
Presentation for CESSDA meeting Paris 2008

Requirements for Cross-cultural
comparison

by

Willem E.Saris

Comparative cross-cultural research

We discuss cross-cultural comparisons of

- means
- relationships

On the basis of

- single questions,
- composite scores and
- latent variables

Basic Requirement: Equivalence

- Comparisons can only be made if the people in the different countries have the same idea about the questions asked
- The measures should be “functionally equivalent”
or
- The response models should be invariant across countries

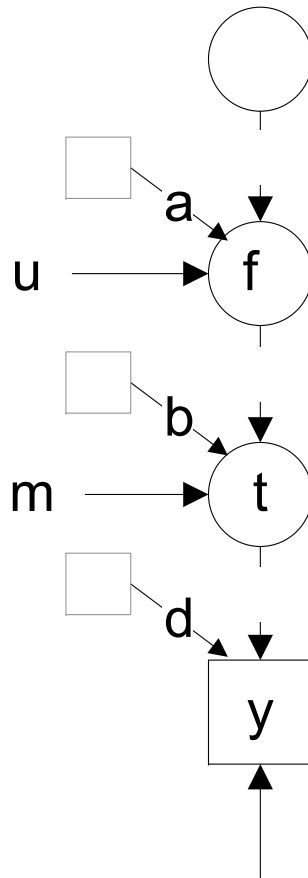
Example: Subjective Competence

Subjective Competence measures “to what extent a person believes that he/she can play a role in politics”

This is a construct with several indicators:

1. Whether one understands politics (Understand)
2. Whether one thinks that one can play an active role in politics (Active)
3. Whether one has the feeling that politics is to complex (Complex)

A simple response model



Unique component ()

1



Formal model

The cognitive process

$$f_1 = a + cCP_1 + u$$

Relation between the indicator and its true score

$$t = b + vf_1 + m$$

Relation between the true score and the observed score

$$y = t + e$$

It follows by substitution that

$$y = \tau + \lambda CP_1 + \zeta$$

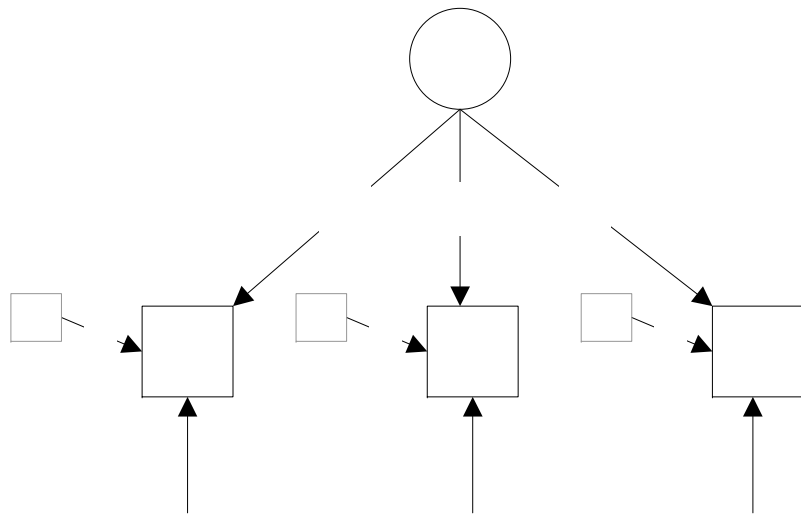
where

$$\lambda = vc$$

$$\tau = b + va$$

$$\zeta = e + m + vu$$

The standard approach to test equivalence across countries



- 1. *Configural invariance*, meaning that the model holds for all the countries involved
- 2. *Metric invariance*, meaning that, besides configural invariance, the slopes are the same in all the countries studied.
- 3. *Scalar invariance*, meaning that, besides metric invariance, the intercepts are the same across all countries being compared.
- Metric invariance is sufficient for comparison of relationships.
- The comparison of means requires scalar invariance

CP₁

$$\lambda_{11} \quad \lambda_{21} \quad \lambda_{31}$$

Test of these requirements

Results of the tests of different requirements of invariance for the concept of "subjective competence" based on data from three countries

Invariance restrictions	chi ²	df	Probability
Scalar	84.5	8	.00
Metric	13.3	4	.01
Metric except for λ_{21} in Spain	8.0	3	.39

- Metric and Scalar invariance have been rejected
- What now ?

