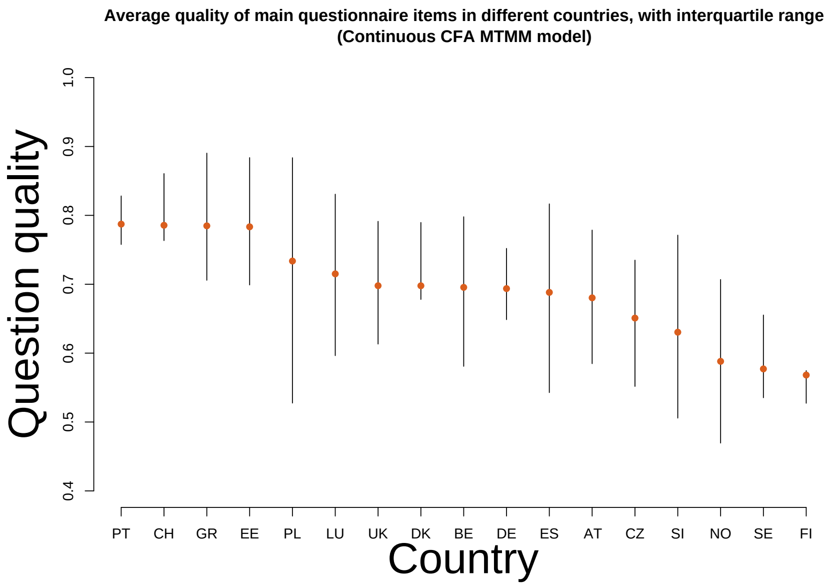
Categorization errors and differences in the quality of questions across countries

Daniel Oberski
Willem Saris
Jacques Hagenaars

Faculty of Social and Behavioural Sciences
Tilburg University

Survey Research Centre
ESADE Barcelona, Universitat Ramon Llull


Average quality of main questionnaire items in different countries, with interquartile range
(Continuous CFA MTMM model)
Question quality
0.4
0.5
0.6
0.7
0.8
0.9
1.0
PT
CH
GR
EE
PL
LU
UK
DK
BE
DE
ES
AT
CZ
SI
NO
SE
FI
Country

- Big differences in question quality are found across countries
- Previous explanations for the differences we (Oberski ea 2007) sought for were
  - the complexity of the sentences in different languages
  - translation errors
  - different implementations of the experiments used to estimate the quality
- None of these possible reasons could sufficiently explain the large differences found
- Therefore we now consider another possible explanation:
- Categorization error.

- Big differences in question quality are found across countries
- Previous explanations for the differences we (Oberski ea 2007) sought for were
  - the complexity of the sentences in different languages
  - translation errors
  - different implementations of the experiments used to estimate the quality
- None of these possible reasons could sufficiently explain the large differences found
- Therefore we now consider another possible explanation:
- Categorization error.

- Big differences in question quality are found across countries
- Previous explanations for the differences we (Oberski ea 2007) sought for were
  - the complexity of the sentences in different languages
  - translation errors
  - different implementations of the experiments used to estimate the quality
- None of these possible reasons could sufficiently explain the large differences found
- Therefore we now consider another possible explanation:
- Categorization error.

- Big differences in question quality are found across countries
- Previous explanations for the differences we (Oberski ea 2007) sought for were
  - the complexity of the sentences in different languages
  - translation errors
  - different implementations of the experiments used to estimate the quality
- None of these possible reasons could sufficiently explain the large differences found
- Therefore we now consider another possible explanation:
- Categorization error.

- Big differences in question quality are found across countries
- Previous explanations for the differences we (Oberski ea 2007) sought for were
  - the complexity of the sentences in different languages
  - translation errors
  - different implementations of the experiments used to estimate the quality
- None of these possible reasons could sufficiently explain the large differences found
- Therefore we now consider another possible explanation:
- Categorization error.

