

LEUVEN



Assessing cross-cultural measurement equivalence taking differences in response style into account

Jaak Billiet CeSO – KU Leuven

outline

- 1. Introduction: Response styles and how to measure them?
- 2. Modelling acquiescence (ARS): characteristics of the model
- 3. Modelling ARS in multi-group situations: overview of successful examples = select 1 (nr 4)
- 4. (partial) failure (ess R4 welfare concept : WHY?

discussion

1.

Introduction:

Response styles and how to measure them

Response styles

- Response style = systematic tendency to respond to a range of survey items on some basis other than that which the items were specifically designed to measure (Paulhus, 1991)
 - is systematic kind of "measurement error"
 - is typical for set of items using same response format (multiple indicators that measure latent variable)
 - can be different according to cultural groups
 - there are ways to detect it, and to control for it

Kinds of response styles

- Tendency to endorse assertions independently from their content = acquiescence (yes-saying) = ARS
- Tendency to deny assertions independently from their content (no-saying) = DRS
- Tendency to choose extreme response categories of response scales independently from content of items = extreme response style = ERS
- Tendency to choose the middle of a response scale =
 midpoint responding independently from their content =
 MRS

Challenge: how to measure style independently from content

How to measure response style (RS)?

Typology of RS measures in 2 dimensions (Weytens, 2006)

- **I.** status of the items on which response style measures are based
 - (A) multifunctional: measure both Content + RS
 - (B) specific RS measure

II. the *treatment of content* in these items

- (1) no specific ex ante control (content of items not deliberatively planned or selected before data collection) response style computed ad hoc on available items
- (2) content can be eliminated with aim to measure style
- (3) content can be manipulated (e.g. opposite meanings in set of items) in order to cancel out the effect of content
- (4) content is randomized so that there is no systematic influence of content on response

How to measure response style (RS)?

I. Treatment of content	II. Function of item set used for response style			
Treatment of content	A. Multi-functional	B. Specific measure for RS		
1. No ex-ante control for content	A1 Try to detect additional RS factor besides content factors Neg: confounding content & RS See example of ESS R4 (attitude Social Security)	B1 Try to measure RS in items without control for content (e.g. # of agreements, MA in other items) neg: not possible to disentangle RS and content		
2. Elimination of content	(A2) Not possible by definition not multi -functional	B2 Try to develop content free items that only measure RS directly (neg: what is studied is guessing, not RS)		
3. Experimental control	Possible in MTMM or in case of positively and negatively worded items (ARS & DRS)	B3 Separate measurement of items and their reversion. Items not further used for substantive reasons, only for RS		
4. Randomization of content	(A4) Not theoretical meaningful since items not used to measure specific content	B4 Used in marketing research Include large additional set of random items in which no correlation is expected. Correlation = response style		

How to measure response style (RS)?

B4

Use large set (e.g. 50 items) of (assumed) mutually independent items

Apart from the target items that measure a content (Greenleaf, 1992; Baumgartner & Steenkamp, 2001; Weytens, 2008)

Measure of RS = correlation between independent items

(one can find out what increases the correlation: extreme, middle, agreement or disagreement)

Disadvantage:

- many additional items just for measuring style

Advantages:

- useful for the four style effects
- possible to include the style measures in substantive regression models with content variables

How to measure response style?

A3.

the MTMM form = repeated measurements with variation in traits and response formats (dependent of the RS one wants to measure)

advantage: directly included in structural models

disadvantages =

- inflation of items
- difficult to distinguish between style and method effect (response scale)



How to measure response style?

- the balanced set of items form advantages = items that are designed for measuring content are used for measuring RS if they are balanced
 - directly included in **structural models** with relations between content variables
 - the <u>latent variable</u> approach with SEM (Billiet & McClendon, 2000)

disadvantage = only useful for ARS and DRS¹ (next part is focused on this approach)

¹ mostly theoretical...

■ the <u>latent class</u> approach (*Moors, Q&Q 2003; Moors, ESR 2004;*

Kankaras & Moors, 2011)

specify an extra latent class identified as RS

advantage: possible to model ARS, ERS, MRS

disadvantage: very large samples, fewer indicators

for concept



2.

