings

Latent variable



* Intercepts

Factor mean

$$\boldsymbol{V1 = ic1 + fl1 * F1 + R1}$$
$$V2 = ic2 + fl2 * F1 + R2$$
$$V3 = ic3 + fl3 * F1 + R3$$
$$V4 = ic4 + fl4 * F1 + R4$$

Expected Variances:

$$\boldsymbol{Var(V1) = fl1^2 + Var(R1)}$$
$$Var(V2) = fl2^2 + Var(R2)$$
$$Var(V3) = fl3^2 + Var(R3)$$
$$Var(V4) = fl4^2 + Var(R4)$$

Expected covariances:

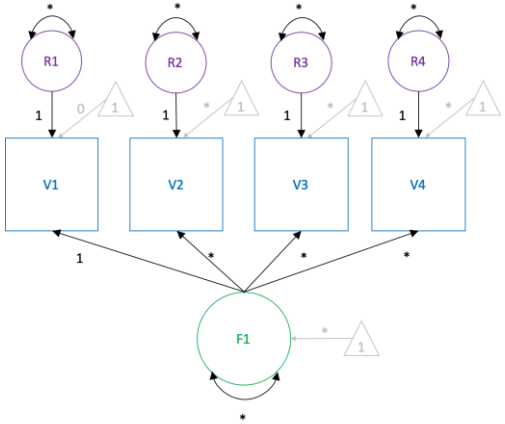
$$\boldsymbol{Cov(V1, V2) = fl1 * fl2}$$
$$Cov(V1, V3) = fl1 * fl3$$
$$Cov(V1, V4) = fl1 * fl4$$
$$Cov(V2, V3) = fl2 * fl3$$
$$Cov(V2, V4) = fl2 * fl4$$
$$Cov(V3, V4) = fl3 * fl4$$

Expected means:

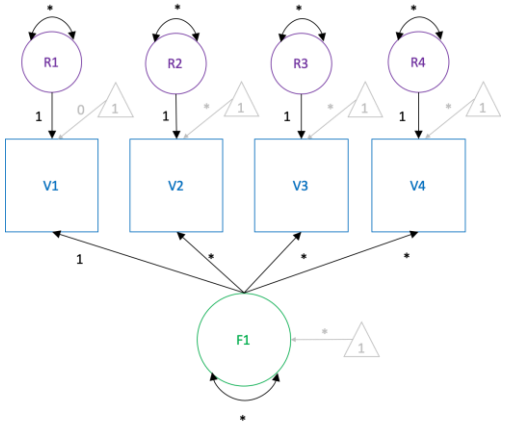
$$\boldsymbol{Mean(V1) = fl1 * 0 + ic1}$$
$$Mean(V2) = fl2 * 0 + ic2$$
$$Mean(V3) = fl3 * 0 + ic3$$
$$Mean(V4) = fl4 * 0 + ic4$$

