Testing measurement Invariance in a CFA framework – State of the art

Philipp Sischka
University of Luxembourg
Contact: Philipp Sischka, University of Luxembourg, INSIDE, Porte des Sciences, L-4366 Esch-sur-Alzette, philipp.sischka@uni.lu

UNIVERSITÉ DU LUXEMBOURG

1. Why testing for measurement invariance?

- To be able to compare latent constructs, the measurement structures of the latent factor and their survey items need to be stable across the compared research units (e.g., Vandenberg, & Lance, 2000).
- Therefore, testing for measurement invariance (also termed measurement equivalence) is a necessary precondition to conduct comparative analyses (Millsap, 2011).
- If invariance is not tested, then, differences between groups, or the lack thereof, in the latent constructs cannot be unambiguously attributed to 'real' differences or to differences in the measurement attributes

2. Sources of non-invariance

Non-invariance could emerge if…

- …the conceptual meaning or understanding of the construct differs across groups,
- …groups differ regarding the extent of social desirability or social norms,
- …groups have different reference points, when making statements about themselves,
- …groups respond to extreme items differently,
- …particular items are more applicable for one group than another,
- …translation of one or more item is improper (Chen, 2008)

3. Forms of Measurement invariance

- Configural invariance (same pattern of loadings of the items on the constructs in each group)
- Metric invariance (same loadings of the items on the constructs in each group)
- Scalar invariance (same loadings and intercepts of the items on the constructs in each group)
- Testing with
  - CFA
  - IRT

4. Order of testing forms of measurement invariance

- Constrained-baseline strategy
  - Starts with a "fully-constrained" baseline model (fixed loadings and interepts)
  - Stepwise freeing of the parameter
- Free-baseline strategy
  - Starts with a "fully-free" baseline model
  - Stepwise fixing of the parameter
- Most common: Free-baseline strategy, because…
  - Constraint-baseline strategy is problematic (if misspecified several subsequent tests are biased) (Lee, 2009)

5. Evaluation of measurement invariance

- $\Delta\chi^2$ difference test
- Changes in approximate fit statistics
  - ΔSRMR
  - ΔRMSEA
  - ΔNCI
  - Most common: ΔCFI (Chen 2007; Cheung, & Rensvold, 2002; Mead, Johnson, & Braddy, 2008)
- Modified CFI (Lai, & Yoon, 2015)
- Magnitude of difference between the parameter estimates (Oberski, 2014)

6. Evaluation of sources of non-invariance

- Statistical significance of the $\Delta\chi^2$ difference test after Bonferroni adjustment
- Modification Indices (e.g., Yoon, & Millsap, 2007)
- Clusterwise simultaneous component analysis (De Roover, Timmerman, De Leersnyder, Mesquita, & Ceulemans, 2014)
- Test for cluster bias (Jak, Oort, & Dolan, 2013)

- If full measurement invariance does not hold, partial measurement can be tested.

