Philipp E. Sischka

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

Testing measurement invariance in a confirmatory factor analysis framework – State of the art

13th Conference of the European Sociological Association 2017, 29.08-01.09

RN 21: Quantitative Methods: Increasing Comparability in Cross-National Research

Measurement invariance – Why?

Why testing for measurement invariance?

- *Measurement structures* of latent factors *need to be stable* across compared research units (e.g., Vandenberg, & Lance, 2000).
- Differences between groups in latent constructs *cannot be unambiguously attributed* to ‘real’ differences if no MI test is conducted.
- Testing measurement invariance (MI) is *a necessary precondition* to conduct comparative analyses (Millsap, 2011).

Possible Causes for Non-invariance

Non-invariance could emerge if…

- …*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)

Different testing frameworks and forms of Measurement invariance

Measurement invariance testing

- within CFA framework
- within IRT framework

Forms of Measurement invariance

- Configural invariance (loadings of the items show same pattern for each group)
- Metric invariance (same loadings for each group)
- Scalar invariance (same loadings and intercepts for each group)

Evaluating Measurement invariance

Evaluation criteria of MI within multiple-group CFA

- χ^2 difference test
- Changes in approximate fit statistics
  - ΔSRMR
  - ΔRMSEA
  - ΔNCI
  - ΔCFI: Most common (Chen 2007; Cheung, & Rensvold, 2002; Mead, Johnson, & Braddy, 2008; Kim et al., 2017)
  - Modified CFI (Lai, & Yoon, 2015)

Measurement invariance testing within many groups

MI testing within many groups

- Exact measurement invariance is *mostly rejected* within many groups.
- *Stepwise post hoc adjustments* based on modification indices *have been severely criticized* (e.g., Marsh et al., 2017).
  - *Many steps* because of many parameter that have to be adjusted
  - Modification indices show high multi-collinearity, making *adjustments* often *arbitrarily*
  - It is not guaranteed that the simplest, most interpretable model is found
- Asparouhov and Muthén (2014) presented a new method for multiple-group CFA: The alignment method

The alignment method in a nutshell

Alignment goal

- Alignment method is a scaling procedure to refine scales and scores for comparability across many groups
- *Goal*: Finding (non-)invariance pattern in large data set
- The alignment method is and iterative procedure that uses a *simplicity function* similar to the rotation criteria used with EFA
- This function will be minimized where there are few large noninvariant parameters and many approximately invariant parameters (rathen than many medium-sized noninvariant parameters).

The alignment method in a nutshell

Alignment approaches and procedure

- Three approaches
  - ML Free approach
  - ML Fixed approach (fixing the mean of one group to 0)
  - Bayesian approach
- Procedure
  - Configural model as starting point (factor means = 0, factor variance = 1)
  - Estimating parameter by freely estimate factor means and variances and iteratively fixing factor loadings and intercepts
  - Final aligned model has same fit as configural model

The alignment method in a nutshell

Alignment fit statistics

- Evaluating degree of (non)invariance and alignment estimations
- Fit statistics
  - Simplicity function value
  - R^2 (between 0 → noninvariant and 1 → invariant)
  - Variance of freely estimated parameter
  - Number (percentage) of approximate MI groups
  - Monte carlos simulation: Reproducibility of the estimated parameters

Study design

Study aim

- Testing measurement invariance of the WHO-5 well-being scale across European countries (one-factor model; fixed-factor scaling method)

Sample overview

- European Working Condition Survey 2015
- 33 European countries
- 41,290 respondents (employees and self-employed)
- 49.6% females, n = 20,493
- Age: 15 to 89 years (M = 43.3, SD = 12.7)

Study design

Instructions: Please indicate for each of the 5 statements which is closest to how you have been feeling over the past 2 weeks.							
	Over the past 2 weeks…	At no time	Some of the time	Less than half of the time	More than half of the time	Most of the time	All of the time
1	... I have felt cheerful and in good spirits.	0	1	2	3	4	5
2	... I have felt calm and relaxed.	0	1	2	3	4	5
3	... I have felt active and vigorous.	0	1	2	3	4	5
4	... I woke up feeling fresh and rested.	0	1	2	3	4	5
5	... my daily life has been filled with things that interest me.	0	1	2	3	4	5

Measure: WHO-5

- Various studies confirmed its high reliability, one-factor structure (e.g., Krieger et al., 2014) predictive and construct validity (Topp et al., 2015).
- WHO-5 is used in health-related domains such as suicidology (Andrews & Withey, 1976), alcohol abuse (Elholm, Larsen, Hornnes, Zierau, & Becker, 2011), or myocardial infarction (Bergmann et al., 2013).