Multigroup CFA

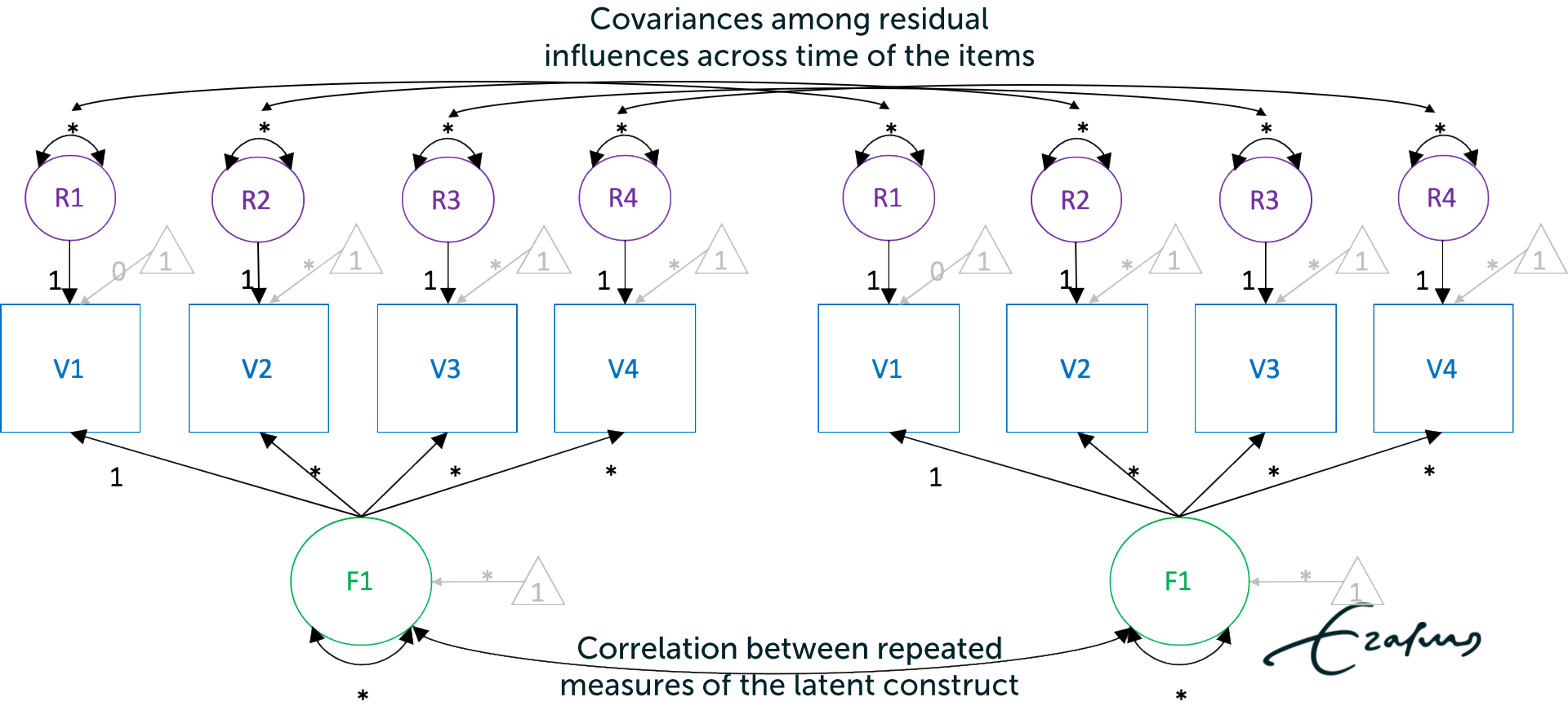
Group 1



Group 2



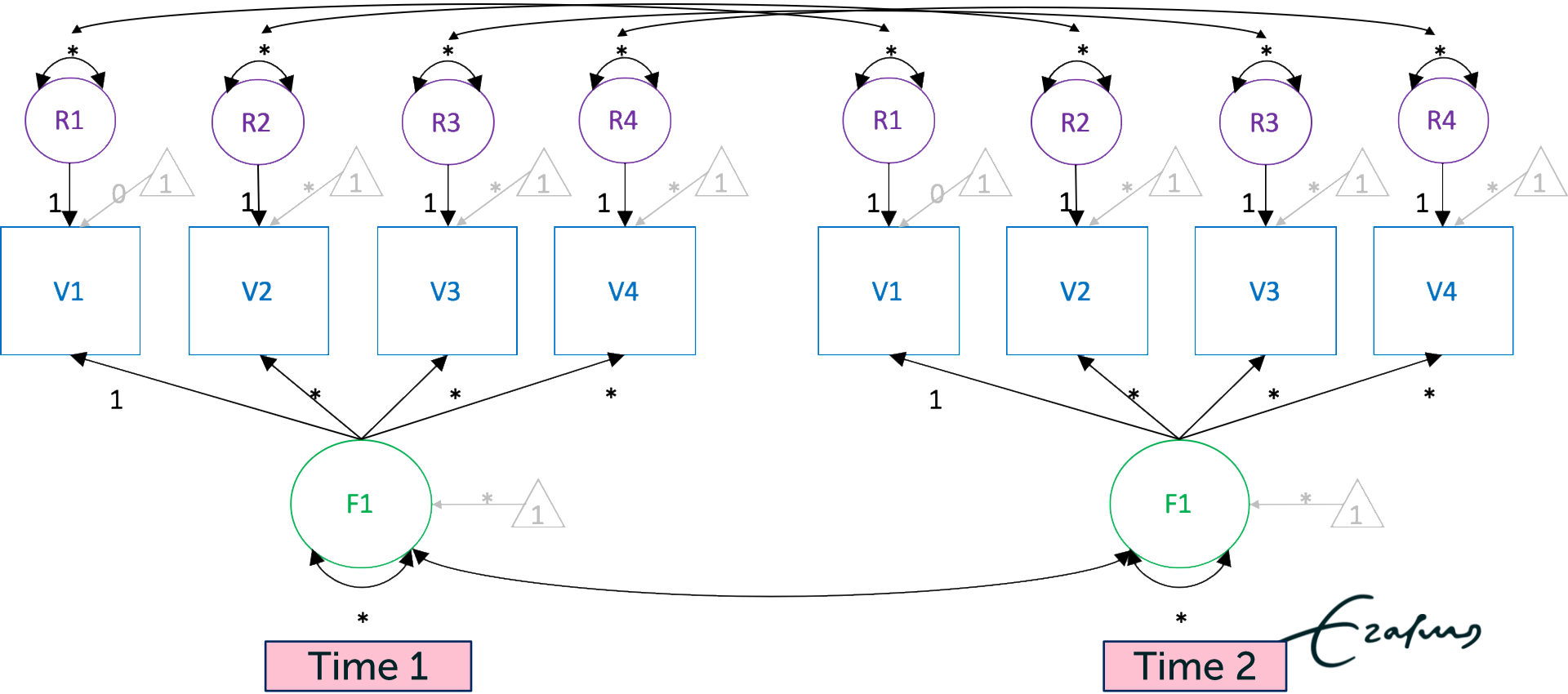
CFA for repeated measures



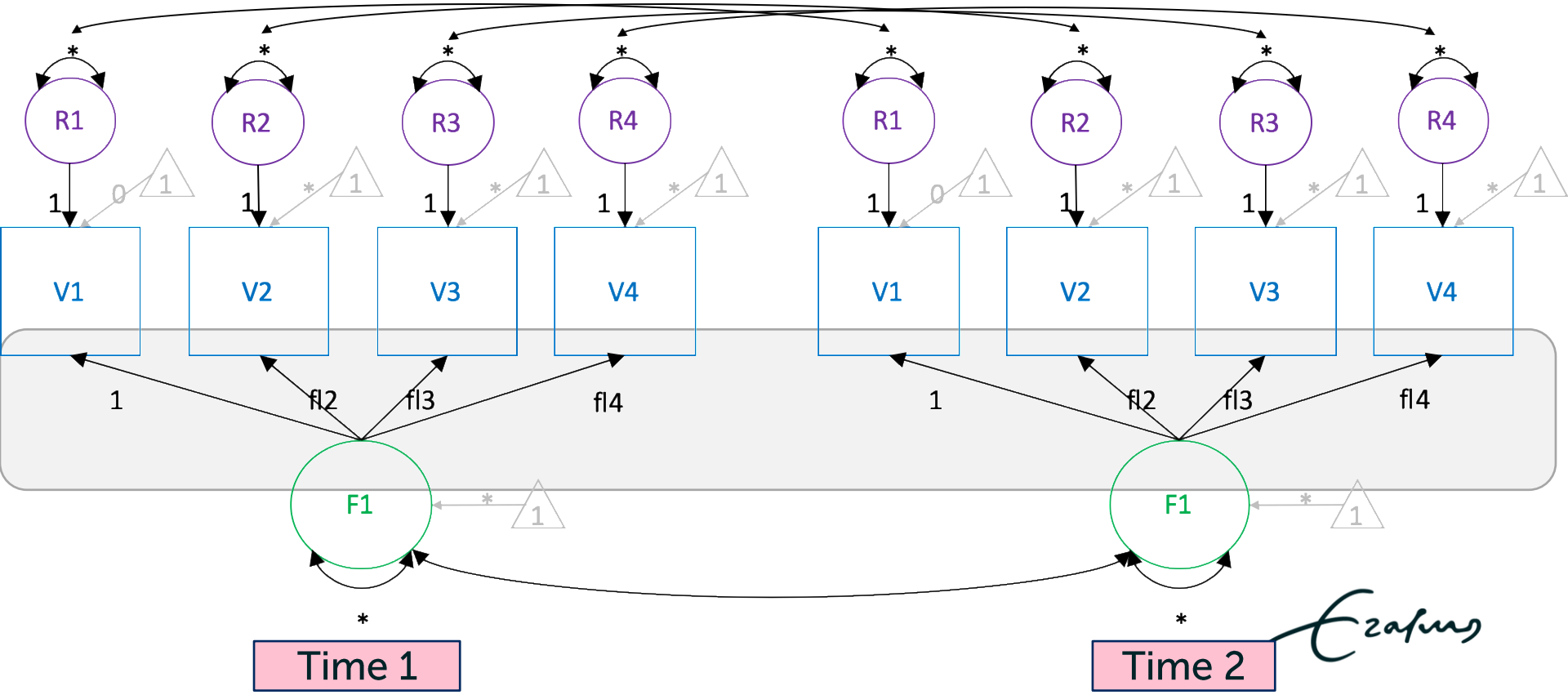
Levels of longitudinal MI

1. Configural invariance: does the same factor model hold across time?
2. Metric invariance: do the factor scores predict the responses on the observed variables equally well across time? (constrain factor loadings)
3. Strong/Scalar invariance: is any difference in response means across time the result of differences in factor means? (also constrain intercepts)
4. Strict invariance: are differences in observed scores across time only due to a true difference on the construct and not to any difference in measurement of the construct? (also constrain residual variances)