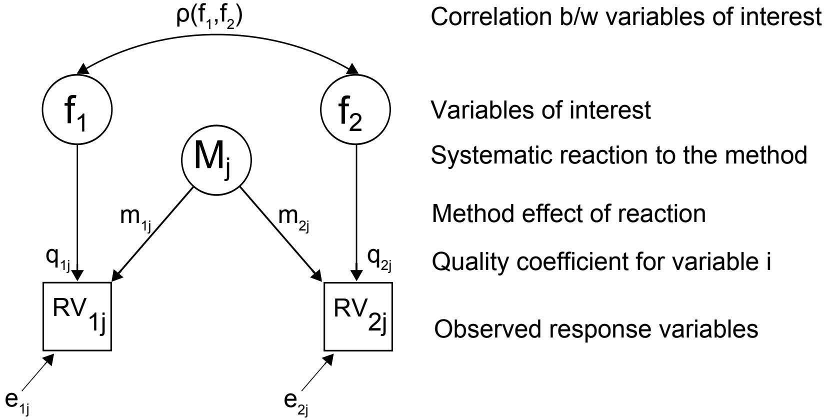
- Big differences in question quality are found across countries
- Previous explanations for the differences we (Oberski ea 2007) sought for were
  - the complexity of the sentences in different languages
  - translation errors
  - different implementations of the experiments used to estimate the quality
- None of these possible reasons could sufficiently explain the large differences found
- Therefore we now consider another possible explanation:
- Categorization error.

- Big differences in question quality are found across countries
- Previous explanations for the differences we (Oberski ea 2007) sought for were
  - the complexity of the sentences in different languages
  - translation errors
  - different implementations of the experiments used to estimate the quality
- None of these possible reasons could sufficiently explain the large differences found
- Therefore we now consider another possible explanation:
- Categorization error.

- Big differences in question quality are found across countries
- Previous explanations for the differences we (Oberski ea 2007) sought for were
  - the complexity of the sentences in different languages
  - translation errors
  - different implementations of the experiments used to estimate the quality
- None of these possible reasons could sufficiently explain the large differences found
- Therefore we now consider another possible explanation:
- Categorization error.

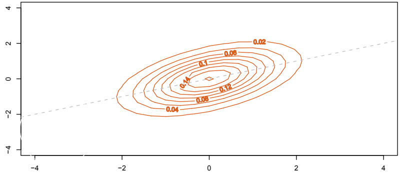
The basic survey response model

The basic survey response model

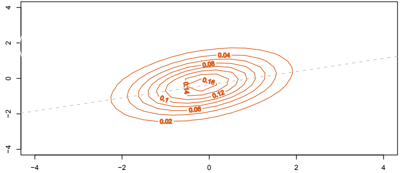


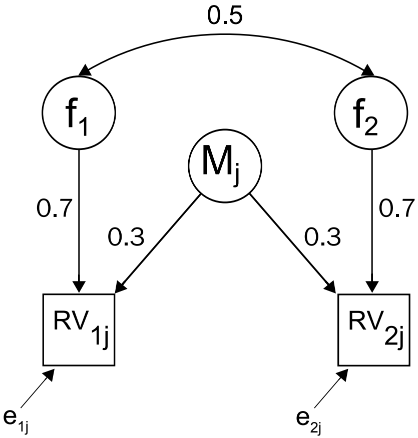
Example

Two variables of interest correlate **0.5**



Measurement error attentuates by a factor 0.5, while method effects of 0.3 add 0.1
Result: oberved resp variables correlate **0.4**





The basic response model

- The square of the quality coefficient q^2 is called the 'total quality' of a measure.
- It is the percentage of variance in the observed variable that can be explained by the latent variable of interest.
- The observed variables are assumed to be continuous.

The basic response model

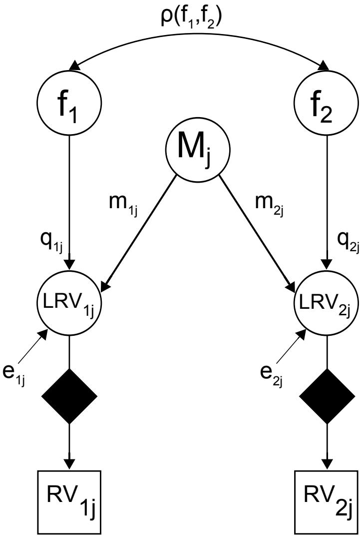
- The square of the quality coefficient q^2 is called the 'total quality' of a measure.
- It is the percentage of variance in the observed variable that can be explained by the latent variable of interest.
- The observed variables are assumed to be continuous.