Results (I)

Table 1. Fit indices of the WHO-5 one-factorial structure from confirmatory factor analysis for the EWCS 2015 wave.

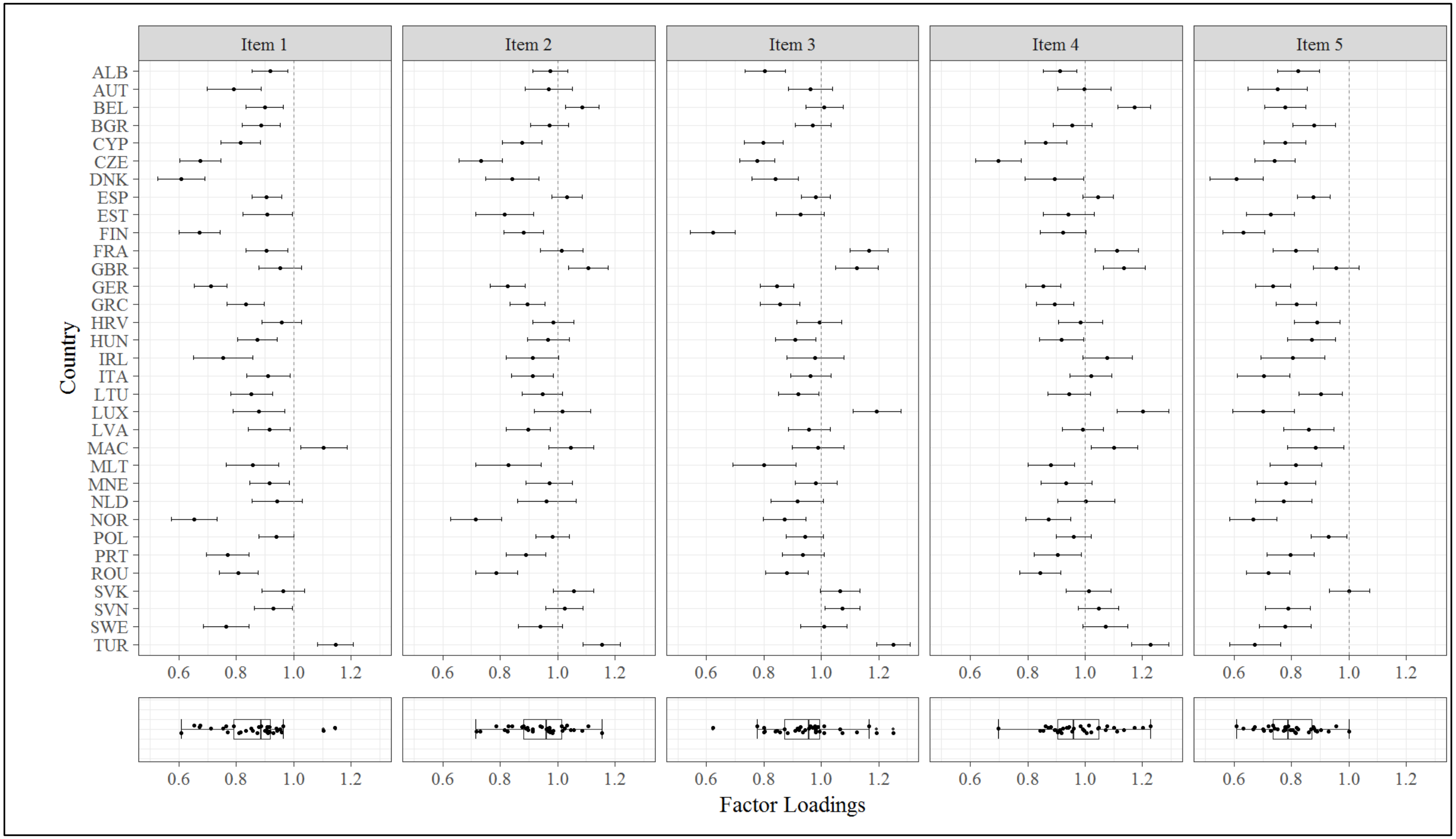
Country	χ^2	*p*	*RMSEA*	*SRMR*	*CFI*	*TLI*
ALB	23.541	.000	.061	.020	.981	.962
AUT	19.355	.002	.053	.019	.985	.969
BEL	79.461	.000	.076	.027	.967	.935
BGR	24.006	.000	.060	.016	.984	.969
CYP	6.740	.241	.019	.011	.998	.997
CZE	24.285	.000	.062	.022	.981	.962
DNK	58.834	.000	.104	.039	.936	.873
ESP	111.872	.000	.080	.028	.960	.921
EST	24.940	.000	.063	.018	.986	.972
FIN	45.805	.000	.091	.027	.959	.918
FRA	83.358	.000	.102	.034	.947	.894
GBR	28.355	.000	.054	.018	.985	.970
GER	34.899	.000	.054	.018	.984	.968
GRC	19.900	.001	.055	.016	.989	.978
HRV	15.340	.009	.046	.015	.989	.978
HUN	22.906	.000	.060	.021	.983	.966
IRL	26.803	.000	.065	.028	.970	.940

