Using a Multilevel Structural Equation Modeling Approach to Explain Cross-Cultural Measurement Noninvariance

Copyright Davidov et al. & CSDI

Eldad Davidov, University of Zurich
Hermann Dülmer, University of Cologne
Elmar Schlüter, University of Cologne
Peter Schmidt, University of Marburg
Overview

1. Introduction
2. What do we mean by invariance?
3. What to do when there is no invariance?
4. Using multilevel techniques to explain measurement noninvariance
5. Empirical illustration
6. Conclusions
1. Introduction

More and more *studies test* the *invariance* (equivalence) *properties* of scales across

- countries,
- cultural groups,
- time points,
- modes of data collection etc.

There are **two** typical scenarios:

- We analyze and find high levels of invariance and we are **happy**.
- Invariance across groups is absent and we are **disappointed** and question, whether it makes sense to continue comparisons.
1. Introduction

In this study we propose using

- mulilevel structural equation modeling (ML SEM)
- to explain noninvariance,
- not to solve it (diagnosis of the problem).

We will exemplify its use with an example from the European Social Survey (ESS).
2. What do we mean by invariance

There are typically three important levels of testing measurement invariance:

● Configural Invariance,
● Metric Invariance,
● Scalar Invariance.
2. What do we mean by invariance

Measurement Invariance:

Group A (Culture, country, time point)

<table>
<thead>
<tr>
<th>Item a</th>
<th>α1A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item b</td>
<td>α2A</td>
</tr>
<tr>
<td>Item c</td>
<td>α3A</td>
</tr>
</tbody>
</table>

Item a → β1A → Univers alism

β2A → Item c

Group B (Culture, country, time point)

<table>
<thead>
<tr>
<th>Item a</th>
<th>α1B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item b</td>
<td>α2B</td>
</tr>
<tr>
<td>Item c</td>
<td>α3B</td>
</tr>
</tbody>
</table>

Item a → β1B → Univers alism

β2B → Item c

All levels of invariance are necessary for meaningful mean comparisons.
2. What do we mean by invariance

But:

Higher levels of invariance (i.e., scalar)
- are *seldom* guaranteed,
- especially when many groups are compared.
3. What to do when there is no invariance?

1) Resort to partial invariance
   (Byrne/Shavelson/Muthén 1989; Steenkamp/Baumgartner 1998).

2) Compare only a subset of countries (or other groups)
   where invariance of the involved concepts does not hold
   (e.g., Byrne/van de Vijver 2010).

3) Even refrain from cross-country comparisons
   when invariance is not found even across a subset of countries.

4) Decrease the number of items and delete those items whose
   parameters are very different across groups. However, when
   this approach is applied, one has to address the question
   whether the meaning of the concept has changed after the
   item reduction (Byrne/van de Vijver 2010).
2. What to do when there is no invariance?

But:

None of these solutions explains where the problem comes from!
4. Using multilevel techniques to explain measurement noninvariance

5) Explain noninvariance by introducing contextual variables and using multilevel analysis (Schlüter/Meulemann 2009).

Noninvariance can be viewed as a useful source of information on cross-cultural differences (e.g., Poortinga 1989; Schlüter/Meulemann 2009).

Although it was already referred to by some authors (e.g., Hox/de Leeuw/Brinkhuis 2010) and although the technique is not new (e.g., Muthén 1985) to the best of our knowledge this possibility has not yet been explicated and applied for the goal of explaining noninvariance systematically.
5. Empirical illustration

Theory of 10 Universal Values (Schwartz 1992)

- Focus of our Analysis: Universalism

- Universalism is measured by 3 Items:
  - Equality
  - Understanding Different People
  - Environment Protection

Data Source: ESS 2004-5
5. Empirical illustration

Test of invariance of universalism across 25 ESS countries by using multiple group confirmatory factor analysis (MGCFA)

Results:

- **Configural and metric invariance** were confirmed by the data:
  - Factor loadings are equal across countries.

- **Scalar invariance was rejected:**
  - Indicator intercepts are different. People use the scale differently.