Criticism

- The test is done to determine whether people have the same ideas about the different questions as indicators for the concept by postulation
- The parameters of the standard model are products of parameters from the cognitive process and the measurement process.
- One should test only the parameters of the cognitive process.

Alternative model

The original model

The cognitive process

$$f_1 = a + cCP_1 + u$$

Relation between the indicator and its true score

$$t = b + vf_1 + m$$

Relation between the true score and the observed score

$$y = t + e$$

The Alternative model

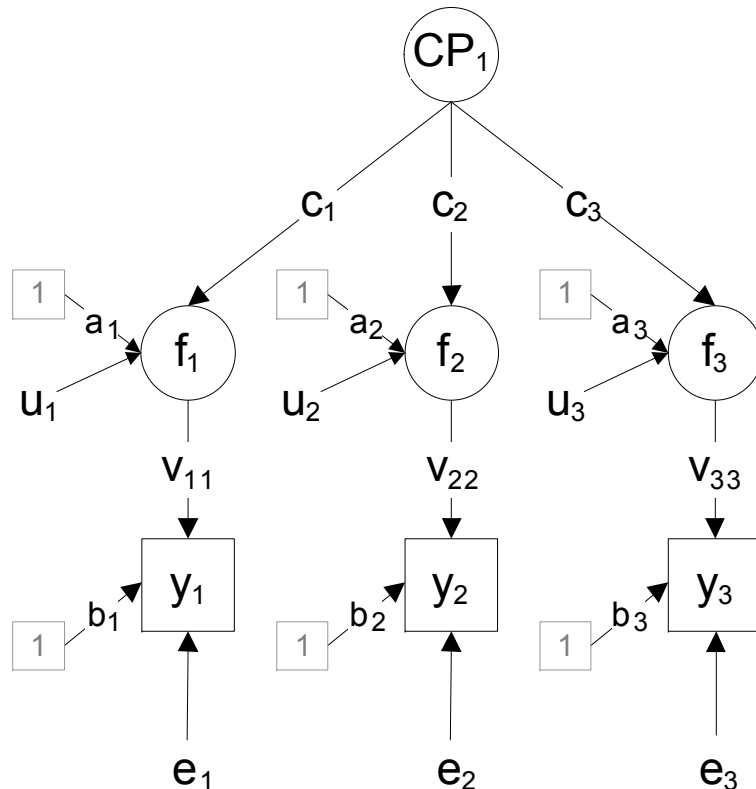
One equation for the cognitive process

$$f_1 = a + cCP_1 + u$$

One for the measurement process

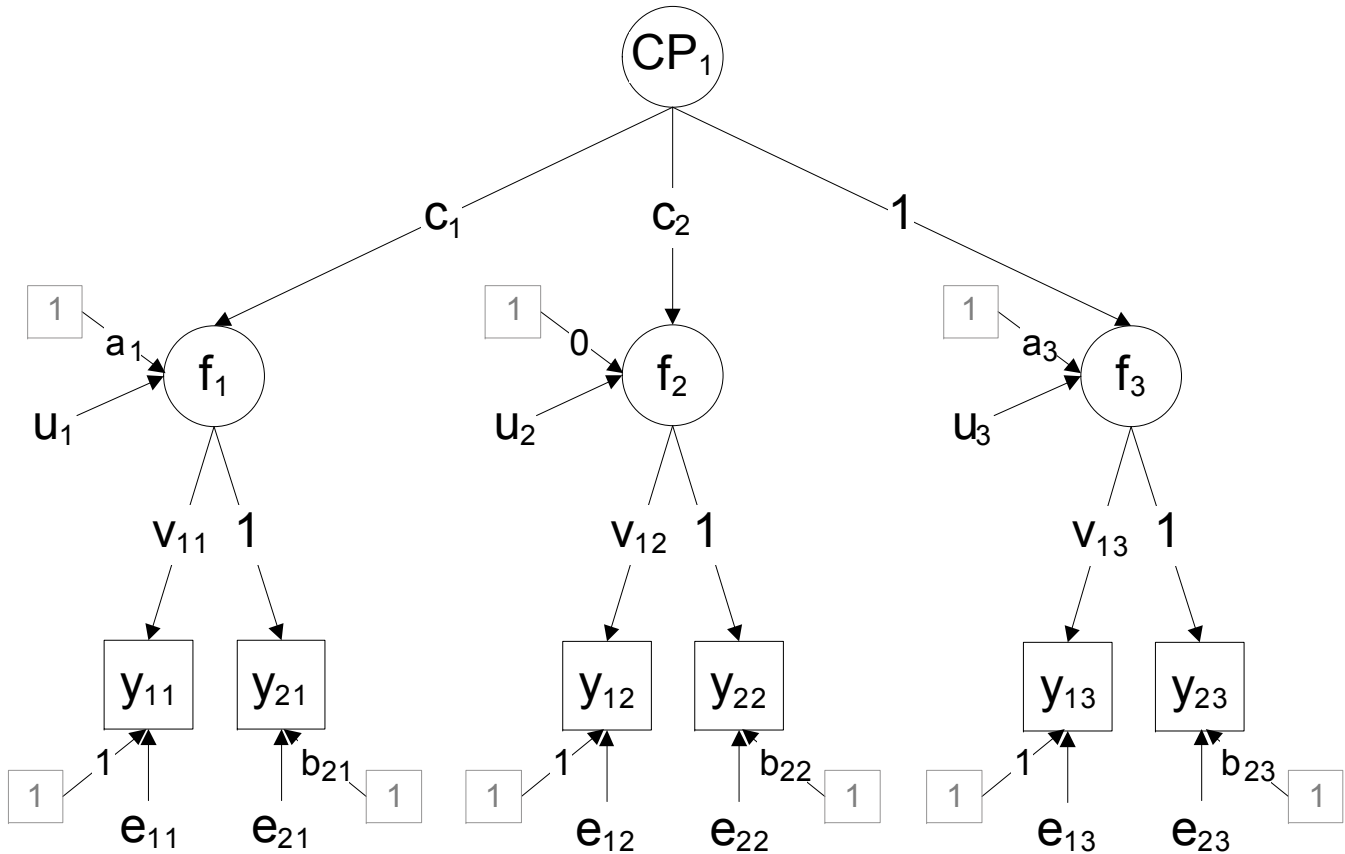
$$y = b + vf_1 + m + e$$

Alternative tests of equivalence



- Cognitive equivalence of measurement instruments should be required.
- For metric invariance we require that c_i is equal across countries and not $C_i V_{ii}$
- For scalar invariance also a_i should be equal across countries

A model to estimate these coefficients



Results of the test

Estimates of parameters of the model

Parameter	UK	NL	Spain
Slopes of the measurement equations			
v_{11}	1.15	1.06	0.98
v_{12}	1.15	1.02	1.00
v_{13}	1.02	0.93	1.13
Intercepts of the measurement equations			
b_{21}	1.28	1.00	0.81
b_{12}	1.00	1.00	1.28
b_{22}	1.22	1.09	1.46
b_{23}	1.06	0.87	1.37
Slopes of the cognitive process equations			
c_1	-.94	-.94	-.94
c_2	.91	.91	.60
c_3	1.00	1.00	1.00
Intercepts of the cognitive process equations			
a_1	3.09	3.09	3.09
a_2	1.00	1.00	1.00
a_3	-.84	-.84	-.84

- Parameters of the measurement equations are rather different
- The parameters of the cognitive process are rather similar except one.