The basic response model

- The square of the quality coefficient q^2 is called the 'total quality' of a measure.
- It is the percentage of variance in the observed variable that can be explained by the latent variable of interest.
- The observed variables are assumed to be continuous.

The basic response model, revised

The basic response model, revised



Correlation b/w variables of interest

Variables of interest

Systematic reaction to the method

Method effect of reaction

Quality coefficient for variable i

Latent response variables (LRV)

A step function with steps at thresholds τ1, τ2, ..., τk

Observed response variables

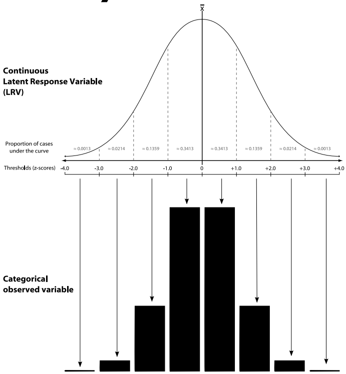
Categorisation of continuous variables

Our model assumes that there are *unobserved* continuous latent response variables (LRV) that have been categorised into the *observed* categorical variables.

Categorisation of continuous variables

Our model assumes that there are *unobserved* continuous latent response variables (LRV) that have been categorised into the *observed* categorical variables.

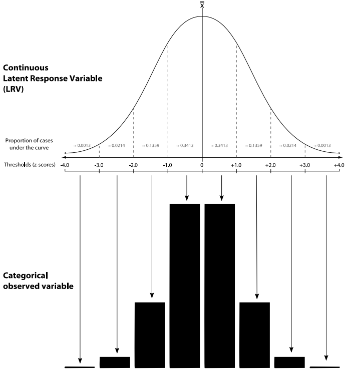
country A



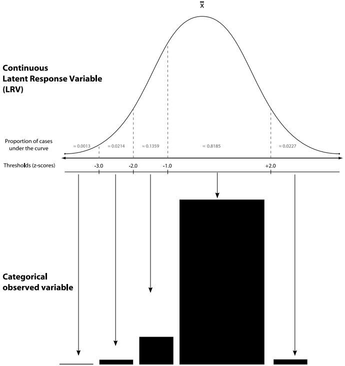
Categorisation of continuous variables

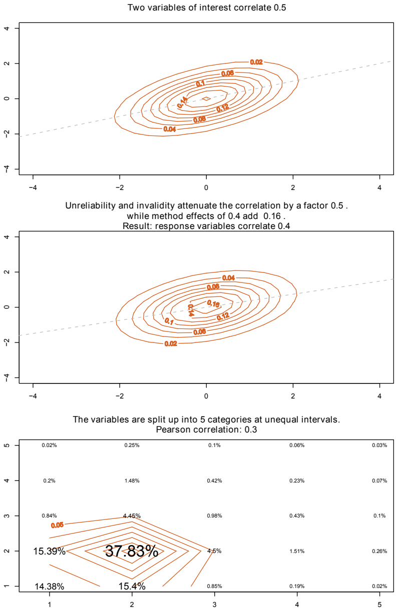
Our model assumes that there are *unobserved* continuous latent response variables (LRV) that have been categorised into the *observed* categorical variables.

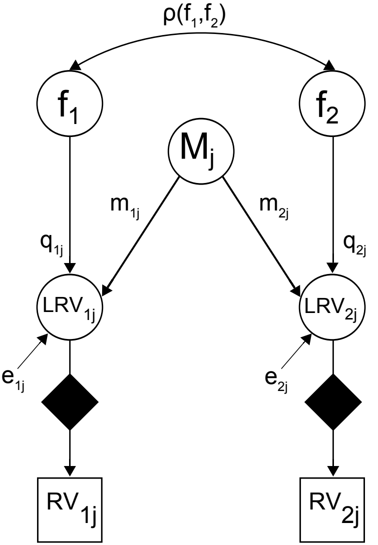
country A



country B

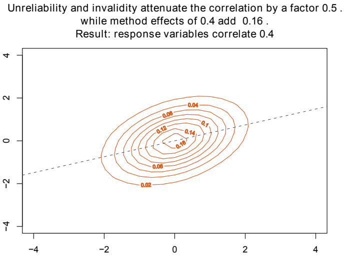


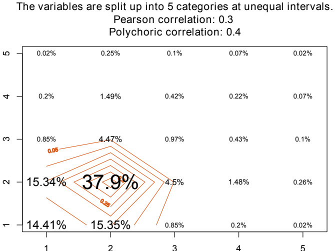

Two variables of interest correlate 0.5
Unreliability and invalidity attenuate the correlation by a factor 0.5 .
while method effects of 0.4 add 0.16 .
Result: response variables correlate 0.4
The variables are split up into 5 categories at unequal intervals.
Pearson correlation: 0.3
37.83%
15.39%
14.38%
15.4%