Country	χ^2	*p*	*RMSEA*	*SRMR*	*CFI*	*TLI*
ITA	18.295	.003	.044	.021	.985	.970
LTU	25.583	.000	.065	.017	.983	.967
LUX	40.638	.000	.085	.028	.961	.922
LVA	13.660	.018	.043	.014	.990	.980
MAC	0.870	.972	.000	.004	1.000	1.008
MLT	17.156	.004	.049	.024	.978	.956
MNE	28.660	.000	.069	.023	.971	.943
NLD	28.490	.000	.068	.023	.971	.942
NOR	35.956	.000	.078	.027	.968	.936
POL	10.380	.065	.031	.012	.995	.990
PRT	11.594	.041	.036	.015	.992	.985
ROU	3.786	.581	.000	.007	1.000	1.002
SVK	16.486	.006	.049	.014	.991	.981
SVN	6.691	.245	.015	.009	.999	.998
SWE	31.245	.000	.072	.025	.974	.947
TUR	41.563	.000	.061	.015	.981	.961

Notes. *MLR* estimator; *df* = 5; *RMSEA* = root mean squared error of approximation; *SRMR* = standardized root mean square residual; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index.

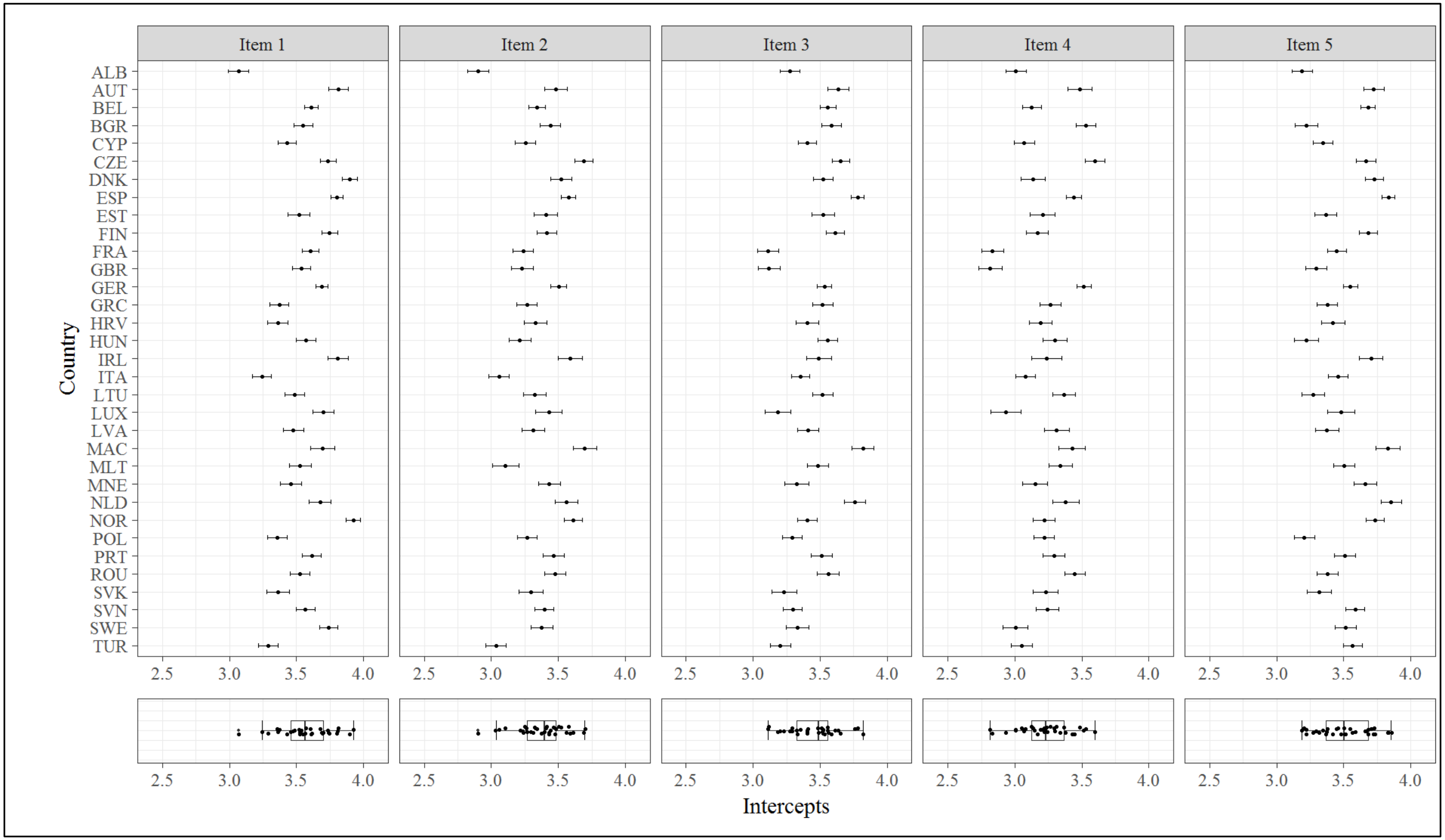
Results (II)

