Equivalence des mesures à travers différents groupes dans un sondage

Melanie Revilla

RECSM, UPF

Measurement equivalence

"Measurement equivalence is conceptually defined as whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute."

Horn and McArdle, 1992: 117

"Measurement invariance or measurement equivalence is a statistical property of measurement that indicates that the same construct is being measured across some specified groups."

Wikipedia

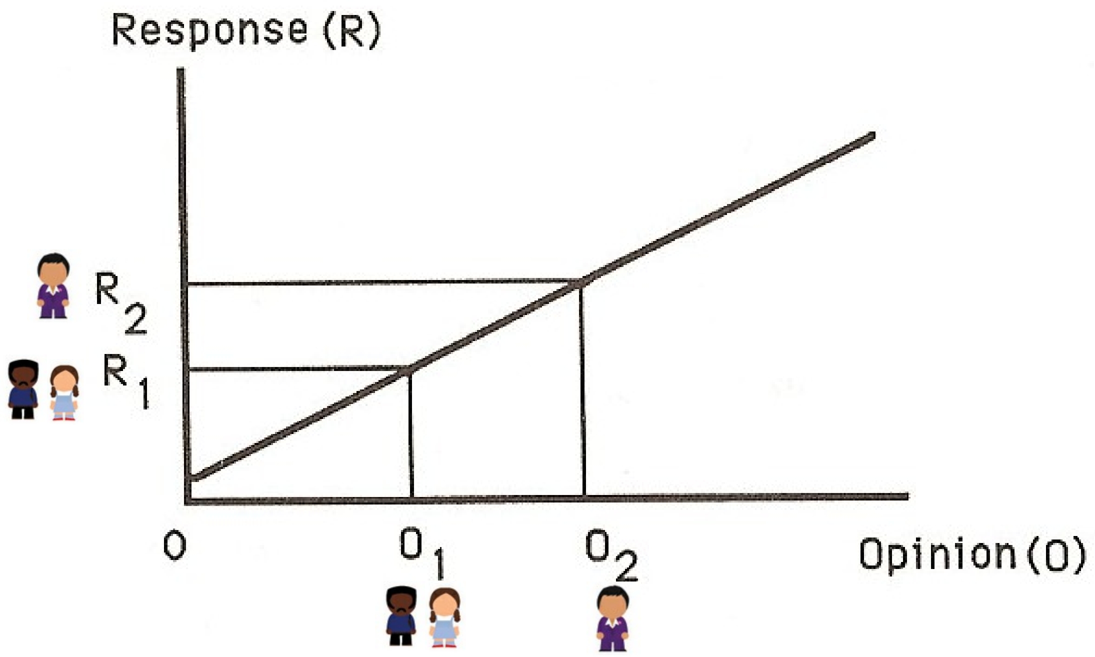
Measurement equivalence

- Usually discussed in the frame of cross-national research (across **countries and languages**):
  - Issue of cultural differences and translation
- But equivalence can also be checked across:
  - Data collection **modes** (e.g. face-to-face vs Web)
  - Data collection **designs** (unimode vs mixed-mode)
  - **Socio-demographic** groups (education levels)
  - **Time** (longitudinal studies)
  - Etc

→ Idea: in different groups people can express themselves in different ways

Measurement equivalence (ME)

- If there is measurement equivalence, 2 persons with the same opinion will give the same answer, whatever their group



Why does it matters?

- Classic assumption to be able to combine answers and compare them
  - Equality of the response function
- Important because observed differences may result from:
  - Non equivalent measures
  - And/or from real substantive differences
- If measurement equivalence does not hold:
  - **cannot** make direct comparison across groups

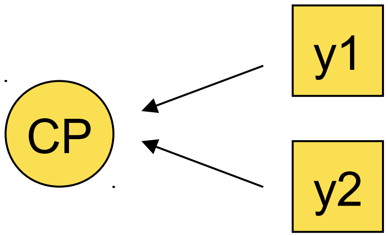
Important preliminary distinction

Different kinds of concepts (Northrop, 1947; Blalock, 1990)

- Concepts-by-intuition (CI)
  - simple concepts whose meaning is immediately obvious.
- Concepts-by-postulation (CP)
  - less obvious concepts which require explicit definitions.
  - also called constructs
- For CP, distinction between:
  - concepts with reflective indicators
  - concepts with formative indicators

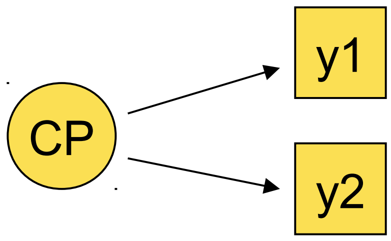
CP with formative indicators

- Concept defined by a combination of indicators
- These variables are not necessarily correlated
- The combination defines the CP
- Requirement: the definition should take into account all necessary components



