# Assessing measurement invariance: Can we make a dead-end road into a highway?

Eldad Davidov, Daniel Seddig, Jan Cieciuch, Peter Schmidt

Fourth Annual Likert Workshop
Cross-Cultural Psychology and Survey Research
March 11, 2022
University of Michigan

**Plan of the presentation**

**1) Measurement invariance**

**2) Exact and approximate measurement invariance**

**3) Alignment optimization**

**4) Application: value measurement in the European Social Survey and its invariance properties (e.g., Davidov, Schmidt and Schwartz, 2008)**

**1) Measurement invariance**
- psychometric property of a questionnaire

The questionnaire is measurement invariant
when it measures
- **the same construct**
- **in the same way**
- **across different groups**, such as countries, cultures or other geographical regions, conditions of data collection or time points

**Measurement invariance**
is a precondition for any meaningful comparison of means, correlates and regression coefficients of the measured construct across groups (e.g., Meredith 1993)

**Examples for measurement noninvariance**

- **Democracy attitudes** in the **World Value Study**
- **Immigration attitudes** in **Denmark** in the **European Social Survey**
- **Religiosity** in **Turkey** in the **European Social Survey**

**International surveys**

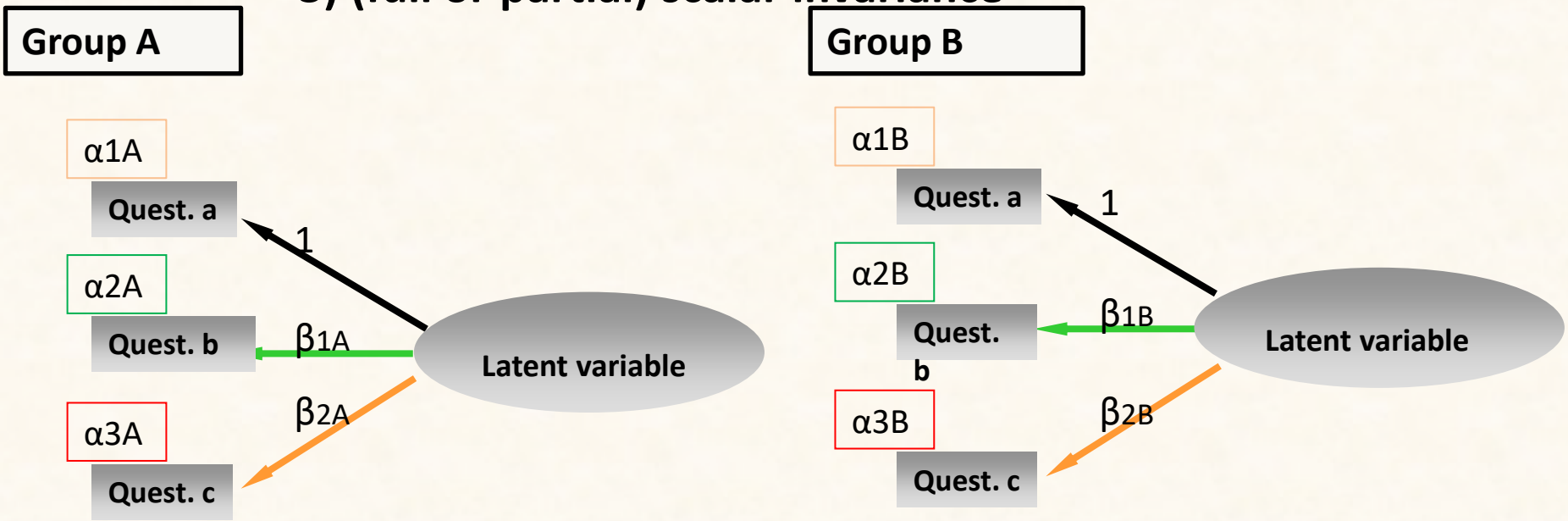
Try to achieve invariance by similar modes and data collection procedures, or careful translations.