Modelling acquiescence (ARS): characteristics of the model



Modelling acquiescence

- Focus further on ARS in the balanced set approach
 (A3)
- Previous work on ASR:
 - known since 1927 (Cronbach...)
 - Explained as:
 - impression management (positive image of oneself) (Ross & Mirowsky, 1984)
 - desire to satisfice (minimum cognitive activity) (Krosnick 1991, 1992, 2005)
- Related to background variables:
 - Education = lower educated more yes-saying
 - Age = older respondents more yes-saying
 - Gender?
 - characteristics of society? (cultural norm not to say NO)

How to measure ARS?

- In case of **single items**: providing opposed assertions in split ballot, or in repeated measurements or avoid by using forced choice items (*Schuman & Presser, 1981*)
- In case of **multiple indicators** per theoretical variable (as commons source of variation: *congeneric measures*)
 - **1. Index of yes-saying** (# of times YES in sets of items about various contents) (Watson, 1992) problem = not independent from content variables
 - 2. Use strict reversals: count double agreements (difficult...)
 - **3. Use balanced sets of items** per concept: balanced = positively and negatively worded items
 - use composite scores after reversing half of items (yes-saying in middle)
 - OR apply **structural equation models** for content variables plus additional style factor (Billiet & McClendon, SEM 2000)

ARS with SEM: Expectations and conditions

Expectations about the model:

- In **one** balanced set of items
 - Acquiescence can be identified as a common factor apart from content
 - the common style factor will have a non-zero variance which is smaller than the variance of the content factor
- In two or more balanced sets of items per content (concept)
 - One common style factor should be found in two (or more) balanced sets of indicators
 - If ARS: should correlate strongly with # agreements
 - If ARS: stable over time

Conditions:

- substantial number of double agreements in (quasi) balanced sets
- (quasi) balance within each set
- for test on stability: panel data (see paper JB & ED 2008)

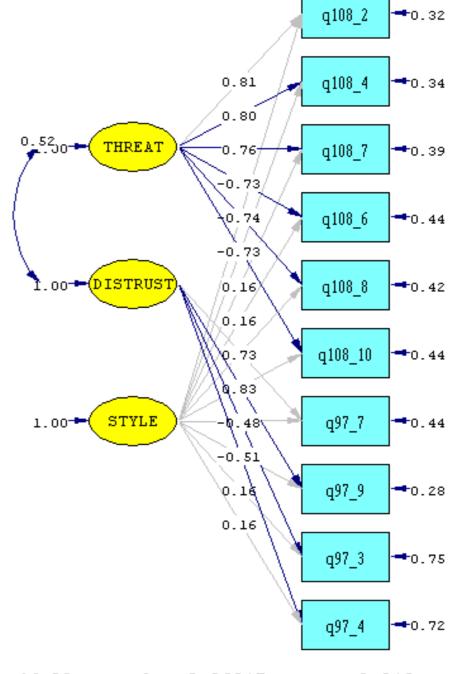
example

- Balanced set with six items on ethnic threat and four items on distrust in politics in Flanders and Wallonia (1995: ISPO data)
- Exploration in one Flemish subsample (N= 986) and confirmation in other Flemish subsample (N = 1,114) and in Walloon sample (N=1,200)
- Scoring: completely agree = 5
 completely disagree = 1
 (otherwise negative slopes for STYLE)
 Do not reverse item-scores in model
- Test of possible models: model with STYLE preferable
- Model specifications: see Billiet & McClendon, SEM 2000

The items

Item number	Balanced sets of items
v108_2	In general, immigrants are not to be trusted.
v108_4	Guest workers endanger the employment of the Belgians.
v108_7	Muslims are a threat for our culture and customs.
v108_6	The immigrants contribute to the prosperity of our country.
v108_8	The presence of different cultures enriches our society.
v108_10	We should kindly welcome the foreigners who come to live here.
v97_7	The politicians have lost the ability to listen to ordinary people like me.
v97_9	Once they are elected, most politicians feel themselves too good for
	people like me.
v97_3	If people like me make their views know, the politicians generally take
	them into account.
v97_4	Most of our politicians are able people who know what they are doing.