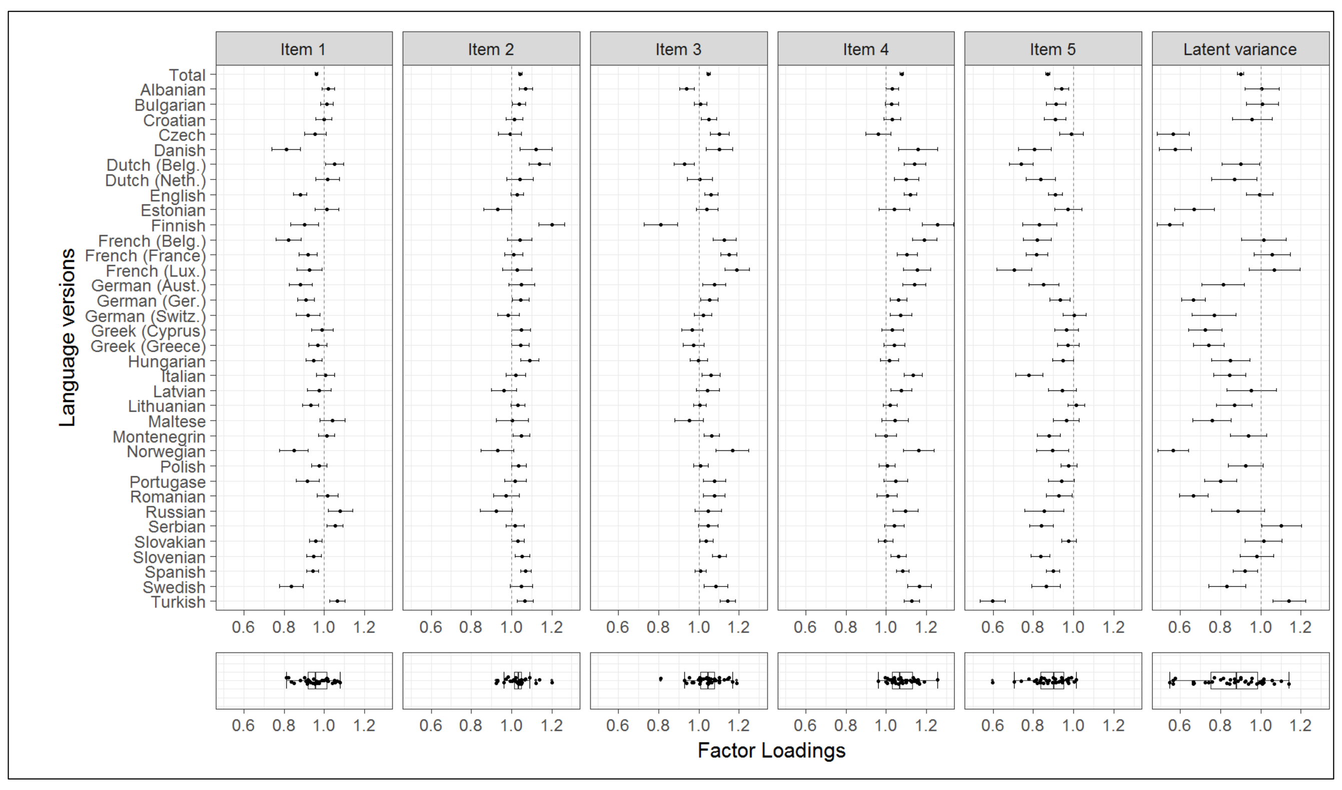
7. New developments

- Bayesian (approximate) measurement invariance test (e.g., Verhagen, & Fox, 2013)
- Alignment Method (Asparouhov, & Muthen, 2014; Marsh et al., 2017)
- Bootstrapping of Fit indices (Yuan, Hayashi, & Yanagihara, 2007; Zang, & Savalei, 2016)

8. Study design

- European Working Condition Survey 2015
  - 35 countries
  - 35 language versions
  - 42,779 respondents (employees and self-employed)
  - 49.6% females, n = 20,930)
  - Age: 15 to 89 years (M = 43.3, SD = 12.7)
- Aim: Testing language measurement invariance of the WHO-5 well-being scale

9. Results (I)



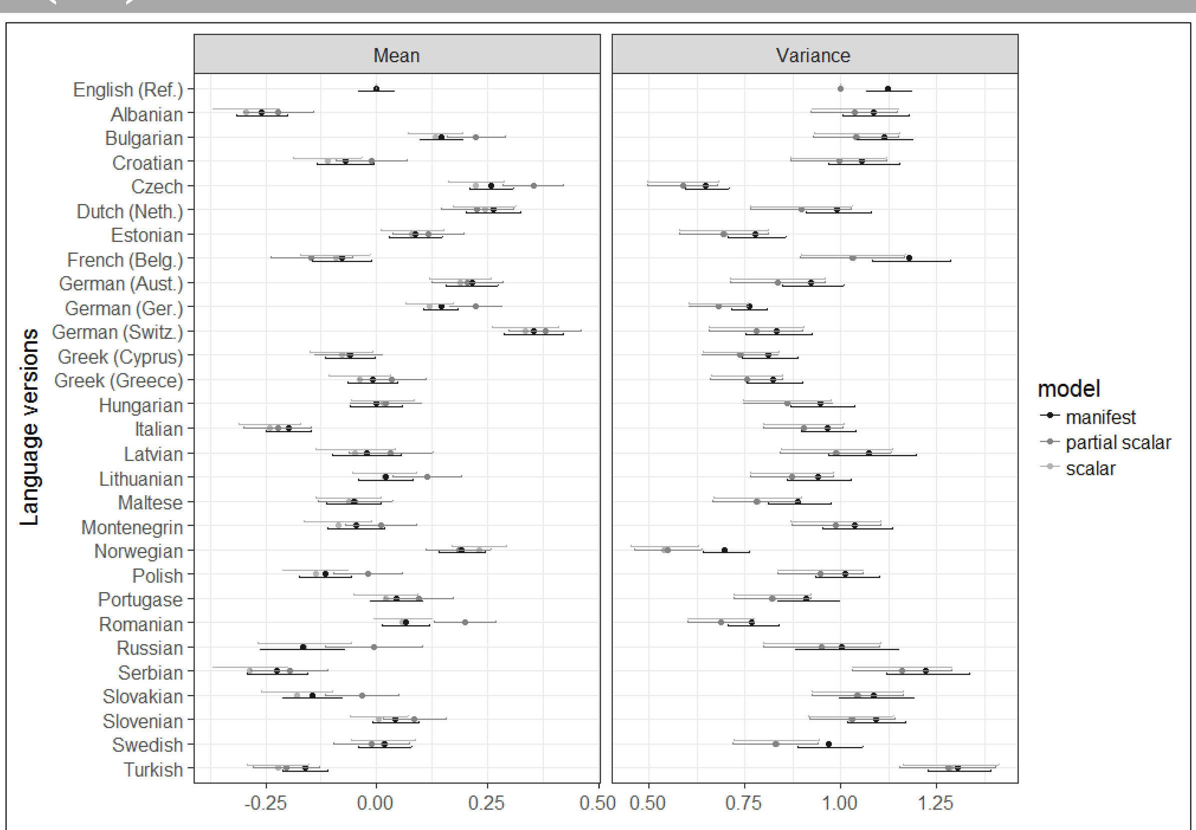
10. Results (II)