But – we do not know if these efforts are successful, if we do not examine the measurement invariance properties.

## 2) The most often used quantitative approach to test for exact measurement invariance:

**Multigroup Confirmatory Factor Analysis - MGCFA** (Bollen 1989, Jöreskog 1971)

1) configural invariance
2) (full or partial) metric invariance
3) (full or partial) scalar invariance




**Model fit comparison between more and less constrained models**

**The problem with this approach:**

**…it establishes rather rarely full invariance…**

**As a consequence…**

**…comparisons across groups may not be allowed**

**Maybe the method examining for exact full or partial invariance is too strict?**

# Daniel Seddig
# Approximate measurement invariance
## What does „approximate” mean?

- Instead of assuming that parameter differences across groups are **exactly zero**, parameter differences in this approach follow a **distribution**, allowing for some “**wiggle room**”
- The distribution may be assumed to have a **mean of zero** and a **small variance**
- The variance indicates the degree to which parameter differences across groups are **allowed to deviate from zero**

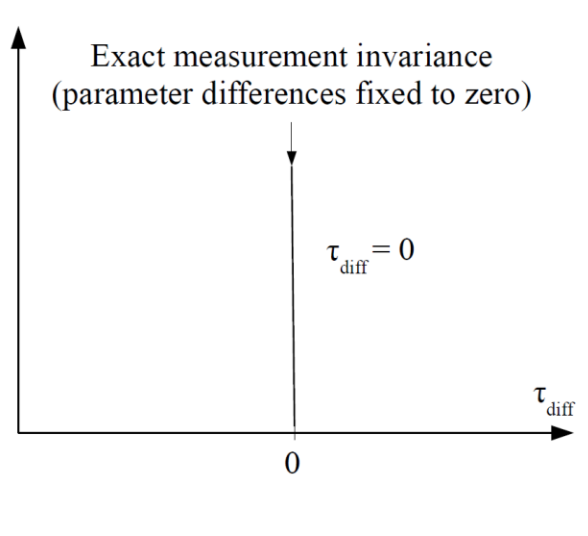
**(Muthén & Asparouhov, 2013; Van de Schoot et al., 2013)**

# Approximate measurement invariance

**Exact measurement invariance**

$$\tau^{g} - \tau^{g'} \sim N(0,0)$$

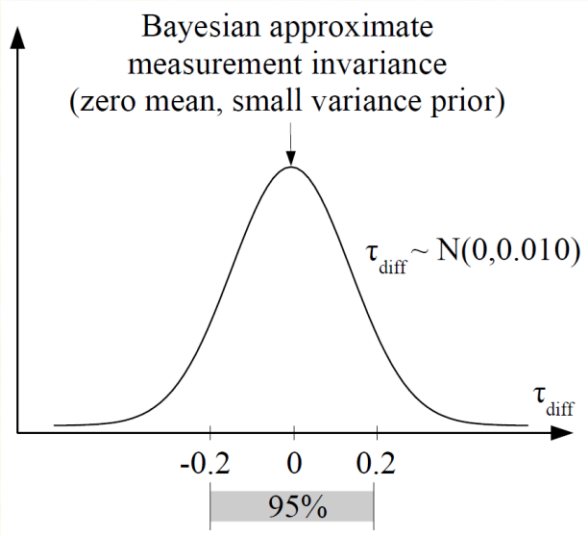
**mean = zero**
**variance = zero**




**Approximate measurement invariance**

$$\tau^{g} - \tau^{g'} \sim N(0,\sigma)$$

**mean = zero**
**variance = $\sigma$**



# Approximate measurement invariance

- The smaller the variance $\sigma$ the more it "pushes" small parameter differences across groups closer to zero
  - small = highly informative
  - large = less informative

**How large may the variance be?**

- Suggestions based on simulations (Van de Schoot et al., 2013)

**variance of .05 for the parameter differences**

→ **small enough** to not distort the substantive results (e.g., results of latent mean comparisons)