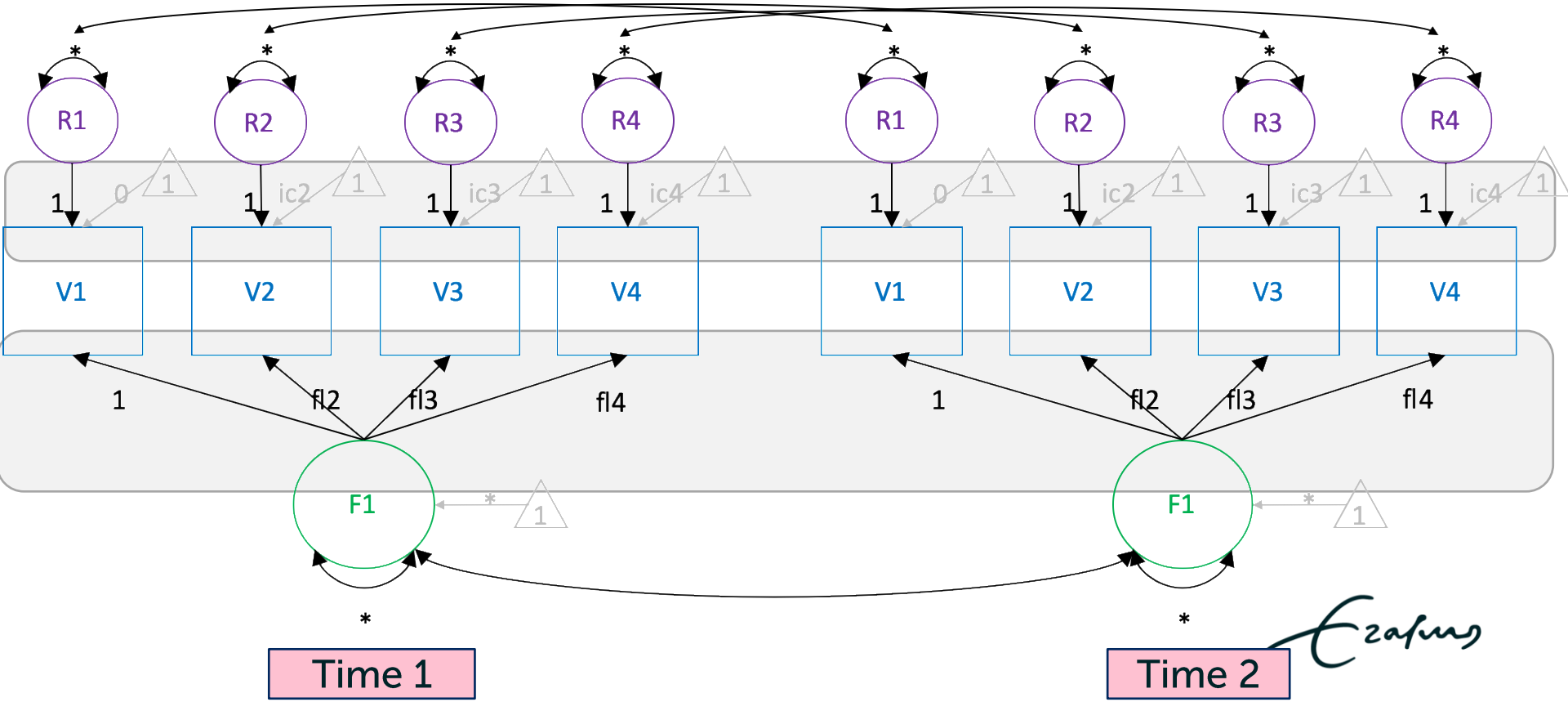
The configural invariance model



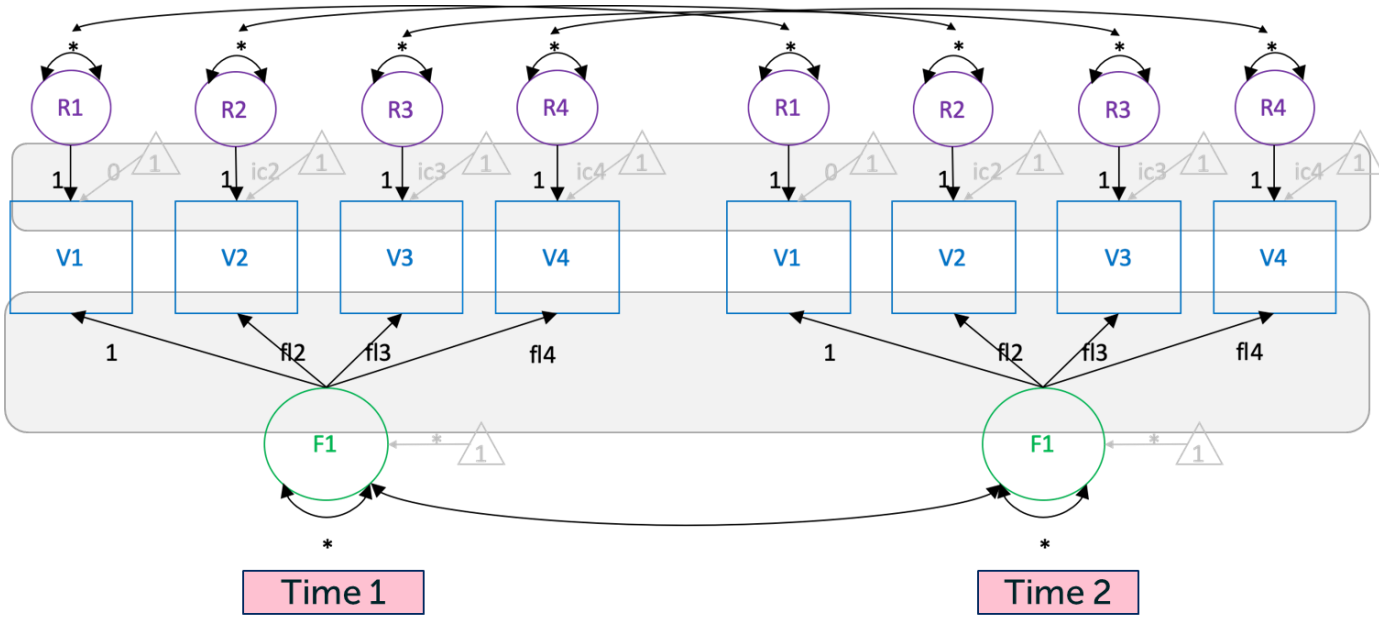
The metric invariance model



The scalar invariance model



The scalar invariance model



Expected means Time 1:

$Mean(V1) = 1 * fmt1 + 0$

$Mean(V2) = fl2 * fmt1 + ic2$

$Mean(V3) = fl3 * fmt1 + ic3$

$Mean(V4) = fl4 * fmt1 + ic4$

Expected means Time 2:

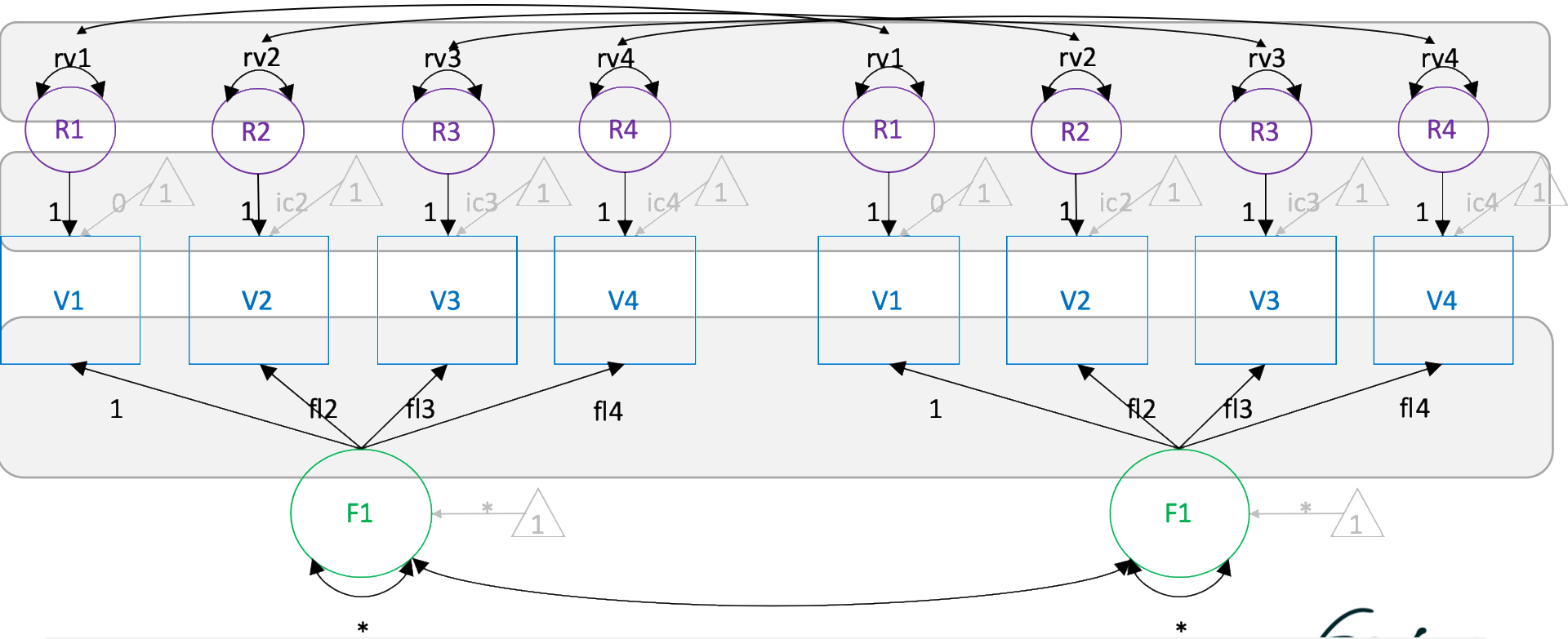
$Mean(V1) = 1 * fmt2 + 0$

$Mean(V2) = fl2 * fmt2 + ic2$

$Mean(V3) = fl3 * fmt2 + ic3$

$Mean(V4) = fl4 * fmt2 + ic4$

The strict invariance model



- Differences in observed means across time are only the result of differences in factor means
- Differences in observed (co)variances across time are only the result of differences in factor variances

Model fit and comparisons

- ❑ Model fit: standard model fit indices such as χ^2, RMSEA, CFI and TLI
- ❑ Model comparisons: $\Delta\chi^2$, ΔRMSEA (.03) and ΔCFI (.01)
  - (Cheung & Rensvold, 2002; Meade et al. 2008; Chen, 2007)

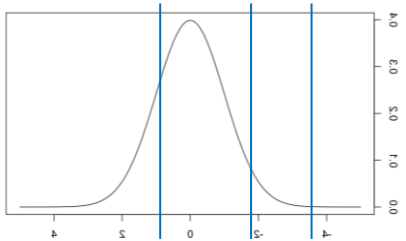
- Still a lot of debate:
  - Influence sample and model size
  - Cutoff points for good/bad fit
  - Equivalence testing more suited?

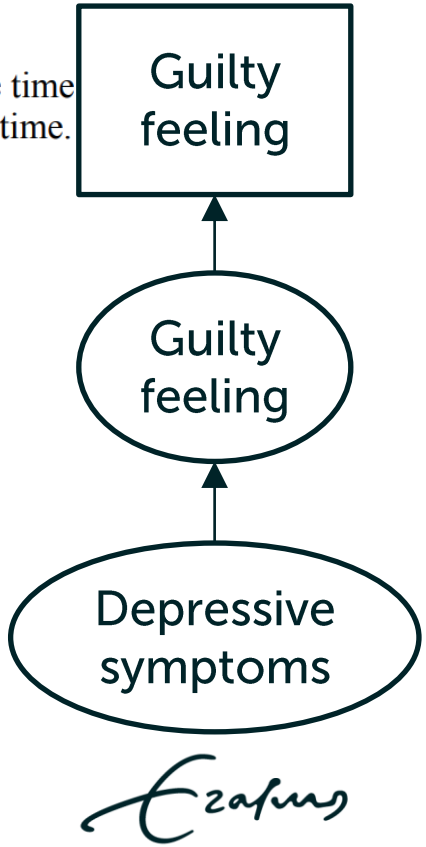
Continuous versus categorical indicators

- Continuous indicators:
  - Four levels of MI

0	I don't feel particularly guilty
1	I feel guilty a good part of the time
2	I feel quite guilty most of the time.
3	I feel guilty all of the time.

- Categorical indicators:
  - Threshold model is needed
  - Three levels of MI





An application: Parenting self-efficacy

Almost to be submitted

generaties²

VU VRIJE UNIVERSITEIT AMSTERDAM

During pregnancy | First two years after giving birth

Research questions

- How does Parenting Self-Efficacy develop in the transition to parenthood?

- Do changes in mean Parenting Self-Efficacy represent changes in true underlying scores?
- Do we measure the same underlying PSE construct prenatally versus postnatally?
- Does the experience of being a parent in the first two years of parenthood affect how mothers answer the scale to assess PSE?

The Self-Efficacy in the Nurturing Role (SENR) Scale

Answer categories from

1 (*not at all representative of me*)

to

7 (*strongly representative of me*)