More specifically:

- The intercept of the indicator Environment is particularly different across countries.
5. Empirical illustration

Theoretical Expectations (Inglehart 1990, 1997):

H1: The value of Universalism is more important in Postmodern advanced industrial countries than in less developed Modern countries.

- Intergenerational value change from Materialist to Postmaterialist value priorities in advanced Modern societies (Inglehart 1990):
  - Postmaterialist needs (belonging, self-expression) become more important than more fundamental Materialist needs (physical/economical security).

- Shift from Materialist to Postmaterialist value priorities is a key indicator for a broader syndrome of Postmodernization (Inglehart 1997):
  - Postmodernization includes a shift from an emphasis on self-denying achievement orientations toward individual choice of lifestyles, greater tolerance for ethic, cultural and sexual diversity, and an increasing emphasis on protection of environment and quality of life in general.
5. Empirical illustration

Theoretical Expectations (Inglehart 1990, 1997):

H2: Environment protection is perceived more important in less developed countries than in Postmodern advanced industrial countries.

- Inglehart (1997: 242) also states that in less developed countries where air/water pollution are far worse than in advanced industrial societies, environment protection is less a quality of life issue but much more a matter of physical health.
5. Empirical illustration

Theoretical Expectations (Inglehart 1990, 1997):

Thus:
People may ‘react’ differently to the environment item, because of differences in the level of human development (HDI) between the countries.

Human Development Index (HDI 2004):
Geometric Mean of
- Standard of Living: Logarithm of the GDP pc in PPP $US
- Educational Attainment: Index for Adult Literacy Rate and Gross Enrolment Rate in Primary/Secondary/Tertiary Schools
- Health: Life Expectation at Birth
5. Empirical illustration

1st step: Two-Level Confirmatory Factor Analysis

Country Level: Between

Respondent Level: Within

AIC: 368050.207
RMSEA: .003
SRMR (within): .000
SRMR (between): .062
5. Empirical illustration

2nd step: Two-Level Structural Equation Model

Country Level: Between

- HDI 2004
- Universalism Between
  - Environment
  - Underst. Diff. People
  - Equality

Respondent Level: Within

- AIC: 368042.483
- RMSEA: .000
- SRMR (within): .000
- SRMR (between): .045

Note: *: p<.05
6. Conclusion

Our empirical results

confirm our hypotheses concerning the impact of a country’s level of economic/technical development (HDI)

- on the value of Universalism (H1)
- on the citizens attitude toward environment protection (H2, intercept noninvariance)

The level of human development explains why scalar invariance was absent for the item Environment.
6. Conclusion

- ML SEM is a rather straightforward technique to explain variation in an item’s intercept.

- Although ML SEM does not provide a resolution to the problem of noninvariance, it is a useful tool for the diagnosis of the causes.

- Diagnosis and systematic identification of the sources of item bias are an essential step before measurement scales can be improved for cross-cultural analysis.
Thank you very much for your attention!!!
## Descriptive Findings

### Variances, Covariances, and ICC for the Indicators of Universalism

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Pooled Sample Correlations and Covariances</th>
<th>Within and Between Countries Correlations and Covariances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Equality (ipeqopt)</td>
<td>1.079</td>
<td>.344</td>
</tr>
<tr>
<td>Understanding Diff. People (ipudrst)</td>
<td>.385</td>
<td>1.162</td>
</tr>
<tr>
<td>Environment (impev)</td>
<td>.347</td>
<td>.372</td>
</tr>
<tr>
<td></td>
<td>within</td>
<td></td>
</tr>
<tr>
<td>Equality (ipeqopt)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding Diff. People (ipudrst)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment (impev)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


| ICC                                | .036          | .035          | .046          |

Note: Bold entries in the upper diagonal are the correlations, entries in the diagonal are variances, and entries in the lower diagonal are covariances; the total sample includes 43,779 respondents from 26 countries.
## Two-Level CFA and Two-Level SEM for Universalism