Conclusions

- The standard invariance tests can lead to a rejection of the model, while the cause of the problem is not that the indicators have a different interpretation across countries,
- but that instead the respondents in the different countries use the scales in different ways.
- Therefore, we prefer the less restrictive requirements of *cognitive equivalence* over the standard requirements.

Comparison of means and relationships *for single questions*

- There is a strong research tradition that concentrates on differences in means and relationships between responses to single questions across countries.
- This, however, is a very questionable activity, if the questions are not previously checked for equivalence.

Possible differences in means

- The measure “Complex “ is scalar invariant. Therefore one should expect the same observed means if the latent means are equal across countries:
- Given the latent mean = 3 in all three countries
- The expected mean in UK is 1.50
- The expected mean in NL is 1.33
- The expected mean in Spain is 1.00

Normally this difference will be seen as substantive but it is due to differences in the measurement

Underestimation of the differences in relationships

- The correlation between the variables “complex” and “understand” is $-.44$ in Greece and $-.514$ in the Check Republic.
- After correction for measurement error, the difference is much larger.
- In Greece the correlation becomes $-.59$,
- In the Check Republic it is $-.77$.

- In this case the differences would have been underestimated.

Overestimation of the differences in relationships

- However, the opposite also occurs.
- In the Czech Republic and Slovenia, we get for the observed correlations, respectively, $-.514$ and $-.449$.
- After correction for measurement errors, these correlations are exactly equal and have the value $-.77$.
- In this case the differences would be overestimated

Conclusion

- The means and relationships of single requests in cross-cultural research **cannot be compared** unless the measurement instruments are equivalent.
- Correction for measurement error is essential
- Given the usual lack of information, we advise to proceed with caution when attempting to compare results based on single requests.

Comparison of means of *composite scores*

- After correction for differences in the measurement equations the composite score for a CP_1 can be computed as a unweighted sum (C_1) of the three latent variables $f_1 - f_3$:

$$C_1 = (f_1 + f_2 + f_3)$$

or

$$C_1 = (a_1 + c_1 CP_1 + u_1) + (a_2 + c_2 CP_1 + u_2) + (a_3 + c_3 CP_1 + u_3)$$

and

$$\mu_{C_1} = (a_1 + a_2 + a_3) + (c_1 + c_2 + c_3) \mu_{CP_1}$$

- This shows that the means can only be compared if the **intercepts and slopes are invariant** across countries (cognitive scalar invariance)

Comparison of relationships of *composite scores*

- The covariance between the variables of interest CP1 and CP2 is denoted by σ_{CP1CP2}

- When

$$C_1 = f_{11} + f_{21} + f_{31} \quad \text{and} \quad C_2 = f_{12} + f_{22} + f_{32}$$

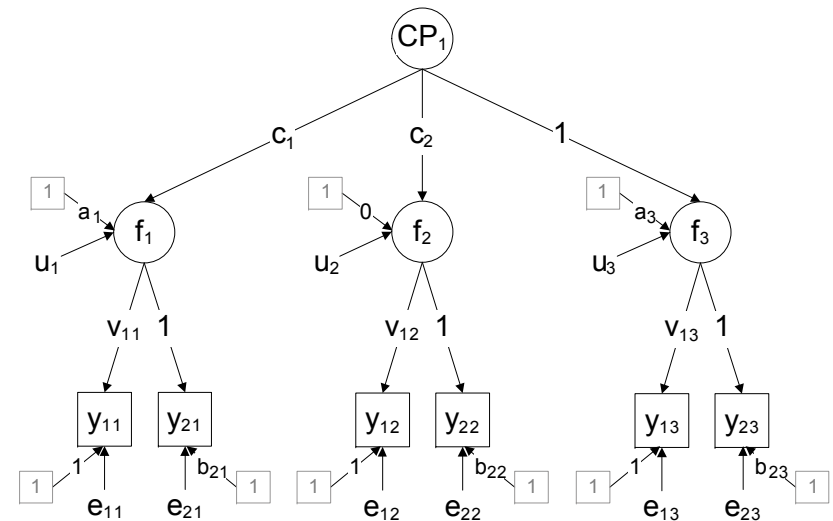
- We can derive:

$$\sigma_{C_1, C_2} = (c_{11} + c_{21} + c_{31}) (c_{12} + c_{22} + c_{32}) \sigma_{CP1CP2}$$