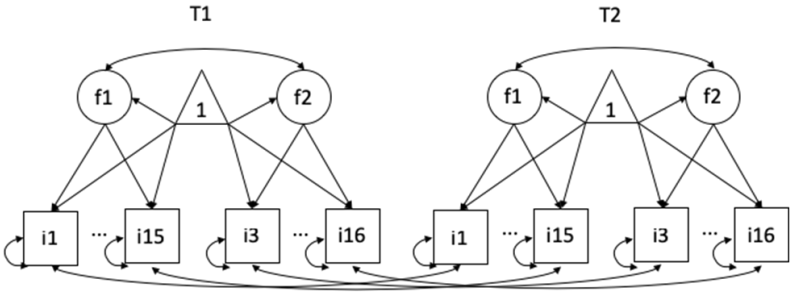
Slightly different wordings for

- Prenatal version
- Toddler version

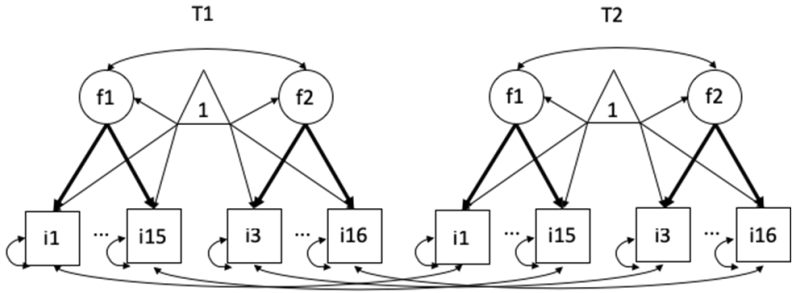
1. I feel confident in my role as a parent.
2. I feel I have quickly caught on to the basic skills of caring for my child.
3. I have difficulties interpreting my baby's cries, knowing whether he or she wants to be fed rather than played with or held.
4. I get uptight if my baby becomes fussy or irritable for longer than a few minutes.
5. I am comfortable playing actively with my baby and getting him or her to smile at me.
6. I feel like I was unprepared in becoming a parent.
7. In most circumstances, even when I am tired, I am able to cope well with meeting my baby's needs.
8. Touching, holding, and being affectionate with my child is comfortable and pleasurable for me.
9. I trust my feelings and intuitions about taking care of my baby.
10. I wonder if I really understand my baby's needs.
11. I am unsure just how much attention I should give my baby.
12. I am able to soothe my baby easily when he or she is crying or fussing.
13. I am concerned that my patience with my baby is limited.
14. I feel comfortable and natural using baby-talk.
15. I find nothing unusually complicated or difficult about feeding, playing with, or providing day-to-day care for a child.
16. The thought of being solely responsible for my child is frightening.

Model for 16 items, 6 timepoints....

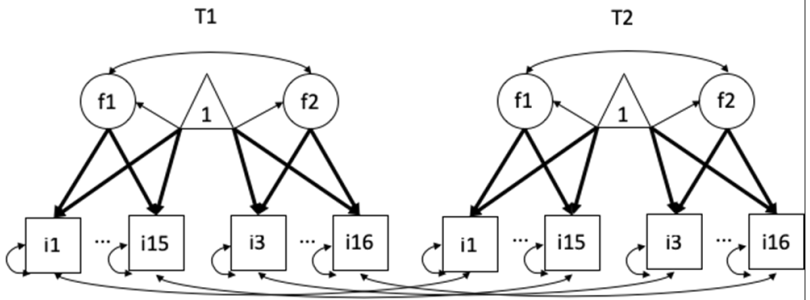
STEP 1: Configural Invariance Model (Same factor model, different estimates)



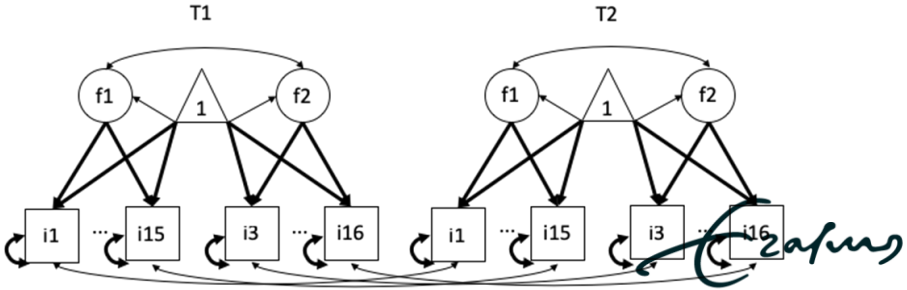
STEP 2: Metric Invariance Model (Same factor loadings)



STEP 3: Scalar Invariance Model (Same factor loadings & intercepts)