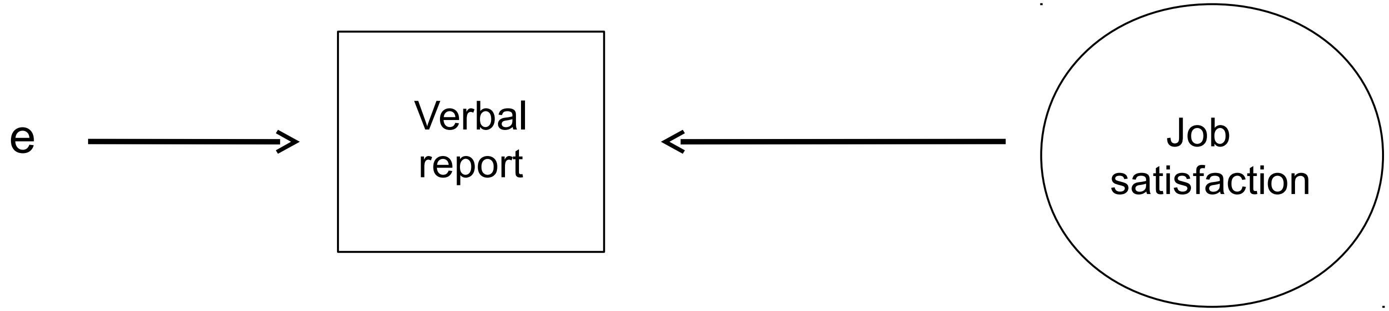
CP with reflective indicators

- CP is indicated by some similar possible indicators (consequences)
- These indicators will always be correlated because all are affected by the CP
- Usually 3 indicators are enough (identified)



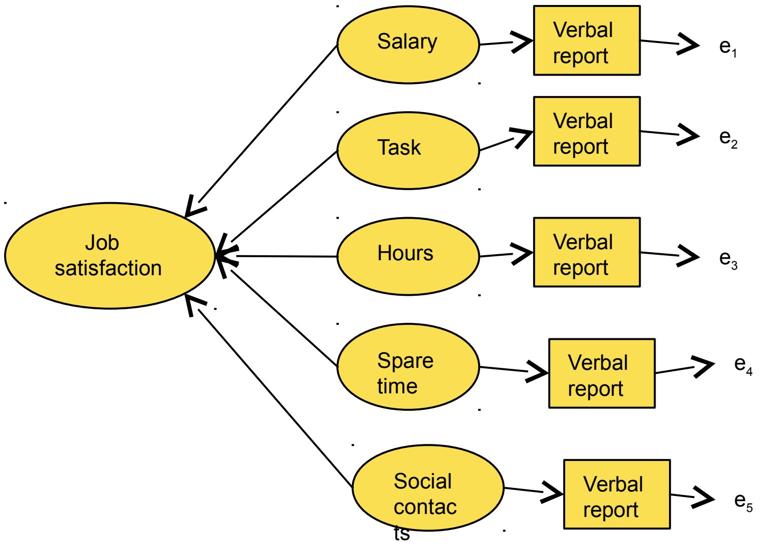
Example of job satisfaction

- Operationalized as a CI
  - *How satisfied are you with your job?*



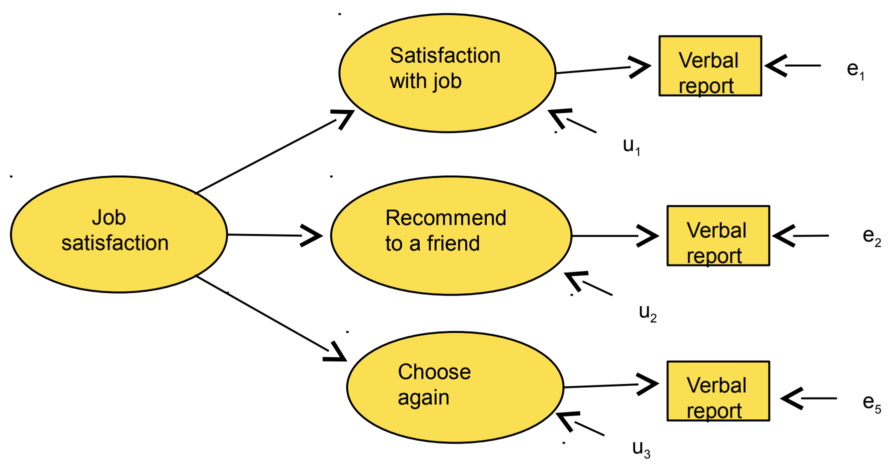
Example of job satisfaction

- Operationalized as a CP with formative indicators
  - *How satisfied are you with :*
    - *your salary?*
    - *promotion possibilities?*
    - *your contacts with colleagues*
    - *your spare time*
    - *……*
- Should be exaustive!



Example of job satisfaction

- Operationalized as a CP with reflective indicators
  - *How satisfied are you with your job?*
  - *Would you recommend your job to a friend?*
  - *Would you chose this job again?*



The responses to all three questions are supposed to be affected by the satisfaction of the respondent with his/her job.