*and*

→ **large enough** to make the assumption of approximate equality realistic

# 3) Alignment optimizatzion

- **An alternative approach in the framework of MGCFA (Asparouhov & Muthen, 2013)**

- **Main idea:**

**→ estimating latent means and variance**

**but *without***

**→ constraining loadings and intercepts to be equal across groups**

# Alignment optimizatzion

***The first step***

**estimate a configural model** with:
- freely estimated intercepts and loadings,
- factor means fixed to zero,
- factor variances fixed to one

Result: a **few large non-invariant parameters** and **many approximately invariant parameters** (rather than many medium-sized non-invariant parameters
→ **Similarity to EFA**:simple factor structure)

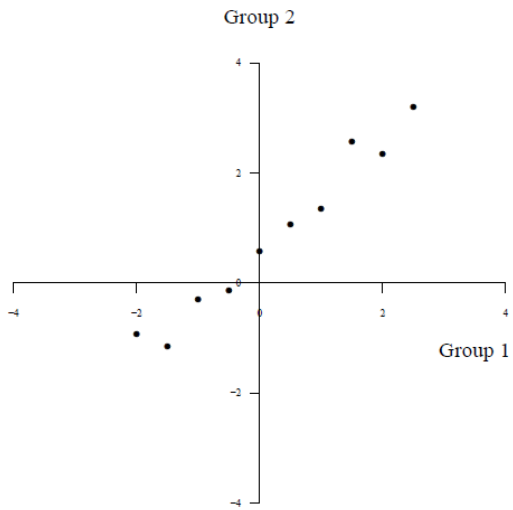
***The second step = aligned model***

**freeing the factor variances and means**
and estimating them in a way
that the **total amount of non-invariance is minimized**
→ *means and variances are chosen to minimize the number of noninvariant loadings and intercepts*

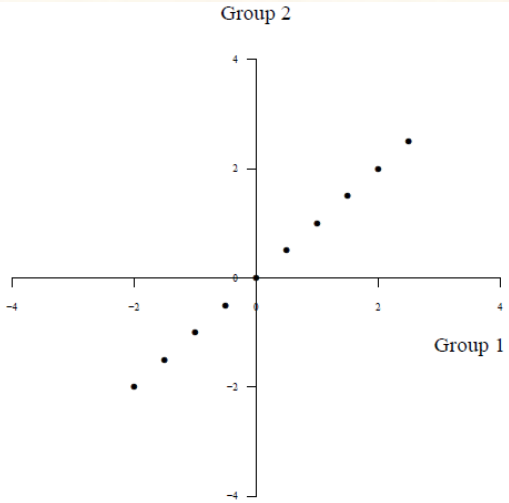
**The same model fit as the configural**

= the Alignment takes into account
(1) differences in loadings and intercepts estimated in the first step
(2) while estimating factor means in the second step.

# Unaligned and aligned intercepts parameters




Unaligned: Configural model (mean=0, variance=1 in both groups)




Aligned: Taking into account the group differences in means and variances

# Alignment optimizatzion

**Frequentiest approach (ML)**

**Bayesian approach**
**two versions**

**First step:**
**configural model**

**First step:**
**non-informative priors**

***(similar to configural model in the frequentiest approach)***

**First step:**
**approximate scalar model**

***(Intercepts and loadings constrained to be approximately invariant; means and variances estimated in all groups but first)***

**Second step:**
**freeing the factor variances and means** and
estimating them in a way that the **total amount of non-invariance is minimized**

**Advantages of Bayesian alignment estimation over ML**

- possibility of using binary indicator variables
- more chance for better model fit at the configural level
- more chance to resolve estimation problems
- measures of approximate measurement invariance

**The advantage of the Bayesian approximate model with alignment estimation over the Bayesian approximate model without alignment…**

*… alignment estimates are obtained by minimizing the number of non-invariant items, while BSEM estimates are obtained by minimizing the variability of estimates across groups → alignment estimates will be simpler to interpret as fewer non-invariant parameters will be found.* (Asparouhov & Muthen, 2013)

But: Bayesian analysis may take a lot of time (sometimes weeks…) and alignment works only on a single factor.