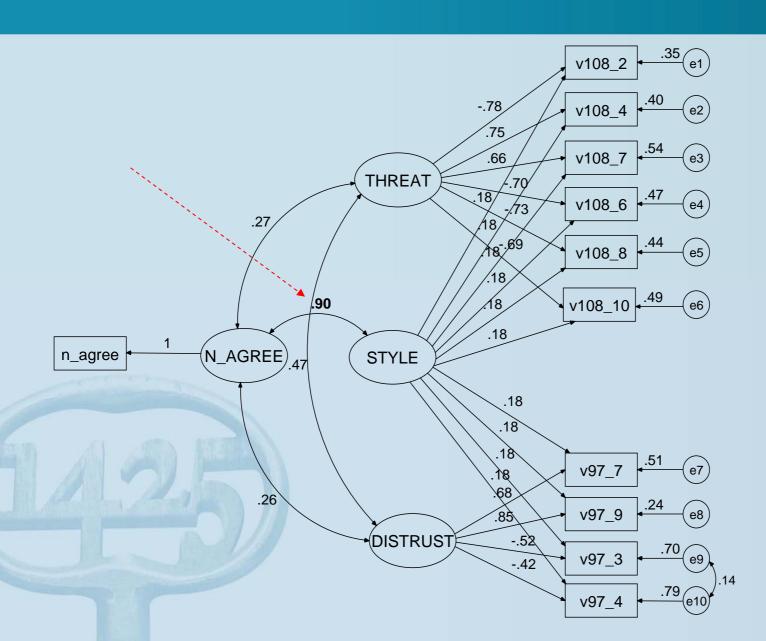
The model Walloon sample



Further comments on this model

- **Is it ARS** or tendency to choose first response alternative on response card, or choose 1) arguments:
 - Negative correlation with education (r = -.23; t = -3.242)
 - Positive correlation with age (r = .15; t = 2.749)
 - Very strong correlation of style factor with # agree in 14 balanced items (+0.90 see next figure) (is however also expected in case of recency effect)
 - In ISSP 1995: "decrease-increase" item within balanced set on immigrants has no significant loading on STYLE factor

see model (in Flemish exploratory sample)



Further comments on this model

- Is it stable over time? (see paper Billiet & Davidov)
 - test in two waves of *Belgium General Election Survey* (Flanders) 1995-1999 (N = 1,112)
 - scalar (& metric) invariant model over waves
 - correlation ARS_{95} - ARS_{99} = **0.56** see structural relations in next slide

Table 2: Scalar invariant models and their Indices of Model Fit (N = 1,112)

Model	Chi-Square	DF	RMSEA	Pclose
Model 1: 2x2 content factors no style	405.57	162	0.037	1.00
Model 2: Style factors, cor(St95, St99)=0	357.20	160	0.033	1.00
Model 3: Correlated style factors St95-St99	344.16	159	0.032	1.00

Table 4: Correlations between Content and Style Factors in Model 3 (T-Values in Parentheses).

	THR95	DISTR95	STYLE95	THR99	DISTR99	STYLE99
THR95	1.000 (37.30)					
DISTR95	0.496 (15.12)	1.000 (19.04)				
STYLE95			1.000 (5.85)			
THR99	0.859 (34.79)	0.504 (16.42)		1.000 (26.23)		
DISTR99	0.523 (17.36)	0.736 (18.34)		0.592 (19.97)	1.000 (20.68)	
STYLE99			0.562 (3.61)			1.000 (5.36)

3.

Modelling ARS in a multi-group situation: short overview of successful examples

Successful examples of ARS & measurement invariance

- **1. M**easurement **e**quivalent (ME) model with <u>two full-balanced</u> sets of items with ASR in Flanders and Wallonia
- 2. ME model for relation between <u>one full balanced</u> set (Ethnic treat) and a concept (sub-national consciousness) measured with <u>mixed</u> response scales: Flanders and Wallonia (1999 BGES)
- 3. ME model with <u>one quasi-balanced</u> set (asylum items in ESS 2002) in search of detection of "*lost in translation*" in four countries (French language)
- 4. ME model with two unbalanced sets of items and with mixed set (relation between sub-national consciousness, ethnic threat and xenophobia in Flanders and Wallonia) (2007 BGES)

BGES = Belgian General Election Survey

Example 1: two balanced concepts in two samples (Dutch 1900; & French 1100)

6 ethnic threat and 4 distrust items Model with invariant slopes (metric invariant

Table. Comparsion of a metric invariant model for ethnic threat and political distrust in Flanders and Wallonia without and with a Style factor

Models	Chi-square	df	RMSEA	P-value of close fit	NFI
Model 1: No Style factor	431.01	76	0.059	0.001	0.968
Model 2: With Style factor	193.46	75	0.033	1.00	0.986