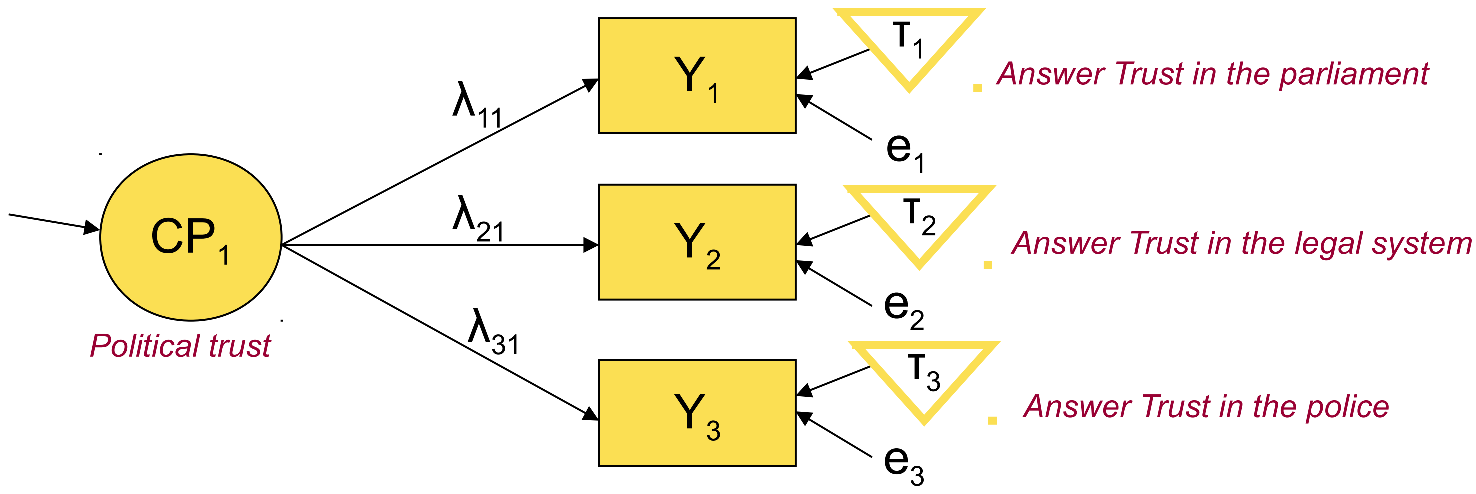
Testing for measurement equivalence

- Classic procedure only exist to test for measurement equivalence in the case of CP with reflective indicators
  - We start with this case

Testing for measurement equivalence

CP / reflective indicators

Basic Confirmatory Factor Analysis model

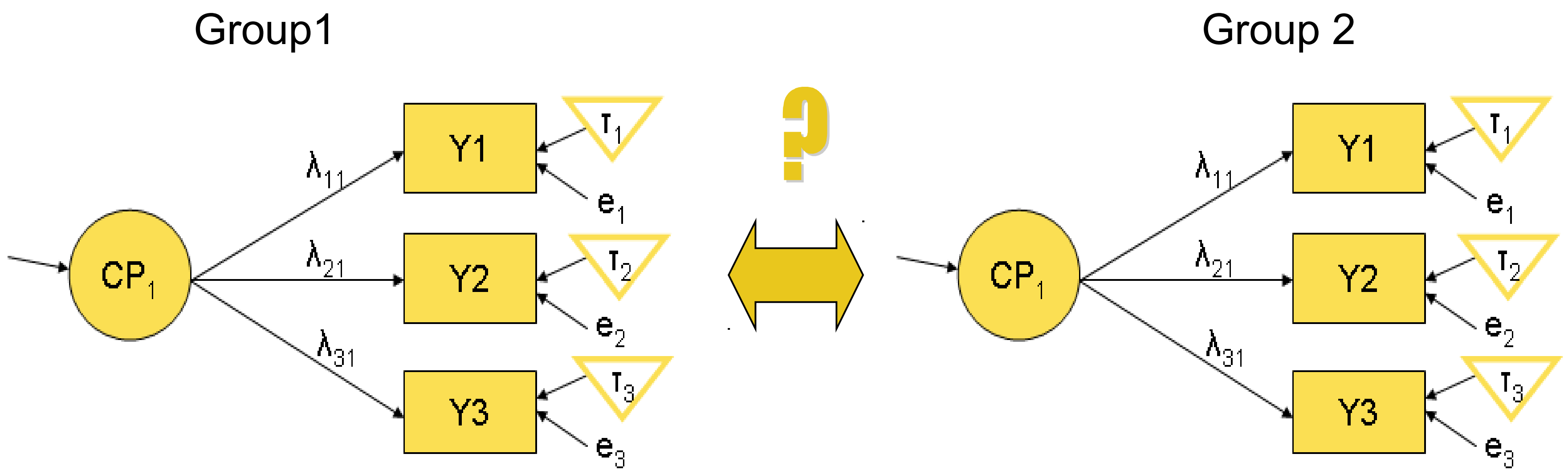


intercepts · slopes · error terms

≈ regression equation → $$Y_i = \tau_i + \lambda_{ij} CP_1 + e_i \qquad i = 1,2,3$$