- Simulations (Pokropek et al. 2019) have shown that
  1. Partial invariance is not worse than approximate invariance or alignment
  2. But alignment is easier when the number of groups is very large.

## 4) Eldad Davidov - application:
## measurement invariance of values in the European Social Survey

## Schwartz's theory of basic human values

**Basic values -**
***Beliefs about the importance of abstract goals as guiding principles in life***

**1) Structure: a circumplex continuum**


**2) Content: 10 value types**
**measured by 21 items in ESS**

**3) Very commonly used for cross-cultural research**

# Application:
# measurement invariance of values in European Social Survey

## Schwartz's measurement of basic human values in the ESS

| *How much like you is this person?* | **Not like me at all** | **Not like me** | **A little like me** | **Some-what like me** | **Like me** | **Very much like me** |
|---|---|---|---|---|---|---|
| BE *It's very important to him to help the people around him. He wants to care for their well-being.* | 1 | 2 | 3 | 4 | 5 | 6 |
| UN *He strongly believes that people should care for nature. Looking after the environment is important to him.* | 1 | 2 | 3 | 4 | 5 | 6 |
| TR Tradition is important to him. He tries to follow the customs handed down by his religion or his family. | 1 | 2 | 3 | 4 | 5 | 6 |
| AC *It's important to him to show his abilities. He wants people to admire what he does.* | 1 | 2 | 3 | 4 | 5 | 6 |

## Previous findings of measurement invariance tests of values in the ESS

1) Davidov, Schmidt, & Schwartz (2008): The disappointing result

only metric invariance was established for the values
(lack of scalar invariance)
…the comparison of latent means across countries is not allowed

2) Approximate measurement invariance and alignment worked much better. It is possible to compare values in most (but not all) countries.

Davidov, E., Schmidt, P., & Schwartz, S. H. (2008). Bringing values back in. The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly*, *72*, 420-445.

Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P. (2018). Testing for approximate measurement invariance of human values in the European Social Survey. *Sociological Methods & Research*.

# Conclusions

**Take home messages:**

**1) Various techniques** to examine leasurement invariance**: full or partial exact invariance** in **MGCFA , approximate invariance**

2) It is **difficult to establish full scalar invariance in international surveys**, **easier across groups within cultures** (gender groups, age groups, education groups etc.)