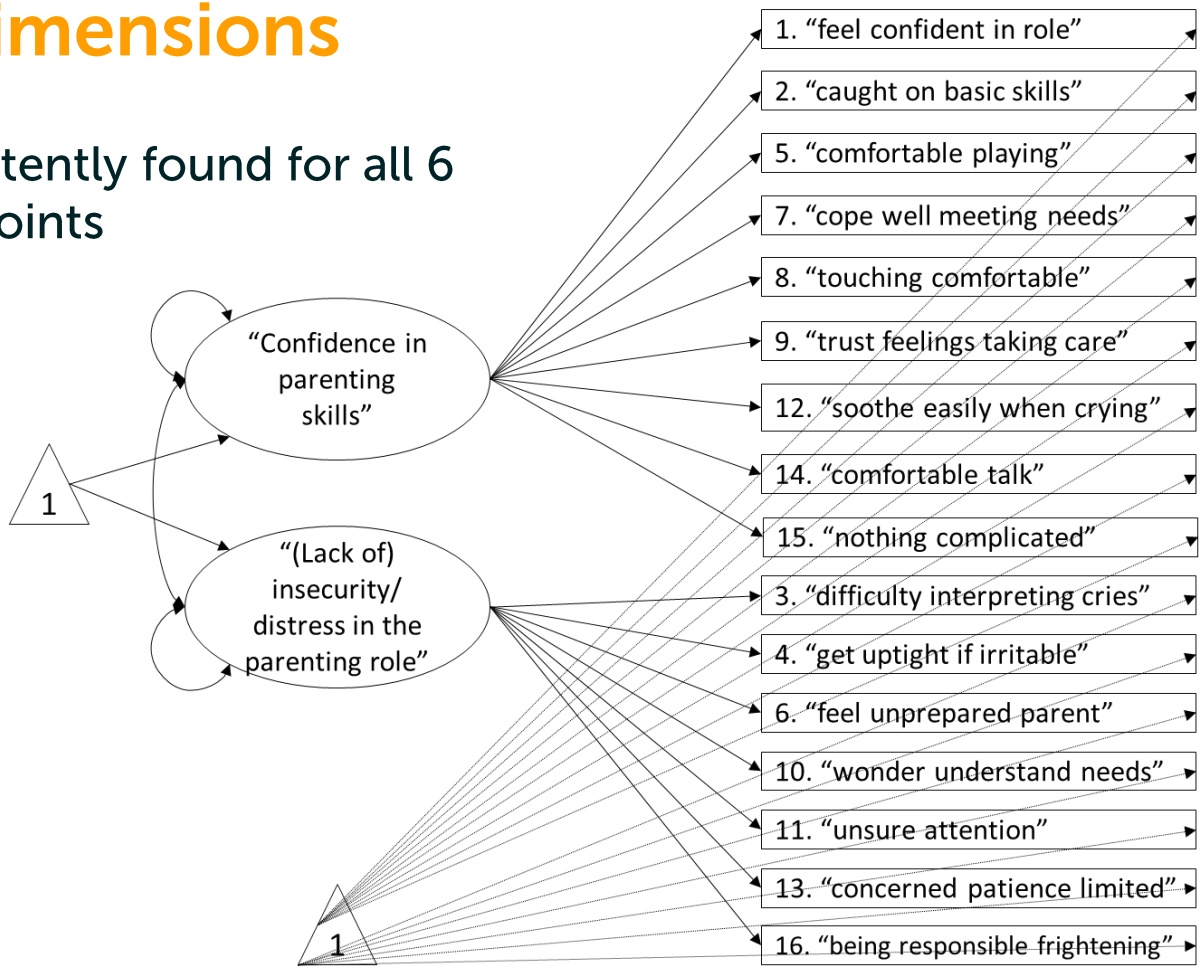


STEP 4: Strict Invariance Model (Same factor loadings, intercepts & residual variances)



Two dimensions

- Consistently found for all 6 time points



Longitudinal MI

Table 4
Model fit results of the measurement invariance analyses

Model	$\chi^2(df)$	RMSEA	CFI	SRMR	ΔCFI	ΔRMSEA	ΔSRMR
Prenatal (T1-T3)							
Configural invariance	3031.10 (1017)	0.033**	0.929*	0.047**			
Metric invariance	3085.01 (1045)	0.032**	0.928*	0.050*	-0,001+	-0,001+	0,003+
Scalar invariance	3229.20 (1073)	0.033**	0.924*	0.052*	-0,004+	0,001+	0,002+
Strict invariance	3452.47 (1105)	0.034**	0.917*	0.065*	-0,007+	0,001+	0,013+
Postnatal (T4-T6)							
Configural invariance	2523.43 (1017)	0.030**	0.907*	0.043**			
Metric invariance	2567.15 (1045)	0.030**	0.906*	0.048**	-0,001+	0,000+	0,005+
Scalar invariance	3739.70 (1073)	0.037**	0.843	0.059*	-0,063	0,007+	0,011+
Strict invariance	4290.02 (1105)	0.042**	0.804	0.147	-0,039	0,005+	0,088
All time points							
Configural invariance	8179.88 (4158)	0.023**	0.921*	0.044**			
Metric invariance	8684.05 (4228)	0.024**	0.912*	0.063*	-0,009+	0,001+	0,019+
Scalar invariance	12205.83 (4298)	0.032**	0.845	0.075*	-0,067	0,008+	0,012+
Strict invariance	14079.51 (4378)	0.035**	0.809	0.138	-0,036	0,003+	0,063

Note. All χ^2-tests have $p < .001$. *Acceptable fit based on RMSEA<.08, CFI>.90, SRMR<.08. ** Good fit based on RMSEA<.05, CFI>.95, SRMR<.05. + Non-substantial deterioration in fit.

Conclusions

- PSE means can be meaningfully compared *during pregnancy*

- But not during infancy and toddlerhood…
- PSE means postnatally represent combination of true changing scores and response shifts

- Suggestion: develop an abbreviated scale with two dimensions covered and that shows longitudinal MI?
- But psychometric proporties of such a new scale should be evaluated more broadly!

Overview of this talk

- ❑ Studying change
- ❑ Testing for longitudinal measurement invariance
- ❑ Issues in testing for longitudinal measurement invariance
- ❑ Conclusions
- ❑ Further reading and practice

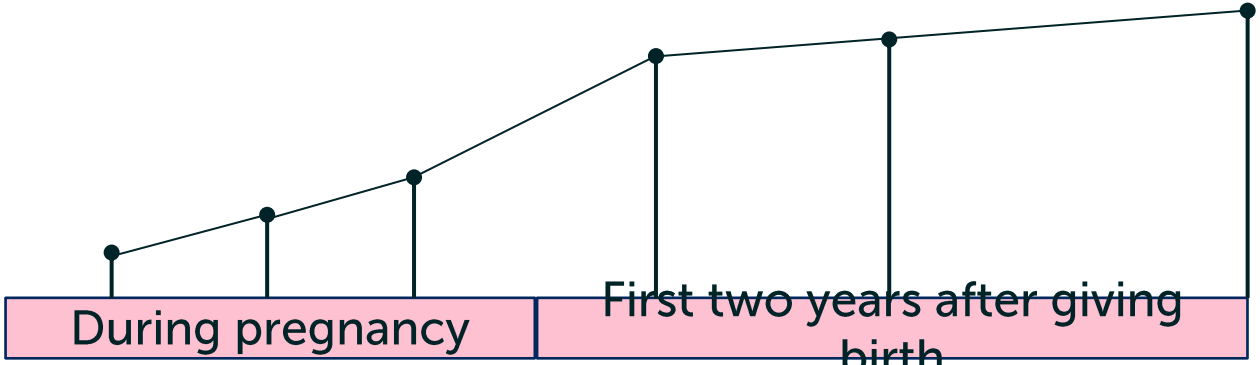
Issues in tesing MI

- Partial MI
  - Allow some loadings/intercepts to be noninvariant
  - To compare means: >=2 loadings/intercepts invariant
- Bayesian MI and approximate MI (blavaan or BSEM)
  - Suitable for complex models and categorical data
- What if scale changes?
  - More/less items
  - Change in content of the items