Dependent variable (Y_i) · Independent variable (CP_1)

Multiple Group CFA approach

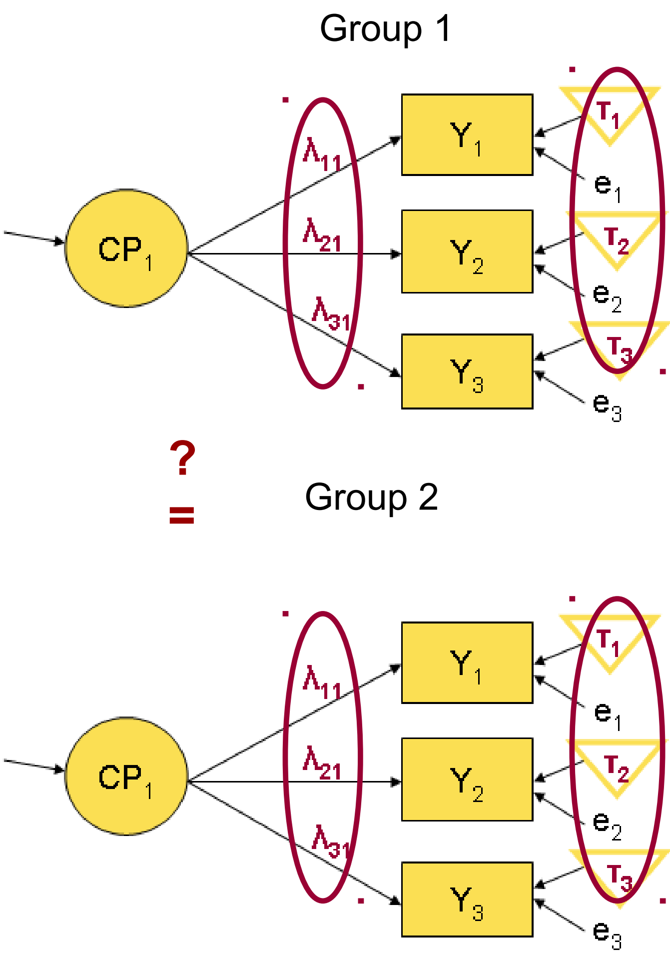


- Multiple group:
  - possible to test for equality of the parameters in the different groups
  - constraints across groups

- Can be extended to more groups

Different levels of invariance (Meredith, 1993)

- **Configural**
  - Same model holds in all groups
- **Metric**
  - Configural holds + Slopes (λ_{ij}) the same in all groups
  - Sufficient for comparison of relationships
- **Scalar**
  - Metric holds + Intercepts (τ_i) the same in all groups
  - Sufficient for comparison of means
- More: error terms, etc…



How to do it in practice?

- Analyses can be done with standard SEM softwares
  - LISREL/M*plus*/R…
  - Based on covariance matrices & means
  - Recommended sample size: >200 in each group
  - 3-step procedure: configural, metric, scalar
  - Syntax quite easy to get estimates

- More tricky but crucial step: testing

Testing the model

- Assessing **global fit**:
  - Chi^2 test
  - Fit indices: RMSEA (<.05), CFI (>.9), etc…
- Limits:
  - Depends on sample size
  - Sensitive to deviations from normality
  - Sensitive to model complexity
- Saris, Satorra and van der Veld (2009)
  - Show that we should test at the parameter level + take into account type II errors (H_0 not rejected despite being false)

Testing the model

- Assessing **local fit**
  - Use MI, EPC and Power

	Low Power	High Power
Insignificant MI	Inconclusive	No misspecification
Significant MI	Misspecification	Inspect EPC

- JRule software
  - Available for Lisrel (van der Veld, Saris, Satorra) and Mplus (Oberski)
  - For each parameter, it tells if it is misspecified
  - /!\ The researcher should decide how large can a deviation be before considering it a misspecification

Testing the model

- Limits:
  - If the model is large and/or you have many different groups: possible to get **many misspecifications**
    - With which one should you start?
    - When should you stop?

- Always free parameters only **one by one**

- Always check if estimates are really different when you free a new parameter in your model
  - Difference may be statistically significant but not **substantially big enough to be meaningful**

What if equivalence does not hold?

- **Partial** invariance (Byrne, Shavelson, Muthén, 1989)
  - If some indicators are equivalent but not all
  - Consistent estimates of the means of the **latent variables** if at least **2** indicators are scalar invariant