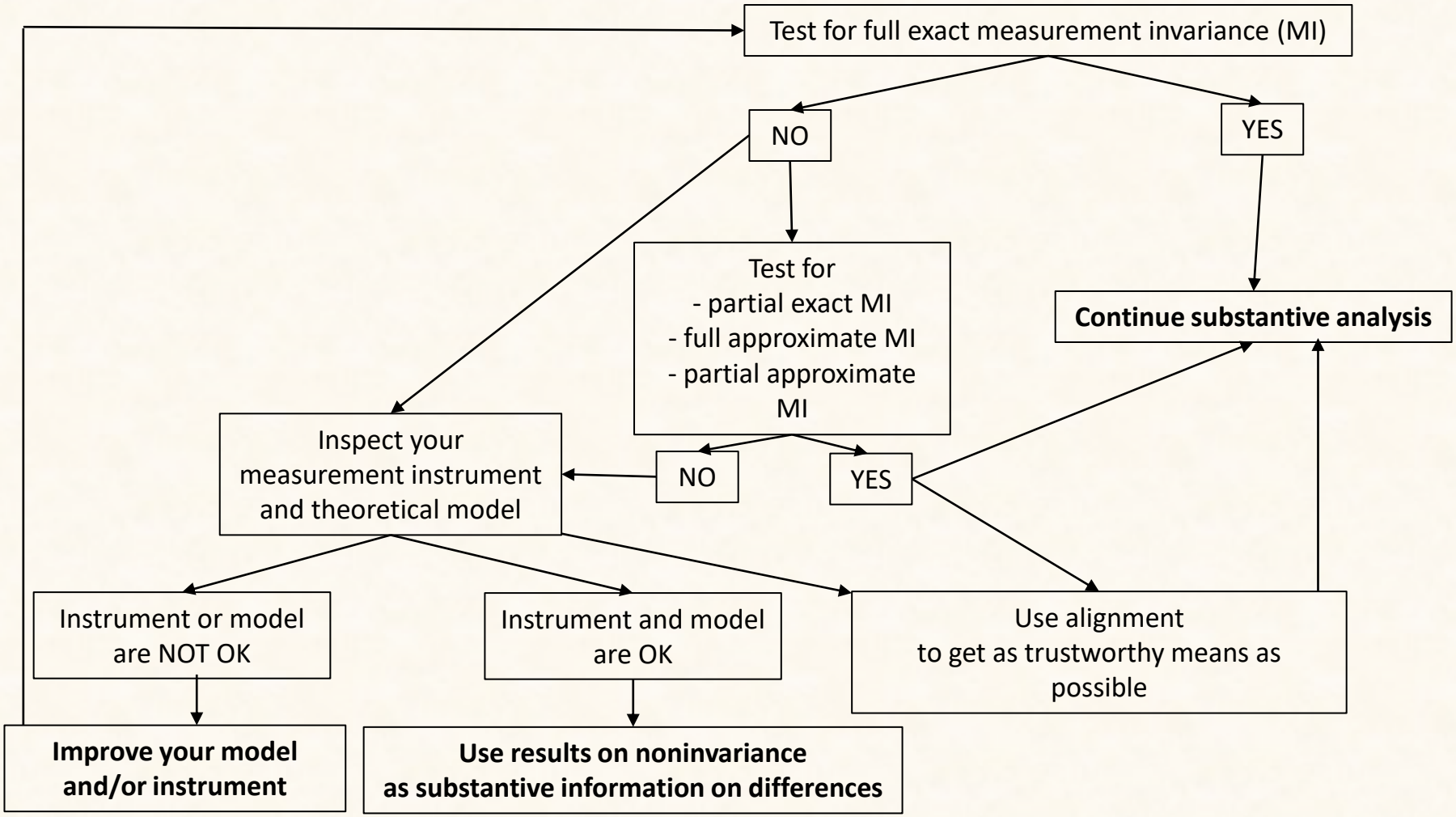
3) Simulations show that **partial scalar invariance** is **not worse than approximate invariance**

4) **As in the example of the human values** measurements: If full exact invariance is not established, one may try alternative less strict tests like partial invariance, approximate invariance or alignment.

***Thank you***
***for your attention!***

| | **Frequentist exact measurement invariance** | **Bayesian approximate measurement invariance** |
|---|---|---|
| Contraints on parameters (loadings and intercepts) | • Parameters are constrained to be ***exactly*** equal | • Parameters are constrained to be ***approximately*** equal (using zero-mean, small-variance priors) |
| Model fit | • ***Change*** in CFI, RMSEA, SRMR<br>• (Change in $\chi^2$)<br>• Detection of model ***misspecification*** | • Posterior Predictive P-Value is ***not significant*** (observed data compared to the posterior predictive distribution)<br>• 95% Credibility Interval contains zero<br>• DIC |

# Measurement Invariance – decision tree




Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R. (2019). How to obtain comparable measures for cross-national comparisons. *Kolner Zeitschrift fur Soziologie und Sozialpsychologie*, *71*, 157-186.