<table>
<thead>
<tr>
<th>Respondent Level (n = 43,779)</th>
<th>Model 1: Two-Level CFA</th>
<th>Model 2: Multilevel SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country Level (n = 26)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| AIC                           | 368050.207             | 368042.483              |
| SRMR Within                   | .062                   | .045                    |
| SRMR Between                  | b                      | z                       |

### Confirmatory Factor Analysis

**Intercept Level 2**
- Equality (ipeqopt): 3.894, 100.478**
- Underst. Diff. People (ipudrst): 3.594, 90.518**
- Environment (impenv): 3.832, 87.229**

**Factor Loadings**

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>z</th>
<th>beta</th>
<th>b</th>
<th>z</th>
<th>beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equality (ipeqopt)</td>
<td>1.00</td>
<td>-.</td>
<td>1.00</td>
<td>1.00</td>
<td>-.</td>
<td>.707</td>
</tr>
<tr>
<td>Underst. Diff. People (ipudrst)</td>
<td>.608</td>
<td>3.666**</td>
<td>.592</td>
<td>.921</td>
<td>3.197**</td>
<td>.637</td>
</tr>
<tr>
<td>Environment (impenv)</td>
<td>.625</td>
<td>3.277**</td>
<td>.549</td>
<td>1.747</td>
<td>4.599**</td>
<td>1.090</td>
</tr>
<tr>
<td>Equality (ipeqopt)</td>
<td>1.00</td>
<td>-.</td>
<td>.568</td>
<td>1.00</td>
<td>-.</td>
<td>.568</td>
</tr>
<tr>
<td>Underst. Diff. People (ipudrst)</td>
<td>1.069</td>
<td>57.275**</td>
<td>.585</td>
<td>1.069</td>
<td>57.275**</td>
<td>.585</td>
</tr>
<tr>
<td>Environment (impenv)</td>
<td>.960</td>
<td>58.203**</td>
<td>.550</td>
<td>.960</td>
<td>58.202**</td>
<td>.550</td>
</tr>
</tbody>
</table>

### Regression

- Universalism (betw.) on HDI 2004: 1.165, 1.871*, .449

**Note:** *p ≤ 0.05; **p ≤ 0.01; Estimator: Full Maximum Likelihood (ML); Estimates for level 2 parameters are indented to the right in the first column. Variances/Residual Variances tested one-tailed. Since we formulated hypotheses for the impact of the HDI on environment and universalism (between), the significance level of both b-coefficients are based on a one-tailed test. The beta-coefficients are fully standardized.
## Two-Level CFA and Two-Level SEM for Universalism

<table>
<thead>
<tr>
<th>Respondent Level (n = 43,779)</th>
<th>Model 1: Two-Level CFA</th>
<th>Model 2 Multilevel SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country Level (n = 26)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>368050.207</th>
<th>368042.483</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRMR Within</td>
<td>.062</td>
<td>.045</td>
</tr>
<tr>
<td>SRMR Between</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Residual Variances

<table>
<thead>
<tr>
<th></th>
<th>Variance</th>
<th>z</th>
<th>Variance</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equality (ipeqopt)</td>
<td>-.</td>
<td>-.</td>
<td>.019</td>
<td>3.478**</td>
</tr>
<tr>
<td>Underst. Diff. People (ipudrst)</td>
<td>.026</td>
<td>3.504**</td>
<td>.024</td>
<td>3.477**</td>
</tr>
<tr>
<td>Environment (impenv)</td>
<td>.035</td>
<td>3.511**</td>
<td>-.</td>
<td>-.</td>
</tr>
<tr>
<td>Equality (ipeqopt)</td>
<td>.703</td>
<td>93.393**</td>
<td>.703</td>
<td>93.391**</td>
</tr>
<tr>
<td>Underst. Diff. People (ipudrst)</td>
<td>.735</td>
<td>88.396**</td>
<td>.735</td>
<td>88.395**</td>
</tr>
<tr>
<td>Environment (impenv)</td>
<td>.711</td>
<td>98.524**</td>
<td>.711</td>
<td>98.525**</td>
</tr>
</tbody>
</table>