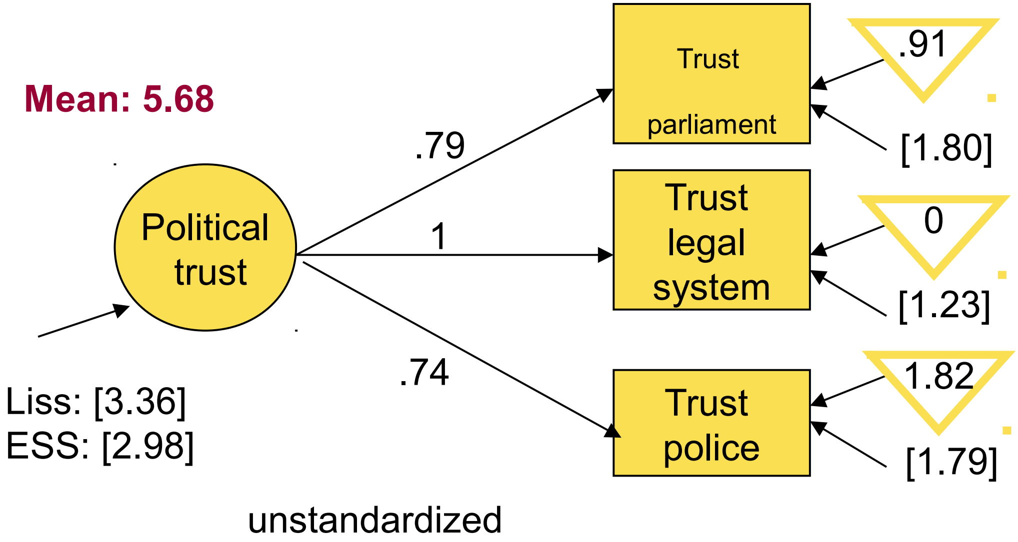
- If you have many groups, often at least several are invariant:
  - Report the results for these ones, and report the deviant cases separately
  - Example of how to do it: Coromina, Saris and Oberski (2008)

Example (Revilla, 2013)

- Comparison:
  - ESS round 4 (F2f) / LISS panel (Web)
  - Same questions / Same fieldwork period
  - The Netherlands

- Configural, metric, scalar invariance achieved

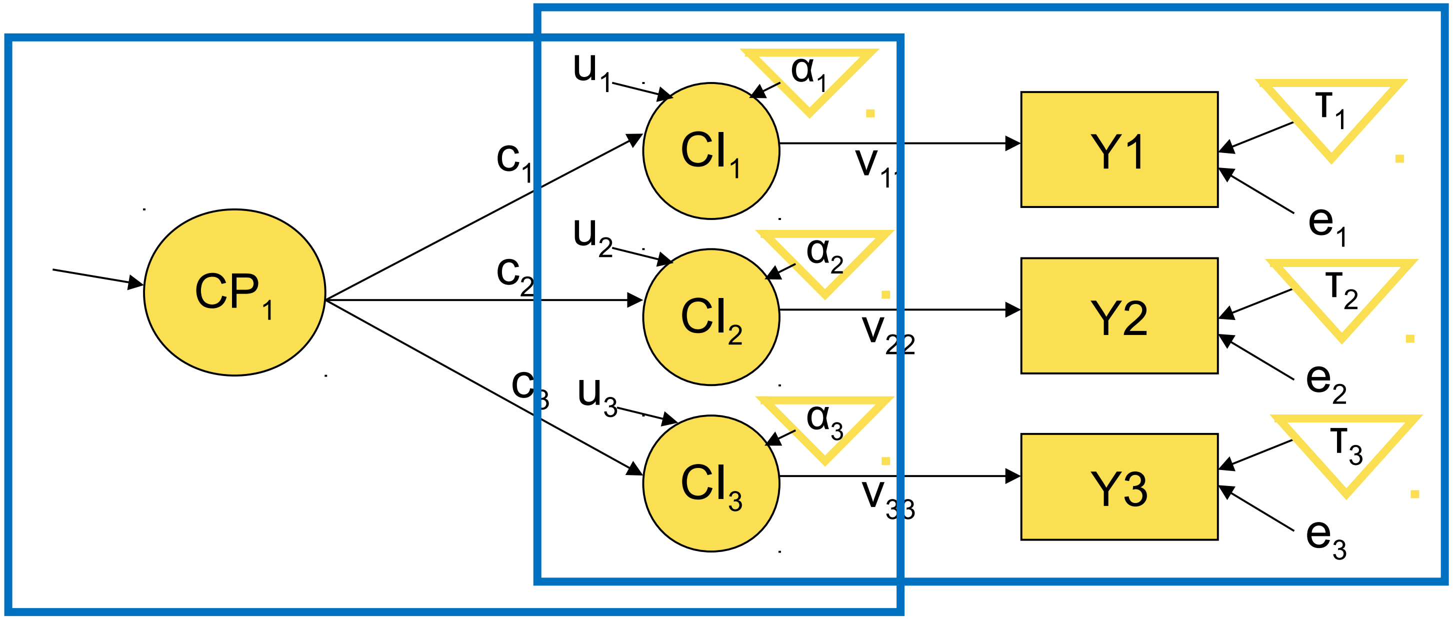
- Possible compare relationships + means across LISS and ESS for that concept



Criticisms (Saris & Gallhofer, 2007)

- Test of **"measurement equivalence"** too strict
  - Need to separate cognitive and measurement processes
  - Cognitive process
    - Understanding of the question
    - Relationship between the higher order variables (CP) and the CI
    - Consistency
  - Measurement process
    - Expressing the answer
    - Relationship between the observed answers and the latent variables they attempt to measure
    - Validity (+ reliability, method effects)