$\rho(f_1,f_2)$
f_1
f_2
M_j
m_{1j}
m_{2j}
q_{1j}
q_{2j}
LRV_{1j}
LRV_{2j}
e_{1j}
e_{2j}
RV_{1j}
RV_{2j}

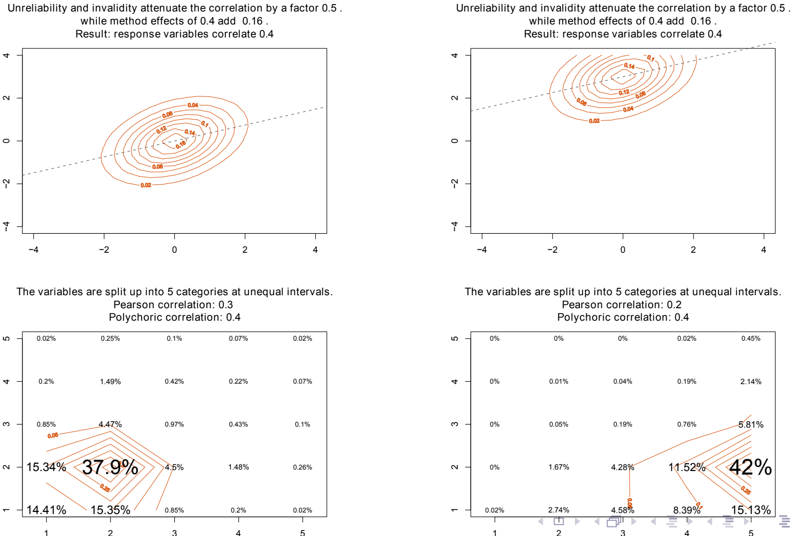
Two countries with equal qualities but different means

Two countries with equal qualities but different means





Two countries with equal qualities but different means



Categorical data in cross-country studies

- The thresholds used earlier are taken from the estimates of a real experiment!
- If the thresholds are different, the observed means and (Pearson) correlations will differ also;
- Even a difference in *means* across countries can cause an observed difference in Pearson correlations;
- If the assumption of normality holds true, the categorical response model (using polychoric correlations) corrects the LRV correlations;
- Whether this model is realistic is the topic of our other presentation...

Categorical data in cross-country studies

- The thresholds used earlier are taken from the estimates of a real experiment!
- If the thresholds are different, the observed means and (Pearson) correlations will differ also;
- Even a difference in *means* across countries can cause an observed difference in Pearson correlations;
- If the assumption of normality holds true, the categorical response model (using polychoric correlations) corrects the LRV correlations;
- Whether this model is realistic is the topic of our other presentation...

Categorical data in cross-country studies

- The thresholds used earlier are taken from the estimates of a real experiment!
- If the thresholds are different, the observed means and (Pearson) correlations will differ also;
- Even a difference in *means* across countries can cause an observed difference in Pearson correlations;
- If the assumption of normality holds true, the categorical response model (using polychoric correlations) corrects the LRV correlations;
- Whether this model is realistic is the topic of our other presentation...

Categorical data in cross-country studies

- The thresholds used earlier are taken from the estimates of a real experiment!
- If the thresholds are different, the observed means and (Pearson) correlations will differ also;
- Even a difference in *means* across countries can cause an observed difference in Pearson correlations;
- If the assumption of normality holds true, the categorical response model (using polychoric correlations) corrects the LRV correlations;
- Whether this model is realistic is the topic of our other presentation...

Categorical data in cross-country studies

- The thresholds used earlier are taken from the estimates of a real experiment!
- If the thresholds are different, the observed means and (Pearson) correlations will differ also;
- Even a difference in *means* across countries can cause an observed difference in Pearson correlations;
- If the assumption of normality holds true, the categorical response model (using polychoric correlations) corrects the LRV correlations;
- Whether this model is realistic is the topic of our other presentation...