^{*} scalar invariance not tested

Example 2. one fully balanced set and mixed response scales for second concept

- Relation between sub-national consciousness and ethnic threat (BGES 1999) (Billiet, Maddens & Beerten, Politcal Psychology 2003)
 - balanced set for eth threat (MIGRANT): 3 pos and 3 neg (see previous examples)
 - National consciousness (NAT_ID) questions on
 - independence of FI/Wal (10 p)
 - split of social security (likert 5p)
 - what level should decide (10p)
 - scale based on first and second identification with FL/Wal-Belg
 - Exclusive identification (exclusive Belg --- exclusive FL/Wal 5p)

Example 2. Comparison of the completely constrained models without (Model a) and with a method factor (Model b)

Models	Chi- square	Df	RMSEA	p-value of close fit	NFI
Model a: factorial invariant: no Style factor	640.71	109	.078	.391	.980
Model b: factorial invariant; Style factor	585.73	108	.068	.682	.982

See next page:

Slope parameter "splitting social security" is not metric invariant Indeed: it is not in favour of the Walloons but according to the Flemish in favour of them.

As expected: much stronger correlation of item with sub-national consciousness in FI than in Wal sample

Example 2: the selected model

Indicators	Λ^1 : Flanders			Λ^2 : Wallonia	Λ ² : Wallonia	
	NAT_ID	MIGRANT	STYLE	NAT_ID	MIGRANT	STYLE
			(all fixed)			(all fixed)
1. Independ	.57 (fixed)			.57 (fixed)		
2. Soc_sec	.58 (19.33)			.29 (8.58)		
3. Decide	.73 (21.76			.73 (21.76)		
4. First_id	.72 (21.60)			.72 (21.60)		
5. Exclus_id	.77 (21.49)			.77 (21.49)		
6. Distrust		.81 (fix)	.17		81 (fix)	17
7. Employ		.78 (39.98)	.17		.78 (39.98)	.17
8. Culture		.74(35.85)	.17		.74(35.85)	.17
9. Prosperity		74 (-42.16)	.17		74 (-42.16)	.17
10. Enriching		75 (-38.39)	.17		75 (-38.39)	.17
11. Welcome		713 (-37.34)	.17		713 (-37.34)	.17
Correlations	NAT_ID	MIGRANT	STYLE	NAT_ID	MIGRANT	STYLE
NAT_ID	1.0 (10.90)			1.0 (10.59)		
MIGRANT	.09 (1.96)	1.0 (33.10)		10 (-2.6)	1.0 (31.21)	
STYLE	.0	.0	1.0 (5.09)	.0	.0	1.0 (5.09)

Example 3. four countries quasi balanced

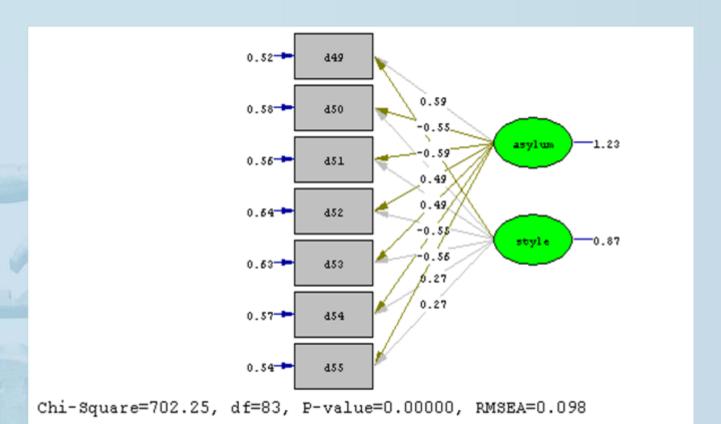
Asylum items in R-ESS 2002 4 samples with questionnaires in French = FR, LU, canton Genèva of CH and Walloon sample of BE Reason for test: translation problem with item D51 expected in France

- D49 [Country] has more than its fair share of people applying refugee status (-)
- D50 People applying refugee status allowed to work while cases considered (+)
- D51 Government should be generous judging applications for refugee status (+)
- D52 Most refugee applicants don't fear persecution in own countries (-)
- D53 Refugee applicants kept in detention centres while cases considered (-)
- D54 Financial support to refugee applicants while cases considered (+)
- D55 Granted refugees should be entitled to bring close family members (+)

(completely disagree 1 – completely agree 5)

Example 3...