Proposition (Saris & Gallhofer, 2007)

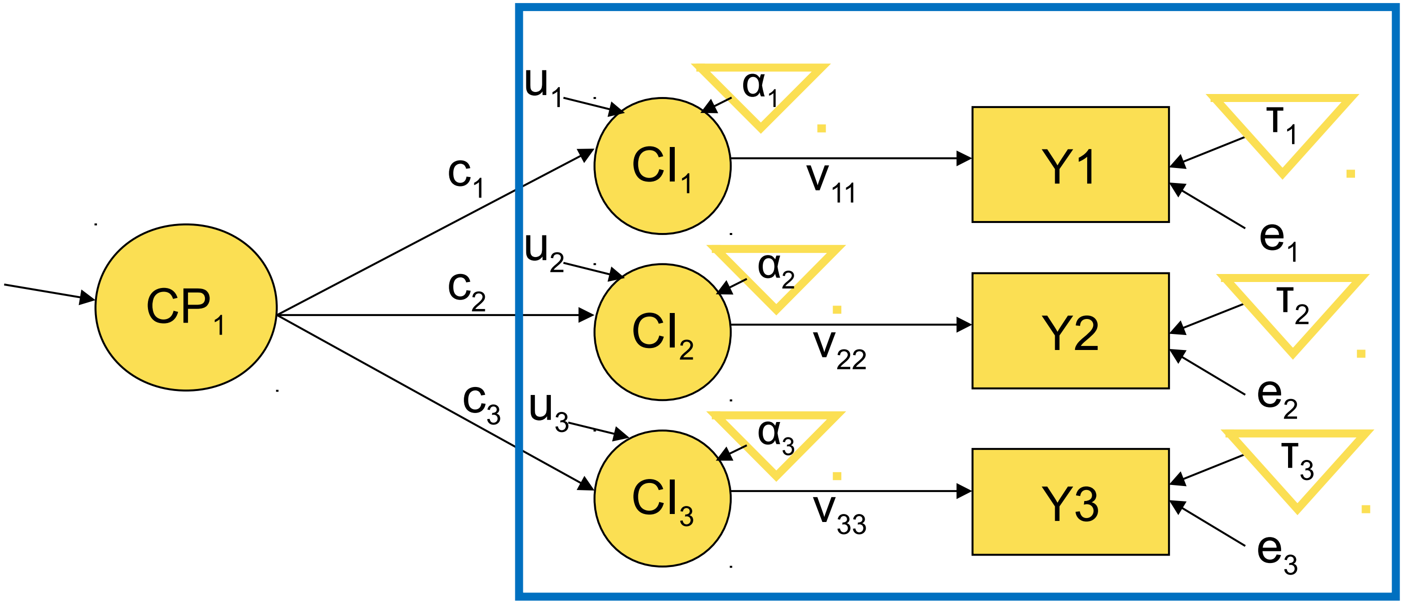


- Measurement process: $Y_i = \tau_i + v_{ij}\, CI_i + e_i \quad i = 1,2,3$

  (Method effect could be included too: distinction true score-factor)

- Cognitive process: $CI_i = \alpha_i + c_i\, CP_1 + u_i \quad i = 1,2,3$

Proposition (Saris & Gallhofer, 2007)



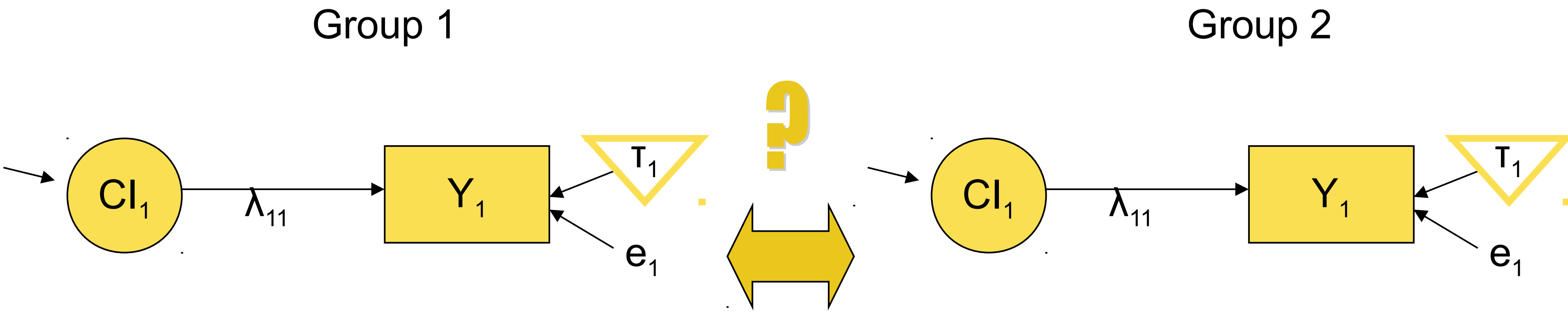
- Testing equivalence measurement process

⇕

- Testing equivalence for CI

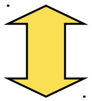
Can we test for measurement equivalence for CI or CP with formative indicators?