How can the quality and thresholds be estimated in different countries?

First trait measured with three methods

CARD 73 Using this card, please tell me how true each of the following statements is about your current job.

		Not at all true	A little true	Quite true	Very true	(Don't know)
G64	There is a lot of variety in my work.	1	2	3	4	8

iS19 The next 3 questions are about your current job. Please choose one of the following to describe how varied your work is.

Please tick one box.

Not at all varied	☐	1
A little varied	☐	2
Quite varied	☐	3
Very varied	☐	4

iS32 Please indicate, on a scale of 0 to 10, how varied your work is, where 0 is not at all varied and 10 is very varied.

Please tick the box that is closest to your opinion

Not at all varied										**Very varied**
0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Three traits measured with first method

CARD 73 Using this card, please tell me how true each of the following statements is about your current job.

		Not at all true	A little true	Quite true	Very true	(Don't know)
G64	There is a lot of variety in my work.	1	2	3	4	8
...						
G66	My job is secure	1	2	3	4	8
...						
G70	My health or safety is at risk because of my work.	1	2	3	4	8

Three traits measured with second method

iS19 The next 3 questions are about your current job. Please choose one of the following to describe how varied your work is.
Please tick one box.

Not at all varied ☐ 1
A little varied ☐ 2
Quite varied ☐ 3
Very varied ☐ 4

iS20 Please choose one of the following to describe how secure your job is.
Please tick one box.

Not at all secure ☐ 1
A little secure ☐ 2
Quite secure ☐ 3
Very secure ☐ 4

iS21 Please choose one of the following to say how much, if at all, your work puts your health and safety at risk.
Please tick one box.

Not at all at risk ☐ 1
A little at risk ☐ 2
Quite a lot at risk ☐ 3
Very much at risk ☐ 4

Three traits measured with third method

iS32 Please indicate, on a scale of 0 to 10, how varied your work is, where 0 is not at all varied and 10 is very varied.
Please tick the box that is closest to your opinion

Not at all varied										***Very varied***
0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

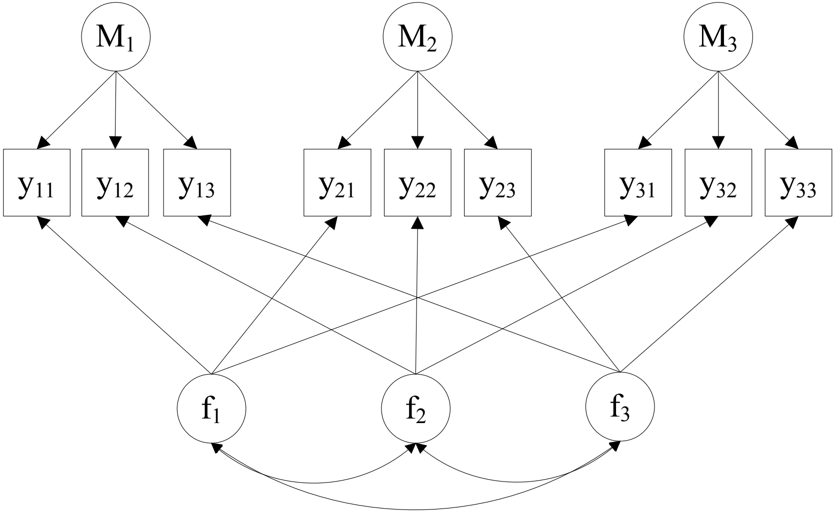
iS33 Now please indicate, on a scale of 0 to 10, how secure your job is, where 0 is not at all secure and 10 is very secure.
Please tick the box that is closest to your opinion

Not at all secure										***Very secure***
0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

iS34 Please indicate, on a scale of 0 to 10, how much your health and safety is at risk from your work, where 0 is not at all at risk and 10 is very much at risk.
Please tick the box that is closest to your opinion

Not at all at risk										***Very much at risk***
0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Classic MTMM model



Conclusions from the analysis of the **continuous** MTMM model

- It is clear that categorization error may cause differences in Pearson correlations across countries;
- This can be expected to have its effect on the estimates of the MTMM model;
- Using the continuous model, quite some differences were indeed found in quality across countries;
- To what extent can these differences be explained by categorization errors?
- In order to quantify the differences between the models across countries we first define a measure called the 'catgorisation factor':