Overview of this talk

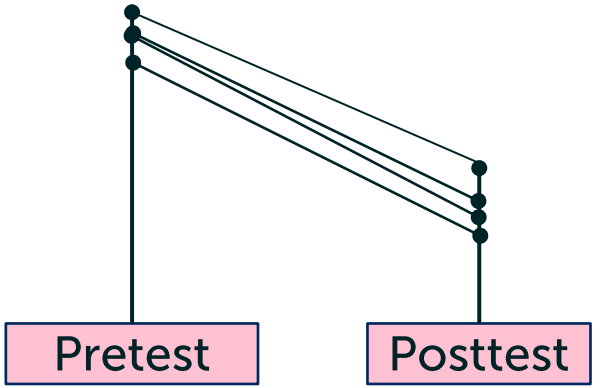
- ❑ Studying change
- ❑ Testing for longitudinal measurement invariance
- ❑ Issues in testing for longitudinal measurement invariance
- ❑ Conclusions
- ❑ Further reading and practice

First testing longitudinal MI, then change



- Parenting competence twodimensional
- Strict MI for prenatal timepoints
- Metric MI for all time points

(De Moor et al., in preparation)



- Response shifts
- Mean comparisons may be confounded

(Fokkema et al. 2013)

Conclusions longitudinal MI

- ❑ Testing for longitudinal MI for constructs measured with multiple indicators.
- ❑ If longitudinal MI does not hold, mean comparisons may be invalid.
- ❑ Some invariance is still allowed.

- ❑ Side effects:
  - ❑ Provides relevant information about what you measure!
  - ❑ Sometimes leads to improvements of your measure.

TAKE HOME MESSAGE:
Always test for longitudinal MI. You now know how to do this!

Overview of this talk

- ❑ Studying change
- ❑ Testing for longitudinal measurement invariance
- ❑ Issues in testing for longitudinal measurement invariance
- ❑ Conclusions
- ❑ Further reading and practice

Further reading

- Papers about MI:
  - Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525-543.
  - Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*(1), 4-70.
  - Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review*, *41*, 71-90.
  - Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.
- Papers about longitudinal MI:
  - Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: An introduction. *Child Development Perspectives*, *4*(1), 5-9.
  - Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological methods*, *22*(3), 486.
  - Fokkema, M., Smits, N., Kelderman, H., & Cuijpers, P. (2013). Response shifts in mental health interventions: An illustration of longitudinal measurement invariance. *Psychological Assessment*, *25*(2), 520.

Further practice

- R/RStudio:
- https://cran.r-project.org/
- https://posit.co/downloads/

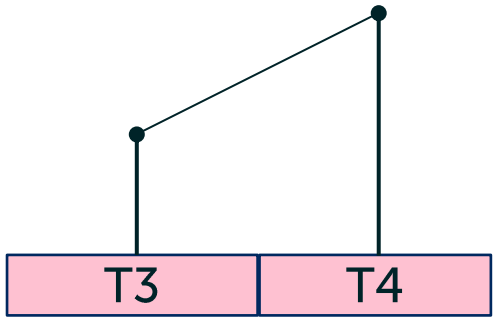
- Lavaan Tutorial:
- https://lavaan.ugent.be/tutorial/

- Mplus:
- https://www.statmodel.com/
- https://www.statmodel.com/MeasurementInvariance.shtml

Next: tutorial!

- ✓ Random selection of 150 mothers, two time points included, 4 items

- ✓ Sum scores indicate significant difference (T3: *M*=23.59 ; T4 *M*=24.33; *p*=.003)
- ✓ But can these means be meaningfully compared?
- ✓ Longitudinal measurement invariance should be tested!



Next: tutorial!

✓ Install R/Rstudio
✓ Open Rscript_Longi_MI.R

Follow the steps in the script

1. Install packages
2. Read in the dataset
3. Compute descriptives

4. Run the longitudinal MI models
5. Fill in the empty code!
6. Answer the questions about the models on the following slides

Questions

Step 1. Configural Invariance

1. Does the one factor model fit the data well at both timepoints?
2. How many parameters do you estimate in this model (count *freely estimated* factor loadings, intercepts, residual and factor variances and factor means, factor covariances and residual covariances)
3. Can you reconstruct the *degrees of freedom* of the model (#observed statistics - #parameters)?

Questions

Step 2. Metric Invariance

1. When inspecting the factor loadings in the configural invariance model, which item shows the largest difference between time points, which the smallest?
2. Are the differences of the factor loadings between time points statistically significant?
3. Does the Metric Invariance model fit well? And compared with the Configural Invariance model?

Questions

Steps 3 and 4. Scalar and Strict Invariance

1. Do these two more restricted model fit the data well?
2. Decide which level of longitudinal MI is achieved.
3. What is you conclusion about whether the means for T3 and T4 can be compared (T3: M=23.59 ; T4 M=24.33; p=.003)?

Erasmus School of
Social and
Behavioural Sciences

Erasmus University Rotterdam