- This result shows that the covariances of composite scores across the countries cannot be compared if not **all slope coefficients are invariant** (Cognitive metric invariance)

Comparison of means and relationships of *latent variables*

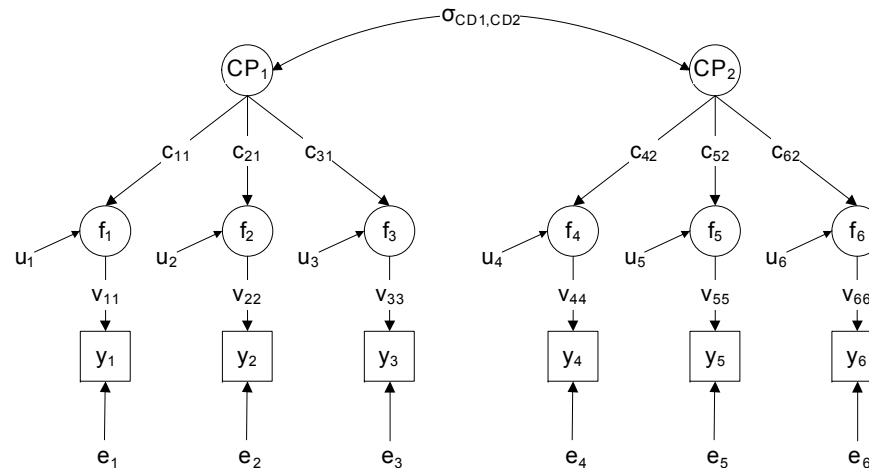
- It is much easier and safer to estimate means of the latent variables for the comparison.
- Even though the second indicator was not scalar invariant, the estimates of the means are correct because two indicators are sufficient to identify the means.



Extra information

- An advantages of this approach is that one can directly test that the means in the different countries are identical.
- Specifying this restriction we get a $\chi^2=53.8$ with $df=27$ where $Pr=.002$.
- Without the equality constraint $\chi^2= 20.1$ with $df=25$.
- So we can conclude that the means are significantly different from each other.

Compare the relationships between *latent variables*



- If the model is correctly specified, it is also possible to simultaneously estimate the relationships between the latent variables with the quality of the measurement instruments in the different countries with a minor extension of the input for the program, we have used before.

Conclusion 1: Correction for errors is necessary

- Without correction for measurement errors, we run the risk of giving explanations for differences between countries on substantive grounds that could be due to differences in measurement quality of the instruments.

Conclusion 2: Standard test too strict

- Comparison across countries requires metric and scalar invariance but the commonly used requirements for equivalence are too strict.
- We suggest to test on cognitive equivalence i.e. invariance after correction for measurement errors

Conclusion 3: Necessary Information

- A problem with this approach is that the needed information is seldom available.
- The information about the quality can be derived from external sources such as MTMM experiments or SQP predictions.

Conclusion 4: Use of single questions is very questionable

Comparisons based on *single questions* without knowledge of the equivalence of the questions is very doubtful

Conclusion 5: Avoid use of composite scores

- Using *composite scores* for concepts-by-postulation, the comparison across countries requires perfect metric invariance for comparison of relationships and perfect scalar invariance for comparison of means.
- These requirements are very strict and will rarely be satisfied.

Conclusion 6 Use Latent variables

- Comparing means and relationships between *latent variables* across countries does not have to require perfect invariance.
- Consistent estimates of the means and relationships are also possible with partial equivalence.
- Therefore, using latent variables is a more flexible approach than employing composite scores.
- The only disadvantage is that the models become more complex.

Conclusion 7 : It is not so simple

- Overall, the conclusion is that cross-cultural comparisons are not as simple as they seem and
- that the comparison of means is even more difficult than the comparison of relationships between variables.