Artur Pokropek, Eldad Davidov & Peter Schmidt (2019): A Monte Carlo Simulation Study to Assess The Appropriateness of Traditional and Newer Approaches to Test for Measurement Invariance, Structural Equation Modeling

- Design and Major Message
- results of simulations for 804 conditions (Sample Size, factor loadings, Amount of bias etc.)
- gave rise to more than 300,000 individual estimations.
- guidance for applied researchers facing partial
- and approximate MI.
-
- Even large deviations from strict MI may allow precise
- estimations and meaningful comparisons of both means
- and path coefficients.
-
- Using appropriate models, both
- group rankings as well regression coefficients of structural
- equation models could be correctly recovered under specific
- conditions.
-

- PARTIAL MEASUREMENT INVARIANCE(PMI).
-
-
- Although one could choose a different bias size, we chose a realistic noninvariance bias.
-
- Indeed, robustness checks suggested that lowering the sample size to
- 1,000 and/or increasing the partial noninvariance bias
- from 0.2 to 0.3 essentially did not alter the results.
-
- For generating data, we used highly reliable scales with factor
- loadings between 0.65 and 0.85—in real survey data, factor
- loadings may range between 0.35 and 1.0.
-

- BAYESIAN SEM(BSEM)
- Convergence rates for BSEM models with
- relatively high priors (0.01 and 0.05) were low.
-
- Robustness checks analyzing different convergence
- criteria  provided virtually the same conclusions
- as those presented in the paper.

- more strict convergence criteria produced better convergence rates
- but computational time grew exponentially, stretching the time of model estimation into weeks.

- Partial Scalar Invariance and alignment
- PMI models may be rather effective to recover
- both path coefficients and latent means when many or
- even most items are noninvariant.

- The alignment procedure is recommended for recovering latent means:
- cases where there are only few noninvariant parameters and many groups( $g > 5$).

## Applying the approximate measurement invariance of values in the ESS

### Step 1

Testing for full approximate measurement invariance
(mean = 0, variance = .05) in the 1st round of the ESS
for each higher order value separately.

### Step 2

Identifying items with deviating parameters
(loadings and intercepts significantly different from those in other groups).

### Step 3

Testing for partial approximate measurement invariance
(releasing deviating parameters identified in step 2).

### Step 4

Replication in rounds 2-6

Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research, online first.