Figure 1. Unstandardized factor loadings with 95% CI for the one-factor WHO-5 model.



Results (III)

Figure 2. Intercepts and 95% CI for the one-factor WHO-5 model.



Results (IV)

Table 2. Test of measurement invariance and fit indices for WHO-5 one-factor model across countries.

Form of invariance	χ^2	*P*	*df*	*RMSEA*	*SRMR*	*CFI*	*TLI*
Configural invariance	978.730	0.000	165	.063	.022	.979	.959
Metric invariance	1601.885	0.000	293	.060	.063	.967	.963
Scalar invariance	4045.027	0.000	421	.083	.095	.908	.928
Δ Configural – metric	623.155		128	-.003	.041	-.012	.004
Δ Metric – scalar	2443.142		128	.023	.032	-.059	-.035

Notes. *MLR* estimator; *RMSEA* = root mean squared error of approximation; *SRMR* = standardized root mean square residual; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index.

Results (V)

Table 3. Alignment fit statistics.

	Fit function contribution	R^2	Variance	Number (percentage) of approximate MI groups	Difference of alignment and scalar model *M* (*SD*)
FL 1	-193.314	.719	0.005	29 (87.9%)	-0.015 (0.071)
FL 2	-176.841	.886	0.002	32 (97.0%)	-0.016 (0.042)
FL 3	-190.018	.743	0.006	30 (90.9%)	-0.005 (0.075)
FL 4	-184.561	.781	0.004	32 (97.0%)	0.011 (0.062)
FL 5	-209.841	.027	0.008	27 (81.8%)	-0.015 (0.094)
IC 1	-215.504	.612	0.010	21 (63.6%)	0.139 (0.102)
IC 2	-199.284	.797	0.006	23 (69.7%)	0.145 (0.082)
IC 3	-200.148	.695	0.006	16 (48.5%)	0.137 (0.080)
IC 4	-226.765	.430	0.016	14 (42.2%)	0.124 (0.130)
IC 5	-220.895	.562	0.011	22 (66.7%)	0.111 (0.106)

Notes. MLR estimator; FIXED approach; FL = Factor loadings; IC = Intercepts.

Results (VI)