Conclusions from the analysis of the **continuous** MTMM model

- It is clear that categorization error may cause differences in Pearson correlations across countries;
- This can be expected to have its effect on the estimates of the MTMM model;
- Using the continuous model, quite some differences were indeed found in quality across countries;
- To what extent can these differences be explained by categorization errors?
- In order to quantify the differences between the models across countries we first define a measure called the 'catgorisation factor':

Conclusions from the analysis of the **continuous** MTMM model

- It is clear that categorization error may cause differences in Pearson correlations across countries;
- This can be expected to have its effect on the estimates of the MTMM model;
- Using the continuous model, quite some differences were indeed found in quality across countries;
- To what extent can these differences be explained by categorization errors?
- In order to quantify the differences between the models across countries we first define a measure called the 'catgorisation factor':

Conclusions from the analysis of the **continuous** MTMM model

- It is clear that categorization error may cause differences in Pearson correlations across countries;
- This can be expected to have its effect on the estimates of the MTMM model;
- Using the continuous model, quite some differences were indeed found in quality across countries;
- To what extent can these differences be explained by categorization errors?
- In order to quantify the differences between the models across countries we first define a measure called the 'catgorisation factor':

Conclusions from the analysis of the **continuous** MTMM model

- It is clear that categorization error may cause differences in Pearson correlations across countries;
- This can be expected to have its effect on the estimates of the MTMM model;
- Using the continuous model, quite some differences were indeed found in quality across countries;
- To what extent can these differences be explained by categorization errors?
- In order to quantify the differences between the models across countries we first define a measure called the 'catgorisation factor':

The ‘categorisation factor’

- The quality was defined as:
- $$q^2 = \frac{Var(f)}{Var(y)}.$$
- However, we have seen that y is itself a categorization of an unobserved continuous variable (LRV), and therefore the above equation can be ‘decomposed’ into
- $$q^2 = \frac{Var(f)}{Var(LRV)} \cdot \frac{Var(LRV)}{Var(y)}.$$
- We call this ratio of the quality coefficient $q = v.r$ from the categorical model to the same coefficient from the continuous model the ‘categorisation factor’.

The ‘categorisation factor’

$$q^2 = \frac{Var(f)}{Var(LRV)} \cdot \frac{Var(LRV)}{Var(y)}. \qquad (1)$$

- It can be seen that the quality normally estimated from the continuous model is a product of two terms:
- $$q^2_{cont} = q^2_{categ} \cdot c,$$
  where c is a categorisation factor.
- If $c < 1$, the quality in the categorical model is higher than in the continuous model
- If $c > 1$, the quality in the categorical model is lower than in the continuous model

Analysis of the experiments

- We analysed the 4 experiments from the ESS that involved variables with 5 categories or less
- The topics: role of women, doctors, political efficacy, job.

- Compare the country with the highest quality to the country with the lowest quality for that experiment (not discussed here)

Analysis of the experiments

- We analysed the 4 experiments from the ESS that involved variables with 5 categories or less
- The topics: role of women, doctors, political efficacy, job.

- Compare the country with the highest quality to the country with the lowest quality for that experiment (not discussed here)

Analysis of the experiments

- We analysed the 4 experiments from the ESS that involved variables with 5 categories or less
- The topics: role of women, doctors, political efficacy, job.

- Compare the country with the highest quality to the country with the lowest quality for that experiment (not discussed here)

Example: Quality (q^2)and method effects (m) in the efficacy experiment in Denmark

Results of continuous MTMM model, main questionnaire (first method)

	'Efficacy'		
	Complex	Active	Mind
q^2	0.77	0.83	0.79
m	0.00	0.00	0.00

$df = 19, \chi^2 = 40.0, p = 0.003.$

Example: Efficacy experiment in Denmark

Pearson correlations

		Method 1			Method 2		
Method 1	Complex	1.00					
	Active	-0.38	1.00				
	Mind	-0.46	0.41	1.00			
Method 2	Complex	0.60	-0.37	-0.44	1.00		
	Active	-0.39	0.67	0.40	-0.43	1.00	
	Mind	-0.46	0.43	0.62	-0.49	0.48	1.00