# ESS Data

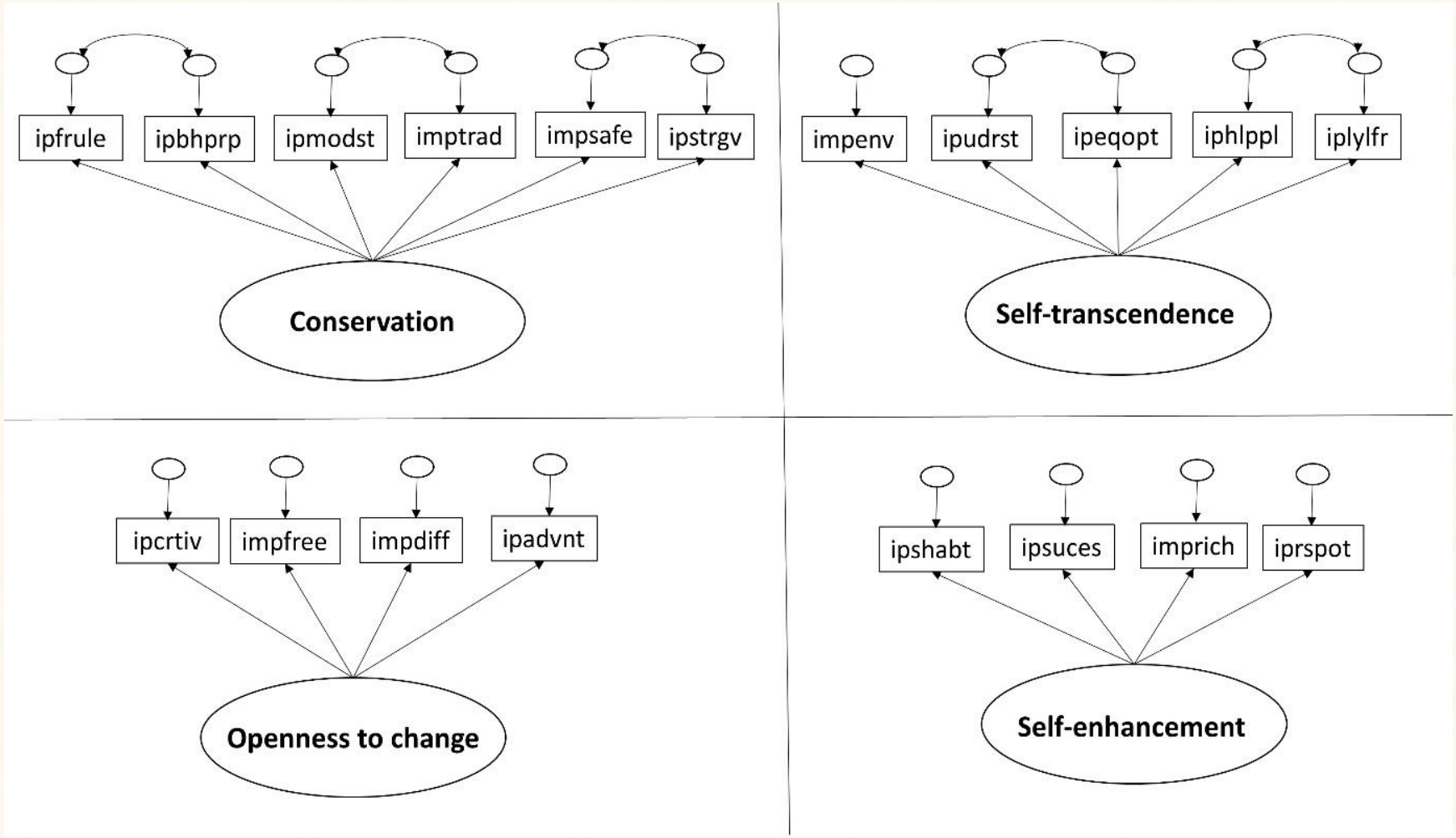
**N = 274,447**

**15 countries that took part in all**

**6 rounds**

| | 1st round | 2nd round | 3rd round | 4th round | 5th round | 6th round |
|---|---|---|---|---|---|---|
| Belgium | 1819 | 1734 | 1767 | 1704 | 1674 | 1809 |
| Denmark | 1457 | 1457 | 1451 | 1554 | 1548 | 1610 |
| Finland | 1758 | 1692 | 1645 | 1898 | 1638 | 2142 |
| Germany | 2785 | 2800 | 2828 | 2697 | 2943 | 2910 |
| Hungary | 1564 | 1407 | 1409 | 1388 | 1404 | 1919 |
| Ireland | 1838 | 1139 | 1582 | 1682 | 2295 | 2498 |
| Netherlands | 2301 | 1824 | 1814 | 1693 | 1754 | 1788 |
| Norway | 1806 | 1543 | 1533 | 1374 | 1518 | 1598 |
| Poland | 1982 | 1621 | 1629 | 1544 | 1675 | 1818 |
| Portugal | 1417 | 1987 | 2117 | 2220 | 2035 | 2062 |
| Slovenia | 1390 | 1297 | 1329 | 1172 | 1238 | 1159 |
| Spain | 1638 | 1544 | 1802 | 2520 | 1862 | 1820 |
| Sweden | 1677 | 1663 | 1585 | 1539 | 1457 | 1799 |
| Switzerland | 2009 | 2084 | 1758 | 1764 | 1467 | 1453 |
| United Kingdom | 1748 | 1806 | 2301 | 2230 | 2315 | 2212 |

# Tested models




Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research, online first.

# Results

| | PPP | 95% Credibility Interval |
|---|---|---|
| **Self-enhancement across ALL countries:** | | |
| 1st Round of ESS | .178 | -29.290 – 82.959 |
| 2nd Round of ESS | .089 | -17.906 – 94.001 |
| 3rd Round of ESS | .035 | -4.388 – 108.102 |
| 4rd Round of ESS | .096 | -19.130 – 92.985 |
| 5th Round of ESS | .014 | *6.500 – 118.831* |
| 6th Round of ESS | .001 | *33.790 – 145.481* |

Conclusion:

**Scalar measurement invariance established in all countries!**

Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research, online first.