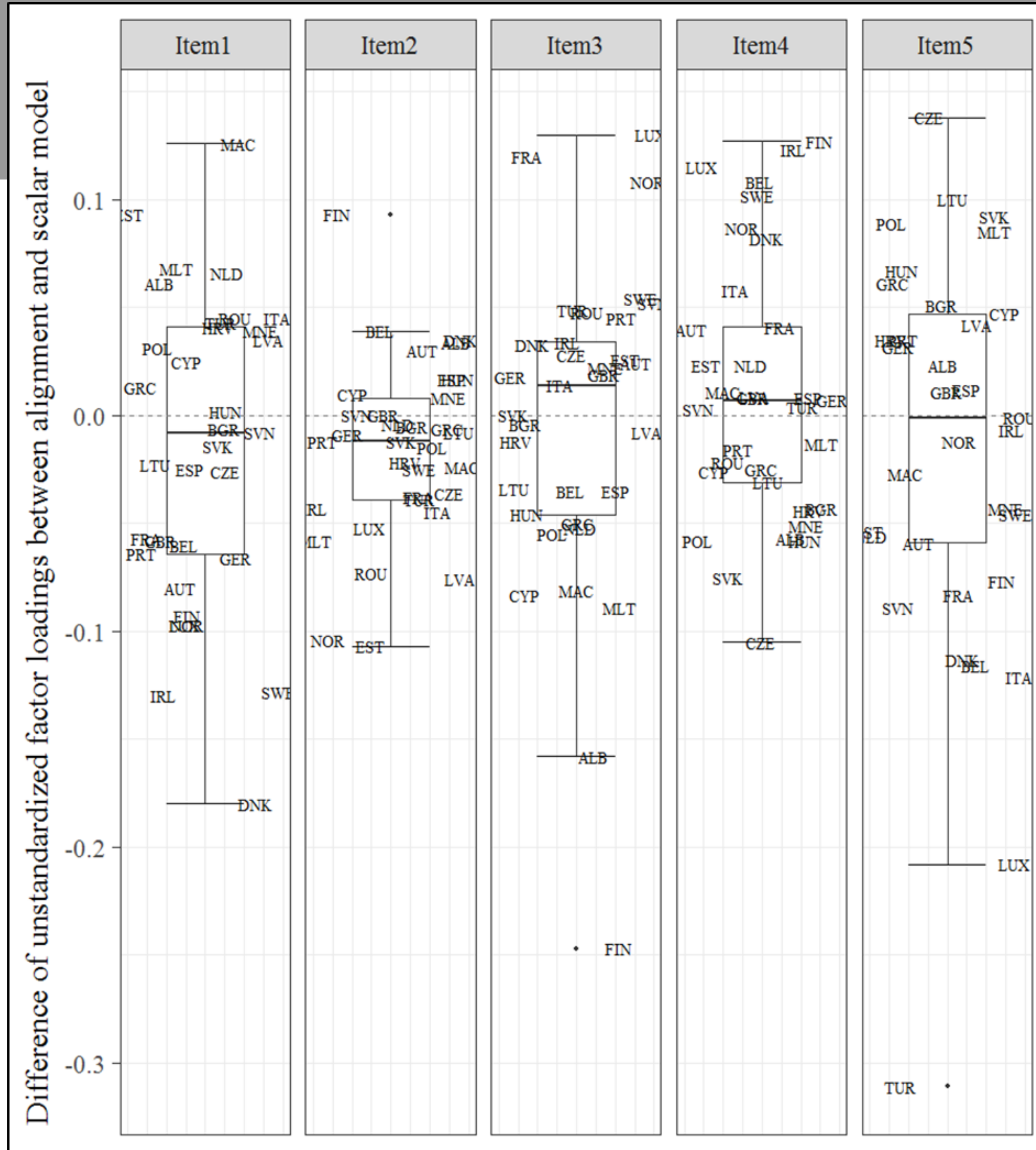


Figure 3. Differences of unstandardized factor loadings between alignment and scalar model.

Results (VII)