Polychoric correlations

		Method 1			Method 2		
Method 1	Complex	1.00					
	Active	-0.44	1.00				
	Mind	-0.51	0.47	1.00			
Method 2	Complex	0.66	-0.45	-0.51	1.00		
	Active	-0.44	0.74	0.46	-0.51	1.00	
	Mind	-0.52	0.51	0.67	-0.56	0.56	1.00

$n \approx 916$

Example: % Increase in the correlations after correction for categorisation

Efficacy experiment: Denmark

		Method 1			Method 2	
Method 1	Complex					
	Active	**16%**				
	Mind	**11%**	**15%**			
Method 2	Complex	10%	22%	16%		
	Active	13%	10%	15%	**19%**	
	Mind	13%	19%	8%	**14%**	**17%**

Mean percentage increase of the polychoric correlations: 14.5%

Example: Quality (q^2)and method effects (m) according to the continuous and categorical models, with categorisation factors

		'Efficacy'	
	Complex	Active	Mind
Continuous analysis			
q^2	0.77	0.83	0.79
m	0.00	0.00	0.00
Categorical analysis			
q^2	0.63	0.70	0.63
m	0.11	0.08	0.11
Categorisation factor			
	1.23	1.18	1.25

Correction for categorisation: conclusions

- The general 'push' is that all coefficients go up, because the polychoric correlations are in general higher than the Pearson correlations;
- But when method factors are taken into account, the coefficients can also go down;
- This happens especially when the method variance is very small (close to zero) in the continuous model, but larger in the categorical model;
- Would then expect countries with high quality in the continuous model to have a lower quality after correction for categorisation and vice versa.

Correction for categorisation: conclusions

- The general 'push' is that all coefficients go up, because the polychoric correlations are in general higher than the Pearson correlations;
- But when method factors are taken into account, the coefficients can also go down;
- This happens especially when the method variance is very small (close to zero) in the continuous model, but larger in the categorical model;
- Would then expect countries with high quality in the continuous model to have a lower quality after correction for categorisation and vice versa.

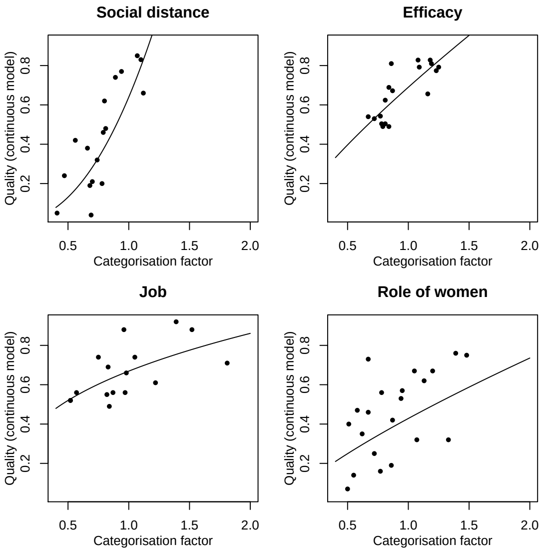
Correction for categorisation: conclusions

- The general 'push' is that all coefficients go up, because the polychoric correlations are in general higher than the Pearson correlations;
- But when method factors are taken into account, the coefficients can also go down;
- This happens especially when the method variance is very small (close to zero) in the continuous model, but larger in the categorical model;
- Would then expect countries with high quality in the continuous model to have a lower quality after correction for categorisation and vice versa.

Correction for categorisation: conclusions

- The general 'push' is that all coefficients go up, because the polychoric correlations are in general higher than the Pearson correlations;
- But when method factors are taken into account, the coefficients can also go down;
- This happens especially when the method variance is very small (close to zero) in the continuous model, but larger in the categorical model;
- Would then expect countries with high quality in the continuous model to have a lower quality after correction for categorisation and vice versa.

The categorisation factor, q_{cat}/q_{cont} versus the quality:



Some implications of the findings

- Using discrete measures (with only a few categories) increased the apparent differences between countries using the continous factor model;
- Differences are smaller after correction for categorization;
- This means that either the continuous model gives misleading results regarding measurement invariance when analysing discrete data
- **OR**
- that the assumption of normality made in the 'revised' (categorical) model is wrong.