### Variance

<table>
<thead>
<tr>
<th></th>
<th>Variance</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent Factor (Universalism betw.)</td>
<td>.038</td>
<td>3.542**</td>
</tr>
<tr>
<td>Latent Factor (Universalism within)</td>
<td>.334</td>
<td>42.894**</td>
</tr>
</tbody>
</table>

### Variance Comp./Residual Var. Level 2

<table>
<thead>
<tr>
<th></th>
<th>Variance</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept Level 2: Universalism (betw.)</td>
<td>.015</td>
<td>1.943*</td>
</tr>
<tr>
<td>Level 1: Universalism (within)</td>
<td>.334</td>
<td>42.894**</td>
</tr>
</tbody>
</table>

Note: * p ≤ 0.05; ** p ≤ 0.01; Estimator: Full Maximum Likelihood (ML);

Estimates for level 2 parameters are indented to the right in the first column. Variances/Residual Variances tested one-tailed. Since we formulated hypotheses for the impact of the HDI on environment and universalism (between), the significance level of both b-coefficients are based on a one-tailed test.
Two-Level CFA

Respondent Level (within):  
\[ y_{ijk} = \alpha_{jk} + \lambda_{wk} \cdot \eta_{Wij} + \varepsilon_{Wijk} \]

Country Level (between):  
\[ \alpha_{jk} = \nu_k + \lambda_{Bk} \cdot \eta_{Bj} + \varepsilon_{Bjk} \]

where
- \( y_{ijk} \) refers to the observed value of respondent \( i \) of country \( j \) on indicator variable \( k \),
- \( \alpha_{jk} \) refers to the intercept of country \( j \) on indicator variable \( k \),
- \( \nu_k \) refers to the intercept (usually called grand mean in multilevel analysis) of indicator variable \( k \),
- \( \eta_{Wij} \) refers to the score of respondent \( i \) of country \( j \) on the within-level latent \( \eta_W \),
- \( \eta_{Bj} \) refers to the score of country \( j \) on the between-level latent variable \( \eta_B \),
- \( \lambda_{wk} \) refers to the within-level factor loading \( \lambda_W \) of indicator variable \( k \),
- \( \lambda_{Bk} \) refers to the between-level factor loading \( \lambda_B \) of indicator variable \( k \),
- \( \varepsilon_{Wijk} \) refers to the within-level error term \( \varepsilon_W \) for respondent \( i \) of country \( j \) on indicator variable \( k \), and
- \( \varepsilon_{Bjk} \) refers to the between-level error term \( \varepsilon_B \) (usually called random term in multilevel analysis) for country \( j \) on indicator variable \( k \).
Two-Level CFA

\[ y_{ijk} = \nu_k + \lambda_{Bk} \cdot \eta_{Bj} + \varepsilon_{Bjk} + \lambda_{Wk} \cdot \eta_{Wij} + \varepsilon_{Wijk} \]

where

- \( y_{ijk} \) refers to the observed value of respondent \( i \) of country \( j \) on indicator variable \( k \),
- \( \eta_{Bj} \) refers to the score of country \( j \) on the between level latent variable \( \eta_B \),
- \( \eta_{Wij} \) refers to the score of respondent \( i \) of country \( j \) on the within level latent \( \eta_W \),
- \( \nu_k \) refers to the intercept (usually called grand mean in multilevel analysis) of indicator variable \( k \),
- \( \lambda_{Bk} \) refers to the between level factor loading \( \lambda_B \) of indicator variable \( k \),
- \( \lambda_{Wk} \) refers to the within level factor loading \( \lambda_W \) of indicator variable \( k \),
- \( \varepsilon_{Bjk} \) refers to the between level error term \( \varepsilon_B \) (usually called random term or \( u_{jk} \)-term in multilevel analysis) for country \( j \) on indicator variable \( k \), and
- \( \varepsilon_{Wijk} \) refers to the within level error term \( \varepsilon_W \) for respondent \( i \) of country \( j \) on indicator variable \( k \).
### One-Level CFA vs. Two-Level CFA for Universalism