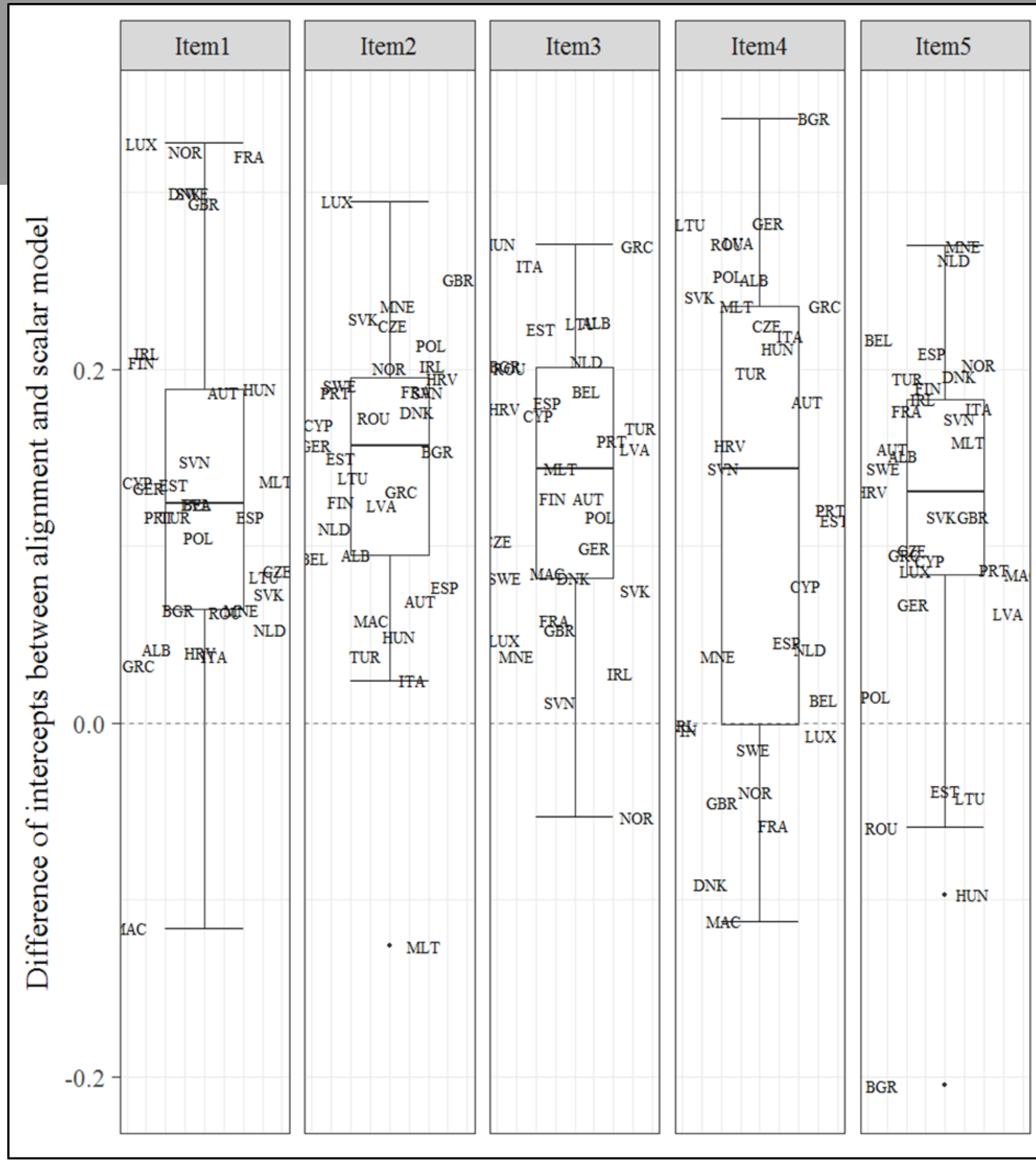
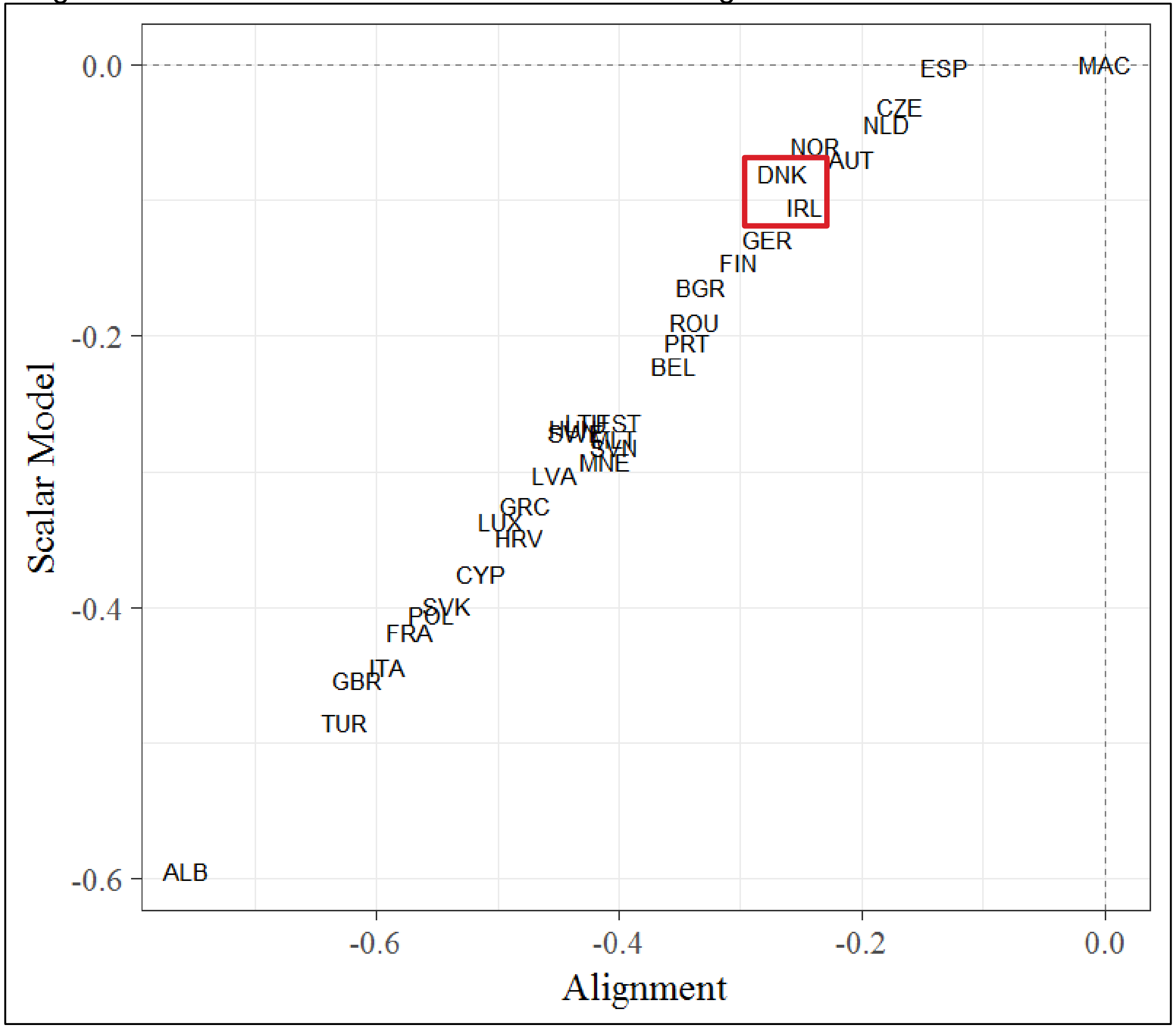