Some implications of the findings

- Using discrete measures (with only a few categories) increased the apparent differences between countries using the continous factor model;
- Differences are smaller after correction for categorization;
- This means that either the continuous model gives misleading results regarding measurement invariance when analysing discrete data
- **OR**
- that the assumption of normality made in the 'revised' (categorical) model is wrong.

Some implications of the findings

- Using discrete measures (with only a few categories) increased the apparent differences between countries using the continous factor model;
- Differences are smaller after correction for categorization;
- This means that either the continuous model gives misleading results regarding measurement invariance when analysing discrete data
- **OR**
- that the assumption of normality made in the 'revised' (categorical) model is wrong.

Some implications of the findings

- Using discrete measures (with only a few categories) increased the apparent differences between countries using the continous factor model;
- Differences are smaller after correction for categorization;
- This means that either the continuous model gives misleading results regarding measurement invariance when analysing discrete data
- **OR**
- that the assumption of normality made in the 'revised' (categorical) model is wrong.

Some implications of the findings

- Using discrete measures (with only a few categories) increased the apparent differences between countries using the continous factor model;
- Differences are smaller after correction for categorization;
- This means that either the continuous model gives misleading results regarding measurement invariance when analysing discrete data
- **OR**
- that the assumption of normality made in the 'revised' (categorical) model is wrong.

That's it for now. Moltes gràcies per la seva atenció!

Tilburg

Barcelona

The latent traits

Estimated correlations between the latent traits under the two different models in the example given before

	Complex	Active	Mind
Continuous model	1		
	-.63	1	
	-.75	.66	1
Categorical model	1		
	-.63	1	
	-.75	.70	1

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error **in this experiment** were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this particular case;
- One cannot expect this to be the case in general, however.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error **in this experiment** were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this particular case;
- One cannot expect this to be the case in general, however.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error **in this experiment** were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this particular case;
- One cannot expect this to be the case in general, however.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error **in this experiment** were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this particular case;
- One cannot expect this to be the case in general, however.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error **in this experiment** were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this particular case;
- One cannot expect this to be the case in general, however.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error **in this experiment** were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this particular case;
- One cannot expect this to be the case in general, however.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error **in this experiment** were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this particular case;
- One cannot expect this to be the case in general, however.

Conclusions

- It was possible to split the measurement error model into three parts:
  - A part due to random errors;
  - A part due to systematic errors;
  - A part due to splitting the variable into just a few categories.
- The estimates one gets can differ, and not always in the way one might expect;
- The correlations between the latent traits corrected for measurement error **in this experiment** were robust to the model specification;
- This suggests either model will provide a correct (or at least similar) inference about the variables of interest in this particular case;
- One cannot expect this to be the case in general, however.

Further study, problems

- Investigate normality assumption (tests indicate possible issues), linearity;
- Unobserved heterogeneity;
- Prediction of the data quality based on characteristics of the question.

Further study, problems

- Investigate normality assumption (tests indicate possible issues), linearity;
- Unobserved heterogeneity;
- Prediction of the data quality based on characteristics of the question.

Further study, problems

- Investigate normality assumption (tests indicate possible issues), linearity;
- Unobserved heterogeneity;
- Prediction of the data quality based on characteristics of the question.

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. `http://www.sqp.nl`

The final goal: Survey Quality Predictor (SQP)

1. Estimate the model for all experiments
2. Save the reliability, validity, and method effect coefficients
3. Relate the coefficients to different aspects of the question
   - Complexity of the sentence: no. words/sentence, avg. no. syllables, ...
   - Response scale: type, no. categories, ...
   - Formulation of the request: agree-disagree, extra information, ...
   - Data collection method: computer assisted, interviewer present, ...
4. Predict the quality of survey questions from their characteristics (SQP)
5. Improve survey questions
6. http://www.sqp.nl

References

Oberski, Daniel, Willem Saris, and Jacques Hagenaars
Why are there Differences in Measurement Quality across Countries?
in: *Measuring Meaningful Data in Social Research*. G. Loosveldt, M. Swyngedouw & B. Cambré (eds.). Leuven: Acco. 2007.

Saris, Willem, Albert Satorra, and William van der Veld
Testing Structural Equation models or Detection of misspecifications?
Forthcoming in *Structural Equation Modeling: an interdisciplinary journal.*