

gesis

Leibniz Institute
for the Social Sciences



Measurement Quality of Agree/Disagree and Item- Specific Response Options

*Natalja Menold, Matthias Bluemke & Anna
Andreenkova*

Agree/Disagree vs. Item-Specific

- Agree/Disagree (AD): „To what extent do you agree or disagree that your health is excellent”
 - ▶ AD is used for a set of items / any content of questions
- Item-Specific (IS): “How would you rate your health – excellent, very good, good, fair, or bad?”
 - ▶ Categories are related to the evaluation domain of a question

e.g. Saris et al., 2010

Background

- Cognitive process is easier in the case of the IS form
- A/D: Possible shortcomings
 - ▶ Agreeing and disagreeing are not necessary linearly related to each other
 - > decrease in validity
 - ▶ Respondents can use different translation into A/D-dimension for different items in a battery
 - > decrease in reliability
 - ▶ AD can be prone to the acquiescence response bias

Saris et al., 2010

Empirical Support

- Higher reliability and validity of IS than of AD form
 - ▶ Counte, 1979 (classical approach; criterion validity)
 - ▶ Scherpenzeel & Saris, 1997; Saris et al., 2010 (MTMM approach)
- However, there are also contradictory results
 - ▶ e.g. Berkowitz & Wolken, 1964 (classical approach)
- No systematic differences in acquiescence
 - ▶ Schuman & Presser, 1981
- No comparisons with respect to cross-cultural measurement invariance

Research Questions (1/2)

- Are there differences between AD and IS forms with respect to cross-cultural measurement invariance?
- *„As proposed by Mellenbergh (1989), “measurement invariance” (MI) requires that the association between the items (or test scores) and the latent factors (or latent traits) of individuals should not depend on group membership or measurement occasion (i.e., time).”*

(van de Schoot et al., 2015, editorial)

Measurement Invariance

- **Metric invariance: comparability of factor loadings**
 - ▶ a variable has the same relationship with the latent variable across countries
 - ▶ prerequisite for the comparability of correlations
- **Scalar invariance: comparability of item intercepts**
 - ▶ a variable value is related to the same latent value
 - ▶ prerequisite of comparability of means

Research Questions (2/2)

- Are different methods to measure reliability and validity responsible for the mixed results?

| | Classical Approach | MTMM |
|-------------|---|--|
| Reliability | <ul style="list-style-type: none"> - correlation between repeated measures (Test-Retest), - item variance to the variance of the multiple component score (e.g. Cronbachs' Alpha) - common item variance to the error variance (Composite Reliability) | <ul style="list-style-type: none"> - Common item variance |
| Validity | <ul style="list-style-type: none"> - prediction of the values of a third variable (Criterion); - correlations with measurements of closely/not closely related measures (Discriminant, Convergent) | <ul style="list-style-type: none"> - relationship between the repeated measures (obtained with different methods) |

Data

- Saris et al. (2010):
 - ▶ MTMM studies in the ESS
 - ▶ A/D provides lower measurement quality (reliability*validity) than IS
- Reuse of the ESS data

ESS 3

3 items: opinions about immigration

| | | | |
|----|---|--------------------------------|------------------------------|
| MQ | IS, 4 categories (cat.), fully labelled | | |
| SQ | AD 5 cat. fully labelled | IS 4 cat. fully labelled | AD 7 cat. end labelled |

ESS 2

3 items: opinions about work

| | | |
|----|--|-------------------------------|
| MQ | AD (not true/true), 4 cat., fully labelled | |
| SQ | IS 4 cat. fully labelled | IS 11 cat. end labelled |

MQ: main questionnaire; SQ: supplementary questionnaire

Results ESS2: Measurement Invariance

- Problem: item 3 (health issues) is hardly related with the other two items (e.g., Saris & Gallhofer, 2014)
- Solution: Analysis of two items only
 - ▶ third item from main questionnaire (Job requires learning new things) is used as an auxiliary variable.
 - ▶ Differences/quality of this variable are/is not tested in the analyses

Results ESS 2: Measurement Invariance

AD

| | Goodness-of-fit | | | | Model difference tests | | |
|--------------|------------------|-------|-------|-----------|---------------------------|--------------|----------------|
| | χ^2 (df) | CFI | RMSEA | BIC | $\Delta\chi^2(\Delta df)$ | ΔCFI | $\Delta RMSEA$ |
| 1 configural | 0.000 (0) | 1.000 | .000 | 132224.20 | - | - | - |
| 2 metric | 75.731** (42) | 0.993 | .032 | 132032.93 | 75.73* (46.72) | .007 | .032 |
| 3 scalar | 1089*** (63) | 0.798 | .145 | 133010.48 | 1034*** (22.63) | .195 | .110 |