Case of CI



$$Y_i = \tau_i + \lambda_{ij} CI_i + e_i$$

- Testing equivalence for CI



- Testing equivalence for single indicator

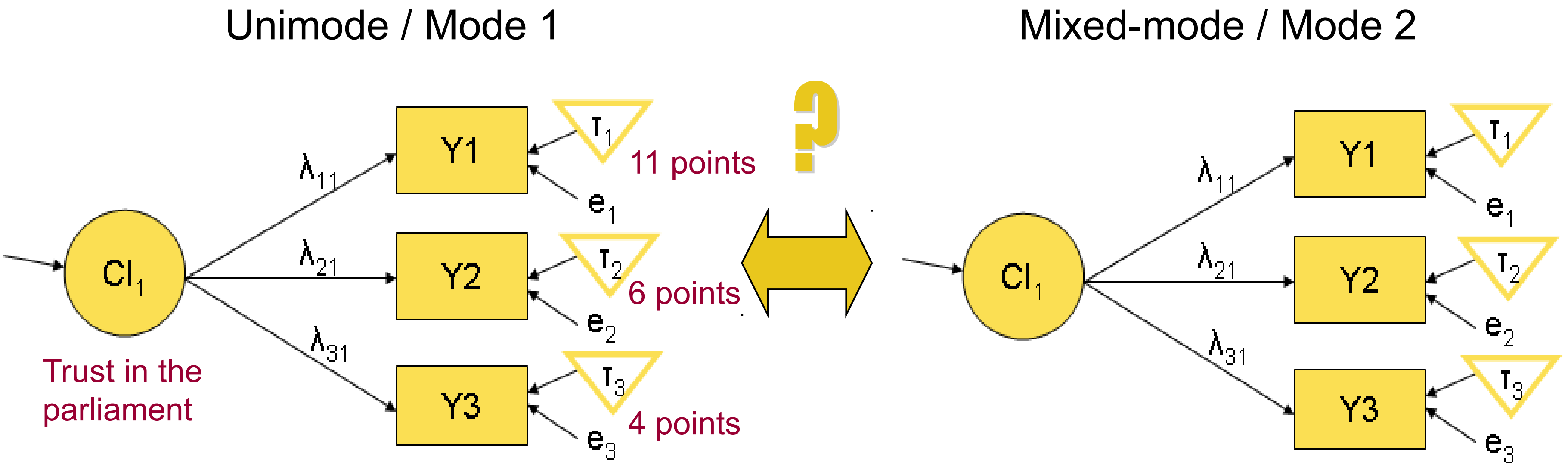
Single ➔ multiple indicators?

- Problem: model just presented not identified

- We need to get several indicators for the CI
  - CI by definition can be asked using a single question
  - But possible to use multiple formulations/methods/scales for this question

- In that way, we can come back to a CFA model with one latent variable with several observed indicators

Can apply again MGCFA



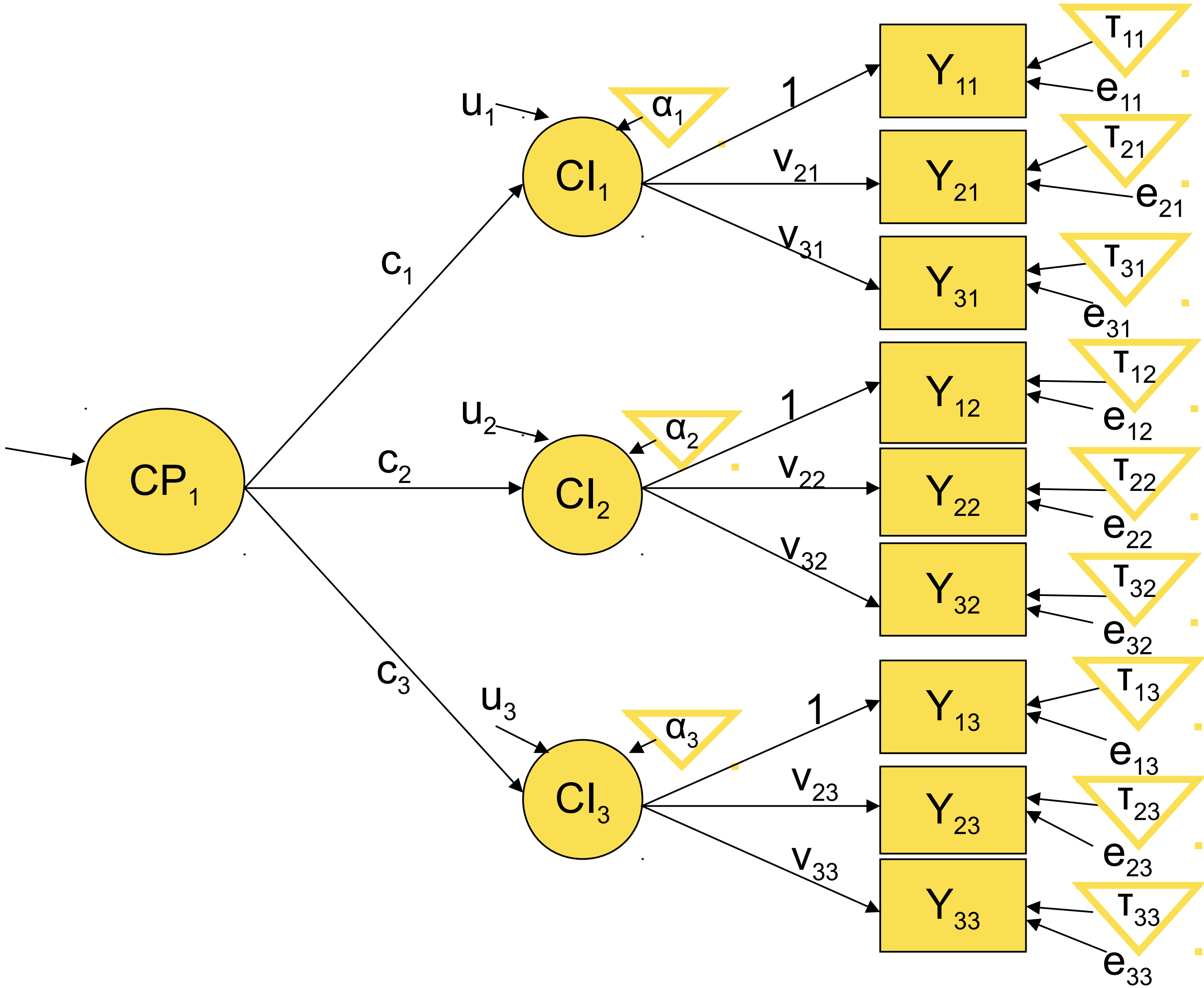
- Then, same procedure to get the estimates and test the model