# Results

| | PPP | 95% Credibility Interval |
|---|---|---|
| **Openness (without hedonism) across ALL countries:** | | |
| 1st Round of ESS | .001 | *37.312 – 149.885* |
| 2nd Round of ESS | .035 | -4.253 – 107.713 |
| 3rd Round of ESS | .059 | -11.201 – 100.421 |
| 4rd Round of ESS | .014 | *7.054 – 119.765* |
| 5th Round of ESS | .051 | -9.382 – 103.396 |
| 6th Round of ESS | .102 | -19.736 – 92.334 |

Conclusion:

**Scalar measurement invariance established in all countries!**

Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research, online first.

# Results

| | PPP | 95% Credibility Interval |
|---|---|---|
| **Self-transcendence across ALL countries:** | | |
| 1st Round of ESS | .000 | 72.03 – 204.32 |
| 2nd Round of ESS | .000 | 100.93 – 231.49 |
| 3rd Round of ESS | .000 | 192.84 – 323.78 |
| 4rd Round of ESS | .000 | 141.69 – 272.43 |
| 5th Round of ESS | .000 | 206.22 – 337.59 |
| 6th Round of ESS | .000 | 218.25 – 349.08 |

Conclusion:

**Scalar measurement invariance NOT established across all countries!**

Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research, online first.

# Results

| | PPP | 95% Credibility Interval |
|---|---|---|
| **Self-transcendence in 12 countries:** | | |
| Belgium, Finland, Germany, Ireland, Netherlands, Norway, Poland, Portugal, Spain, Sweden, Switzerland, United Kingdom | | |
| 1st Round of ESS | .508 | -50.05 – 49.46 |
| 2nd Round of ESS | .419 | -45.17 – 55.39 |
| 3rd Round of ESS | .326 | -38.57 – 61.99 |
| 4rd Round of ESS | .419 | -45.13 – 55.07 |
| 5th Round of ESS | .273 | -34.72 – 65.34 |
| 6th Round of ESS | .505 | -48.32 – 47.27 |

Conclusion:

**Scalar measurement invariance across 12 countries established!**

Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research, online first.

# Results

| | PPP | 95% Credibility Interval |
|---|---|---|
| **Conservation across ALL countries:** | | |
| 1st Round of ESS | .000 | 268.70 – 419.12 |
| 2nd Round of ESS | .000 | 188.11 – 339.10 |
| 3rd Round of ESS | .000 | 159.55 – 310.56 |
| 4rd Round of ESS | .000 | 232.02 – 383.11 |
| 5th Round of ESS | .000 | 263.83 – 413.27 |
| 6th Round of ESS | .000 | 205.24 – 356.04 |

Conclusion:
**Scalar measurement invariance NOT established across all countries!**

Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research, online first.

# Results

| | PPP | 95% Credibility Interval |
|---|---|---|
| **Conservation in 10 countries**: | | |
| Belgium, Finland, Germany, Ireland, Netherlands, Poland, Portugal, Switzerland, United Kingdom, Slovenia | | |
| 1st Round of ESS | .173 | -23.70 – 67.94 |
| 2nd Round of ESS | .131 | -19.20 – 72.09 |
| 3rd Round of ESS | .135 | -20.07 – 71.09 |
| 4rd Round of ESS | .097 | -15.76 – 75.66 |
| 5th Round of ESS | .176 | -23.61 – 66.82 |
| 6th Round of ESS | .067 | -10.68 – 80.66 |

Conclusion:

**Scalar measurement invariance across 10 countries established!**

Cieciuch, J., Davidov, E., Algesheimer, R., Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European Social Survey. Sociological Methods & Research, online first.