IS4

| | Goodness-of-fit | | | | Model difference tests | | |
|--------------|--------------------|-------|-------|----------|---------------------------|--------------|----------------|
| | χ^2 (df) | CFI | RMSEA | BIC | $\Delta\chi^2(\Delta df)$ | ΔCFI | $\Delta RMSEA$ |
| 1 configural | 0.000 (0) | 1.000 | .000 | 80305.06 | - | - | - |
| 2 metric | 74.86** (42) | 0.977 | .031 | 80109.29 | 74.88* (46.76) | .023 | .031 |
| 3 scalar | 375.920*** (63) | 0.785 | .077 | 80297.13 | 311.43*** (22.08) | .192 | .046 |

IS11

| | Goodness-of-fit | | | | Model difference tests | | |
|--------------|--------------------|-------|-------|----------|---------------------------|--------------|----------------|
| | χ^2 (df) | CFI | RMSEA | BIC | $\Delta\chi^2(\Delta df)$ | ΔCFI | $\Delta RMSEA$ |
| 1 configural | 4.54* (1) | 0.997 | .064 | 79803.71 | - | - | - |
| 2 metric | 112.91** (46) | 0.943 | .041 | 79646.18 | 108.89*** (52.94) | 0.054 | .021 |
| 3 scalar | 345.530*** (60) | 0.756 | .075 | 79821.20 | 247.33*** (14.94) | 0.187 | .034 |

Reliability Assessment

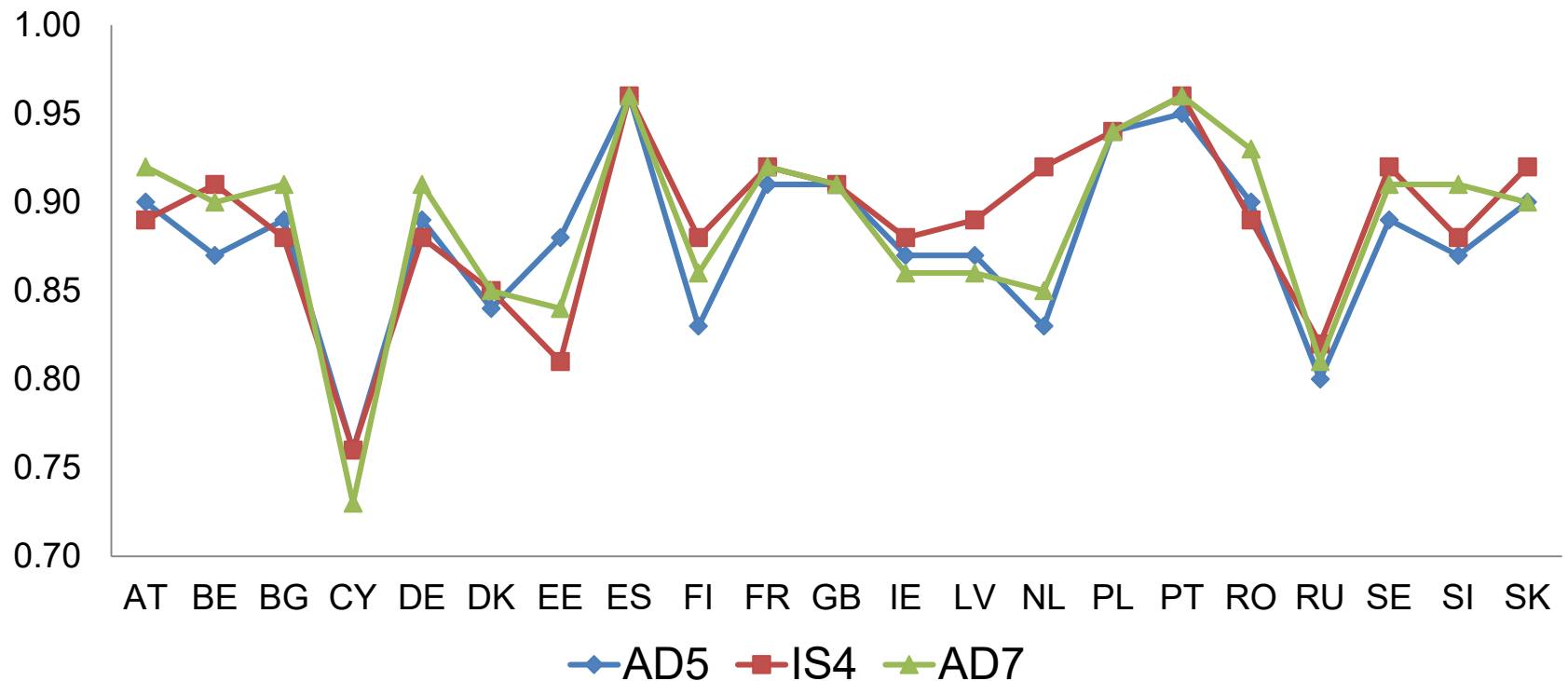
- Composite Reliability
- Differences among groups with MGCFA (e.g. Menold & Raykov, 2015).