Table 4. Test of measurement invariance and fit indices for WHO-5 one-factor model across 29 language versions (N = 33,723).

Form of invariance	χ^2	p	df	$RMSEA$	$RMSEA$ [90% CI]	$SRMR$	CFI	TLI
Configural invariance	688.355	.000	145	.057	[.054; .060]	.017	.989	.979
Metric invariance	1300.094	.000	257	.059	[.057; .062]	.046	.980	.977
Scalar invariance	3687.381	.000	369	.088	[.086; .090]	.065	.936	.949
Partial scalar invariance[a]	1723.693	.000	285	.066	[.064; .068]	.050	.972	.972
Δ Configural – metric	+611.739		+112	+.002		+.029	-.009	-.002
Δ Metric - scalar	+2387.287		+112	+.029		+.019	-.044	-.028
Δ Metric – partial scalar	+423.599		+28	+.007		+.004	-.008	-.005

Notes. *RMSEA* = root mean squared error of approximation; *RMSEA* [90% *CI*] = 90% confidence interval of root mean squared error of approximation; *SRMR* = standardized root mean square residual; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; [a] = Intercepts of Item 1, 3 and 5 freely estimated. Without the Danish, the Belgium Dutch, the Finnish, the France French, the Luxembourg French and the Spanish version.

11. Results (III)



12. Discussion

- Measurement invariance testing is crucial in cross-cultural research
- Without testing MI it remains unclear whether cross-cultural differences (or the lack thereof) in the latent construct are due to differences in the measurement attributes
- If non-invariant items are found, one should try to explain the differential item functioning
- However, especially in studies with many groups full scalar invariance seldom can be established
- Possible solutions are Bayesian (approximate) approaches of measurement invariance tests or the newly developed alignment method

Asparouhov, T., & Muthen, B. (2014). Multiple-Group Factor Analysis Alignment. *Structural Equation Modeling, 21*, 495-508; Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504; Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*, 1005-1018; Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255; De Roover, K., Timmerman, M. E., De Leersnyder, J., Mesquita, B., & Ceulemans, E. (2014). What's hampering measurement invariance: detecting non-invariant items using clusterwise simultaneous component analysis. *Fronties in Psychology*, Advanced online. doi: 10.3389/fpsyg.2014.00604; Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling, 20*, 265-282; Lai, M. H. C., & Yoon, M. (2015). A modified comparative fit index for factorial invariance studies. *Structural Equation Modeling, 22*, 236-248; Lee, J. (2009). *Type I error and power for mean and covariance structure confirmatory factor analysis for differential item functioning detection: Methodological issues and resolutions*. (University of Kansas). ProQuest Dissertations and Theses; Marsh et al., 2017). What to do when scalar invariance fails: the extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*. Advanced online. doi: 10.1037/met0000113; Mead, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and Sensitivity of Alternative Fit Indices in Tests of Measurement Invariance. *Journal of Applied Psychology, 93*, 568-592; Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge. Oberski, D. L. (2014). Evaluating sensitivity of parameters of interest to measurement invariance in latent variable models. *Political Analysis, 22*, 45-60; 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, 2*, 4-69; Verhagen, A. J., & Fox, J. P. (2013). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology, 66*, 383-401; Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling, 14*, 435-463; Yuan, K. H., Hayashi, K., & Yanagihara, H. (2007). A class of population covariance matrices in the bootstrap approach to covariance structure analysis. *Multivariate behavioral research, 42*, 261-281; Zhang, X., & Savalei, V. (2016). Bootstrapping Confidence Intervals for Fit Indexes in Structural Equation Modeling. *Structural Equation Modeling, 23*, 392-408.