Figure 4. Differences of intercepts between alignment and scalar model.

Results (VIII)

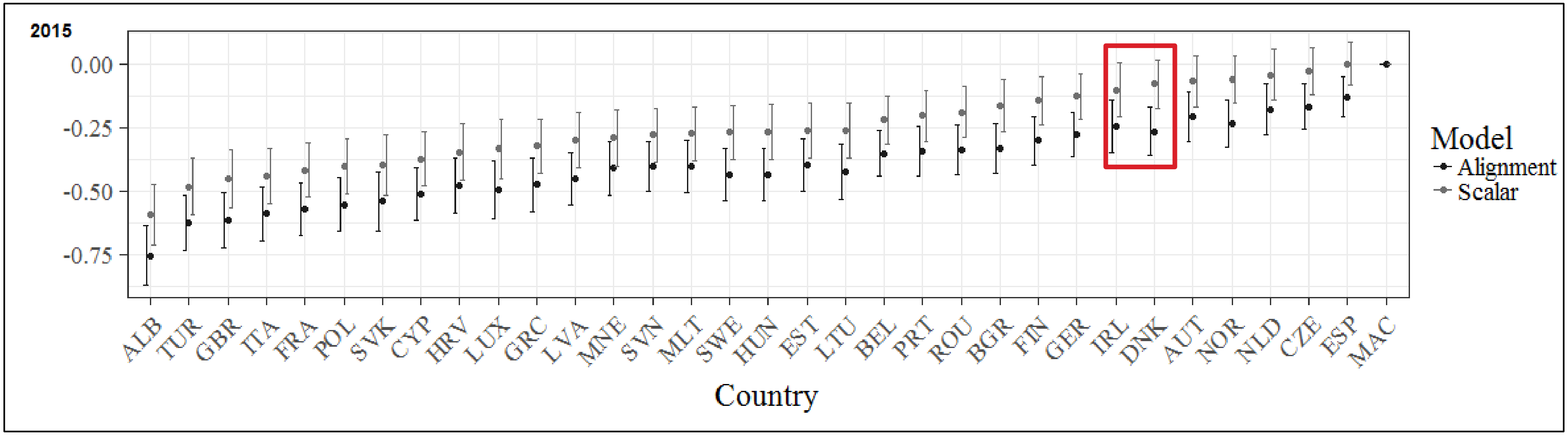
Figure 5. Differences of factor means between alignment and scalar model.



$r = 0.984$

Results (IX)

Figure 6. Differences of factor means between alignment and scalar model.



$d_{paired} = 4.75$

Discussion

Summary & conclusion

- The WHO-5 scale seems partially invariant across countries
- However, using a manifest approach or a full scalar model is probably a bad idea

- Alignment is an exploratory tool that can point out problematic indicators
- As a new tool, its performance has to be evaluated under different conditions (number of compared groups, form of non-invariance, item distribution, etc.)
- However, first simulations studies (Aspharouhov & Muthén, 2014; Flake, 2015; Marsh et al., 2017, Kim et al., 2017) showed promising results

Thank you for your attention!

Email: philipp.sischka@uni.lu

Literature

Andrews, F. M., & Withey, S. B. (1976). *Social Indicators of Well-Being: Americans' Perceptions of Life Quality*. New York: Plenum Press.

Asparouhov, T., & Muthen, B. (2014). Multiple-Group Factor Analysis Alignment. *Structural Equation Modeling, 21*, 495-508.

Bergmann, N., Ballegaard, S., Holmager, P., Kristiansen, J., Gyntelberg, F., Andersen, L. J., …, Faber, J. (2013). Pressure pain sensitivity: A new method of stress measurement in patients with ischemic heart disease. *Scandinavian Journal of Clinical and Laboratory Investigation, 73*, 373–379.

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.

Elholm, B., Larsen, K., Hornnes, N., Zierau, F., & Becker, U. (2011). Alcohol withdrawal syndrome: symptom-triggered versus fixed-schedule treatment in an outpatient setting. *Alcohol and Alcoholism, 46*, 318–323.