<table>
<thead>
<tr>
<th>Respondent Level (n = 43,779)</th>
<th>Model 1a: One-Level CFA</th>
<th>Model 1b: Two-Level CFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country Level (n = 26)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>1a</th>
<th>1b</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>372209.497</td>
<td>368050.207</td>
</tr>
<tr>
<td>SRMR Within</td>
<td></td>
<td>.062</td>
</tr>
<tr>
<td>SRMR Between</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Intercept Level 1
- Intercept Level 2
  - Equality (ipeqopt) 3.894 100.478**
  - Underst. Diff. People (ipudrst) 3.594 90.518**
  - Environment (impenv) 3.832 87.229**
- Equality (ipeqopt) 3.890 783.571**
- Underst. Diff. People (ipeqopt) 3.589 696.455**
- Environment (impenv) 3.845 779.247**

#### Factor Loadings
- Equality (ipeqopt) 1.000 -. .578 1.000 -. .568
- Underst. Diff. People (ipudrst) .608 3.666** .592
- Environment (impenv) .625 3.277** .549
- Equality (ipeqopt) 1.000 -. .578 1.000 -. .568
- Underst. Diff. People (ipudrst) 1.069 60.117** .596 1.069 57.275** .585
- Environment (impenv) .964 60.997** .561 .960 58.203** .550

* p ≤ 0.05; ** p ≤ 0.01; Estimator: Full Maximum Likelihood (ML); the beta-coefficients are fully standardized; the residual variance of equality at level 2 turned out to be insignificant and has been fixed for that reason.
One-Level CFA vs. Two-Level CFA for Universalism

<table>
<thead>
<tr>
<th>Respondent Level (n = 43,779)</th>
<th>Country Level (n = 26)</th>
<th>Model 1a: One-Level CFA</th>
<th>Model 1b: Two-Level CFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>372209.497</td>
<td>368050.207</td>
<td></td>
</tr>
<tr>
<td>SRMR Within</td>
<td>.062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRMR Between</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual Variances</td>
<td>Variance</td>
<td>z</td>
<td>Variance</td>
</tr>
<tr>
<td>Equality (ipeqopt)</td>
<td>-.-</td>
<td>-.-</td>
<td>-.-</td>
</tr>
<tr>
<td>Underst. Diff. People (ipudrst)</td>
<td>.026</td>
<td>3.504**</td>
<td>.035</td>
</tr>
<tr>
<td>Environment (ipeqopt)</td>
<td>.035</td>
<td>3.511**</td>
<td>.035</td>
</tr>
<tr>
<td>Equality (ipeqopt)</td>
<td>.718</td>
<td>93.533**</td>
<td>.703</td>
</tr>
<tr>
<td>Underst. Diff. People (ipudrst)</td>
<td>.750</td>
<td>88.392**</td>
<td>.735</td>
</tr>
<tr>
<td>Environment (ipeqopt)</td>
<td>.730</td>
<td>98.486**</td>
<td>.711</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latent Factor (Universalism betw.)</td>
<td>.038</td>
<td>3.542**</td>
<td>.334</td>
</tr>
<tr>
<td>Latent Factor (Universalism within)</td>
<td>.360</td>
<td>44.711**</td>
<td>.334</td>
</tr>
</tbody>
</table>

*p ≤ 0.05; ** p ≤ 0.01; Estimator: Full Maximum Likelihood (ML); the residual variance of equality at level 2 turned out to be insignificant and has been fixed for that reason
Intercepts and Slopes:
Equal Slopes and Unequal Intercepts
Using mulilevel technique to explain measurement invariance

- The basic idea behind a two-level confirmatory factor analysis is to decompose the variability of the indicator variables into a respondent-level “within” variability and a country-level (group level) “between” variability.

- It allows, in a subsequent step, to account for differences in the parameters between groups by including contextual variables.

- In this way, the multilevel CFA (cf. Muthén 1994; Hox 2002) is extended to a multilevel SEM (cf. Muthén 1994; Selig et al. 2008), which allows the explanation of measurement noninvariance by contextual variables.