Reliability ($\hat{\rho}_x$)

$$\hat{\rho}_x = \frac{(\hat{b}_1 + \dots + \hat{b}_p)^2}{(\hat{b}_1 + \dots + \hat{b}_p)^2 + \hat{\theta}_1 + \dots + \hat{\theta}_p}$$

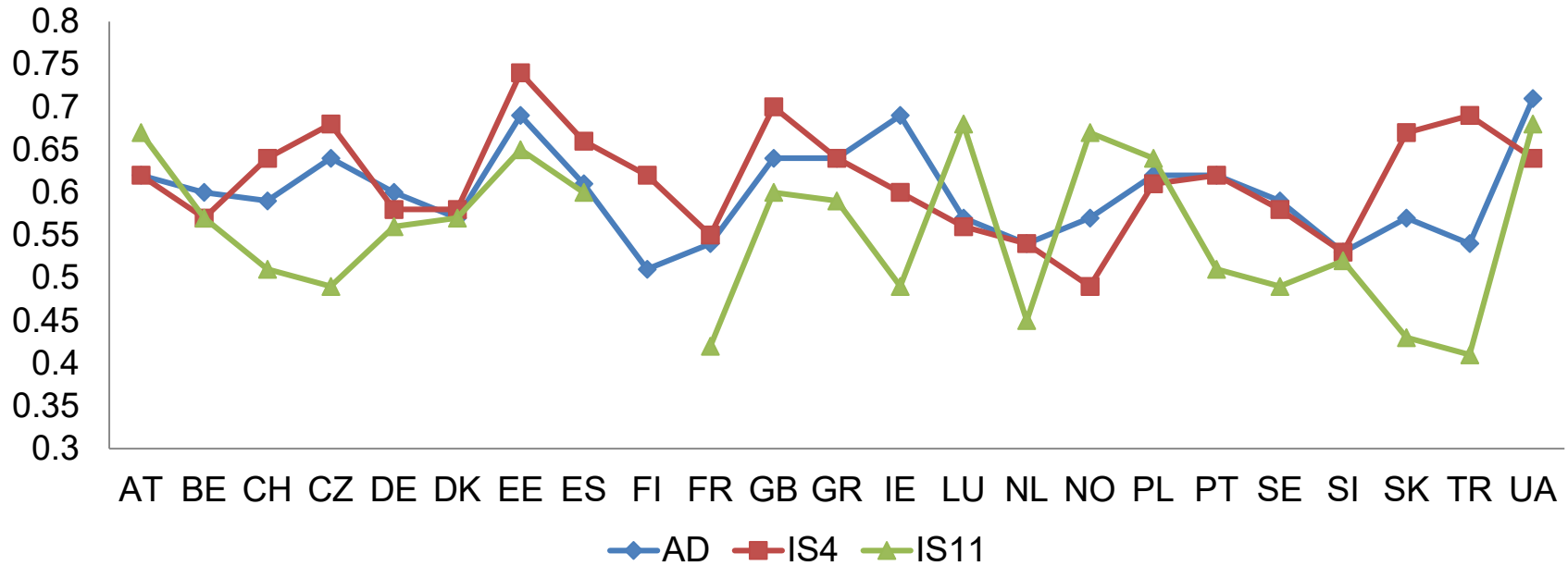
where b_1, \dots, b_p are the factor loadings and $\theta_1, \dots, \theta_p$ are the error variances, obtained from the MGCFA

Reliability ESS3



Significant differences:
AD > IS: AT, DE, EE, RO
IS > AD: BE, FI, NL, SK

Reliability ESS2



Significant/high Differences

AD > IS: IE, NO

IS > AD: CH, FI, NO, SK, TR

4 categories > 11 categories: CH, CZ, EE, FR, GB, IE, NL, PT, SE, SK, TR

Criterion Validity

- ESS3: Regression on „Immigration bad or good for country's economy”
- ESS2: Regression on “How satisfied with life as a whole”
- Relation is not observed for at least one independent variable:
 - ▶ ESS3: AD5: 15 countries; IS: 15 countries; AD7: 12 countries
 - ▶ ESS2: AD: 11 countries; IS4: 20 countries; IS11: 16 countries

higher Validity for AD in the ESS2?

Summary

- Key findings based on comparative cross-national data:
 - ▶ Measurement invariance: ESS 2: better results for A/D form
 - ▶ Reliability: No consistent differences between AD and IS forms
 - ▶ Validity: in ESS2 higher for AD than for IS form
 - it would be easier to provide agreement/disagreement than to evaluate amounts
 - alternative explanation by repeated administration

Contact :

natalja.menold@gesis.org

matthias.bluemke@gesis.org

anna.andreencova@cessi.ru