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.

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

5. Evaluation of measurement invariance

- Δχ² difference test
- Changes in approximate fit statistics
  - ΔSRMR
  - ΔRMSEA
  - ΔCFI
- Most common: ΔCFI (Cheung & Rensvold, 2002; Mead, Johnson, & Braddy, 2008)
- Modified CFI (Lai, & Yoon, 2015)
- Magnitude of difference between the parameter estimates (Oberski, 2014)

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)

9. Results (I)

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

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)

4. Order of testing forms of measurement invariance

- Constrained-baseline strategy
  - Starts with a “fully-constrained” baseline model (fixed loadings and intercepts)
  - 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)

6. Evaluation of sources of non-invariance

- Statistical significance of the Δχ² 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.

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

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).

<table>
<thead>
<tr>
<th>Form of invariance</th>
<th>χ²</th>
<th>df</th>
<th>p</th>
<th>RMSEA (90% CI)</th>
<th>SRMR</th>
<th>CFI</th>
<th>TLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configural invariance</td>
<td>444.355</td>
<td>300</td>
<td>.000</td>
<td>[.059, .060]</td>
<td>.047</td>
<td>.989</td>
<td>.997</td>
</tr>
<tr>
<td>Metric invariance</td>
<td>1380.096</td>
<td>300</td>
<td>.000</td>
<td>[.077, .080]</td>
<td>.046</td>
<td>.980</td>
<td>.977</td>
</tr>
<tr>
<td>Scalar invariance</td>
<td>5677.383</td>
<td>300</td>
<td>.000</td>
<td>[.086, .089]</td>
<td>.045</td>
<td>.916</td>
<td>.949</td>
</tr>
<tr>
<td>Partial scalar invariance*</td>
<td>1723.605</td>
<td>300</td>
<td>.000</td>
<td>[.064, .069]</td>
<td>.050</td>
<td>.972</td>
<td>.972</td>
</tr>
<tr>
<td>Δ Configural - metric</td>
<td>-611.739</td>
<td>300</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Δ Metric - scalar</td>
<td>2197.267</td>
<td>300</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

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; * = 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.