Alternative / solution for CP formative indicators

- Saris, Pirralha and Zavala (under review)
  - Proposed to vary slightly the question itself
  - For instance ask 3 direct questions about job satisfaction
    - “How satisfied are you with your job?”
    - “How much do you like your job?”
    - ”How happy are you with your job?”

- For CP with formative indicators:
  - Same procedure, we need to come back to something similar as for CP with reflective indicators
  - For each of the formative indicators, we need to ask different questions/use different methods

More general model



Implication

- In order to separate the cognitive and the measurement process, we need to get several indicators for each CI

- Can be done by using different methods (e.g. different scales) for the same CI
  - Same persons get the question several times using different methods
  - But 20 minutes at least to avoid memory effects (Van Meurs & Saris, 1990)
  - We cannot do it for all questions of the questionnaire

In summary

- Measurement equivalence can be assessed using Multiple Group Confirmatory Factor Analysis
- Classic procedure exists for CP with reflective indicators
  - Testing is the most crucial step in this case
  - Very important to consider also local fit
  - Limit: the classic procedure does not separate cognitive and measurement process
- For CI, for CP with formative indicators, and if we want to separate cognitive and measurement:
  - We need to repeat questions

Thank you for your attention!

Contact: melanie.revilla@upf.edu

References

Blalock, H.M. (1990). Auxilary measurement theories revisited. In Hox J. J., and J. de Jong-Gierveld (eds.), *Operationalization and Research Strategy*. Amsterdam: Swets and Zeitlinger, 33─49.

Byrne, B.M., R.J. Shavelson, and B. Muthén (1989). "Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance." *Psychological Bulletin* 105(3): 456-466.

Coromina, L., Saris, W.E. and D. Oberski (2008). The quality of the measurement of interest in the political issues presented in the media in the ESS. ASK, 2008, nr 17, ISSN 1234–9224. Available at: http://daob.nl/wp-content/uploads/2013/03/Coromina_Saris_Oberski_IPIMedia_2008.pdf

Horn, J.L., and J.J. McArdle (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3):117-144.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.

Northrop, F.S.C. (1947). *The Logic of the Sciences and the Humanities*. New York: World Publishing Company.

Revilla, M. (2013). "Measurement invariance and quality of composite scores in a face-to-face and a web survey" *Survey Research Methods* 7(1): 17-28.

Saris, W.E., and I. Gallhofer (2007). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. New York: Wiley.

Saris, W.E., Pirralha, A., and D. Zavala Rojas (under review). *Testing the comparability of different types of social indicators across groups*.

Saris, W. E., Satorra, A., and W.M. Van der Veld (2009). Testing Structural Equation Models or Detection of Misspecifications? *Structural equation modeling: A multidisciplinary Journal*, 16(4), 561-582.

Van Meurs, L., and W.E. Saris (1990). Memory Effects in MTMM Studies.' Pp. 134-146 in *Evaluation of Measurement Instruments by Meta-analysis of Multitrait-Multimethod Studies*, edited by Willem E. Saris and Lex van Meurs. Amsterdam, the Netherlands: North Holland.