Literature

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling.* Advance online.

Krieger, T., Zimmermann, J., Huffziger, S., Ubl, B., Diener, C., Kuehner, C., & Grosse Holtforth, M. (2014). Measuring depression with a well-being index: further evidence for the validity of the WHO Well-Being Index (WHO-5) as a measure of the severity of depression. *Journal of Affective Disorders, 156*, 240–244.

Lai, M. H. C., & Yoon, M. (2015). A modified comparative fit index for factorial invariance studies. *Structural Equation Modeling, 22*, 236-248.

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.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

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.

Oberski, D. L. (2014). Evaluating sensitivity of parameters of interest to measurement invariance in latent variable models. *Political Analysis, 22*, 45-60.

Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A Systematic Review of the Literature. *Psychotherapy and Psychosomatics, 84*, 167–176.

Appendix (I)

Table 1. Sample size, percent females, mean and standard deviation of age, and reliability (McDonald's Omega).

	2010				2015			
Country	N	% female	Age M (SD)	ω	N	% female	Age M (SD)	ω
ALB	941	43.7	41.6 (11.9)	.90	995	53.1	39.7 (13.2)	.91
AUT	938	47.9	40.1 (12.3)	.82	1017	51.9	41.6 (12.5)	.88
BEL	3904	45.4	40.4 (11.0)	.85	2578	47.2	42.0 (11.8)	.87
BGR	993	47.2	41.8 (11.5)	.93	1057	50.1	43.3 (11.9)	.92
CYP	989	44.7	41.0 (12.0)	.91	995	48.8	38.6 (12.5)	.89
CZE	964	43.1	41.5 (11.6)	.90	990	50.4	43.0 (11.8)	.87
DNK	1061	47.3	40.4 (13.2)	.74	997	47.2	42.7 (13.5)	.81
ESP	1001	43.2	39.8 (11.0)	.88	3341	47.2	42.0 (11.1)	.90
EST	959	51.6	42.3 (12.7)	.86	990	54.0	43.5 (13.4)	.88
FIN	1016	49.1	42.0 (12.7)	.81	995	50.8	45.0 (12.4)	.84
FRA	3015	47.5	40.3 (11.3)	.88	1520	49.4	41.9 (11.5)	.86
GBR	1548	46.5	40.7 (13.2)	.87	1611	46.5	41.8 (13.5)	.89
GER	2104	46.4	41.4 (12.3)	.85	2076	49.6	43.8 (13.1)	.87
GRC	1029	39.7	41.2 (11.3)	.90	998	43.4	42.3 (11.6)	.90
HRV	1069	45.9	42.6 (12.3)	.92	997	49.0	42.7 (12.3)	.91
HUN	1006	45.9	40.8 (11.2)	.87	1010	51.9	43.9 (11.8)	.89
IRL	993	45.8	39.6 (12.2)	.86	1043	46.7	42.4 (12.4)	.87
ITA	1457	39.8	41.3 (11.1)	.88	1390	46.2	45.0 (11.7)	.87
LTU	939	52.0	41.4 (11.9)	.91	989	53.9	43.9 (12.6)	.92
LUX	972	42.8	39.9 (10.9)	.85	991	48.1	41.2 (10.7)	.85
LVA	985	51.5	41.3 (12.6)	.83	941	52.9	42.7 (13.1)	.88
MAC	1090	38.7	40.2 (11.3)	.89	1001	47.2	40.9 (12.9)	.90
MLT	991	33.1	37.7 (12.3)	.83	999	40.3	39.6 (13.0)	.84
MNE	995	43.2	39.4 (11.9)	.92	999	47.2	39.9 (12.5)	.90
NLD	1013	45.8	40.3 (13.2)	.81	1024	47.5	42.0 (13.9)	.88
NOR	1076	47.5	41.3 (13.3)	.82	1025	49.4	41.8 (13.8)	.82
POL	1435	45.0	39.4 (11.8)	.92	1128	53.0	41.8 (13.0)	.91
PRT	997	46.8	42.4 (13.1)	.92	1008	53.2	45.6 (13.1)	.87
ROU	965	43.1	41.0 (12.5)	.88	1052	48.7	41.2 (11.7)	.87
SVK	987	44.1	40.3 (11.3)	.91	954	50.6	41.9 (11.5)	.94
SVN	1384	45.8	39.9 (11.5)	.88	1586	49.5	42.4 (11.5)	.89
SWE	969	48.2	42.7 (13.0)	.81	1002	48.0	43.3 (13.1)	.86
TUR	2085	27.5	36.4 (12.4)	.91	1991	29.4	37.3 (12.2)	.88