Model	Chisq	df	RMSEA	P(close fit)	Model CAIC
Mo: basic model invariant (A)	1,842.78	92	0.133	0.000	2,291.57
Mo: basic model invariant (A+S)	1,304.60	84	0.116	0.000	1,829.35
M1: free τ^{FR}_{3}	860.92	83	0.093	0.000	1,395.04
M2: free τ ^{LU} ₅	738.80	82	0.086	0.000	1,282.29



Example 4. two unbalanced sets and mixed set in two groups (Flemish and Walloon samples of GBES 2007)

- Measurement of (sub)national consciousness (NAT_ID) in Flanders and Wallonia: set of 4 indicators with mixed response scales
 - (soc_sec item dropped because in 2007 in opposite direction related with the 4 other indicators the two samples)
- Two other concepts (ethnic threat and Islamfobia) all agreedisagree items
 with 6 negatively worded and 2 positively worded items in each set
- Substantive question: is opposite relation between nat_id and ethnic threat a stable finding
- Meth. question: is model with style factor still possible? (12 versus 4 items over de two sets)



Example 4: observed indicators for perceived ethnic threat (ISPO 2007) (5p disagree---agree items)

Item	Ethnic threat
Q114_1	In general, immigrants are not to be trusted (-)
Q114_2	Immigrants contribute to the country's welfare (+)
Q114_3	Guest workers come here to take advantage of our social security system (-)
Q114_4	Immigrants are a threat to our culture and customs (-)
Q114_5	The presence of different cultures enriches our society (+)
Q114_6	Most immigrants are lazy, who try to avoid hard work (-)
Q114_7	Guest workers are a threat to the employment of Belgians (-)
Q114_8	Immigrants' way of life is irreconcilable with Western Europeans' way of life (-)

Example 4: observed indicators for Islamphobia (ISPO 2008) (5p disagree---agree items)

Item	Islamfobia
(D32_1)	The Islam can contribute to the European culture (+)
D32_2	Muslim men dominate their wives (-)
D32_3	Muslims do attach great importance to their children's education (-)
D32_4	If it really matters Muslims turn against Europe (-)
D32_5	The Islamic culture and history are more violent than others (-)
D32_6	Islamic values are a threat to the European culture (-)
(D32_7)	Most Muslims have respect for our culture and our way of living (+)

Observed indicators for (sub)national consciousness (ISPO 2008)

4	
Item	(Sub)national identity
First_id	4-point scale (0 = first identification with Belgium 3 = first identification with Flanders/Wallonie)
Exclus_VW	5-point scale (1 = exclusive Belgium 5 = exclusive Flemish/Walloon
Decide	11-point scale (0 = Belgium must decide 10 = Flanders must decide)
Split_B	5-point scale (1 = Unitarian Belgium state 5 = split the state



LEUVE ONIVERSI

Table: Equivalent measurement model (scalar and metric invariance) in the Flemish and Walloon samples

part I: measurement part - standardized factor loadings (response style =
acquiescence = tendency to agree with all)

ı						
	Items	Ethnic threat	Islamophobia	(Sub)national	identity	Response style
		(in both samples)	(in both samples)	Flemish W	<i>lalloon</i>	(in both samples)
	Q114_1	0.797				0.112
	Q114_2	-0.751				0.112
	Q114_3	0.817				0.112
	Q114_4	0.873				0.112
	Q114_5	-0.781				0.112
	Q114_6	0.791				0.112
	Q114_7	0.746				0.112
	Q114_8	0.829				0.112
	D32_1		-0.773			0.112
	D32_2		0.627			0.112
	D32_3		0.608			0.112
ø	D32_4		0.836			0.112
	D32_5		0.813			0.112
Á	D32_6		0.902			0.112
A	D32_7		-0.705			0.112
	First_id			0.793		
	Exclus_VW			0.789		
	Decide			0.632	0.749	05
	Split_B			0.750		35

Table 1: Equivalent measurement model in the Flemish and Walloon samples. **Part II**: structural model

Stand. cov.	Ethnic threat	Islamophobia	(Sub)national identity	Response style
Flanders				1 ,
Threat	1.000			
Islamophobia	0.790	1.000		
(Sub)national	0.259	0.319	1.000	
Resp. style	- \	\		1.000
Stand. cov	Ethnic threat	Islamophobia	(Sub)national identity	Response style
Wallonia			(2002)	response style
Threat	1.000			
Islamophobia	0.790	1.000		
(Sub)national	-0.243	-0.240	1.000	
Resp. style				1.000

Example 4: conclusions and questions

- Full scalar and partial metric invariant After drop of "splitting of social security" item, NAT_ID is (see model)
- Correlation is negative in Wallonia and positive in Flanders (as expected according to theoretical expectations)
- Partial metric equivalence is indication that meaning of NAT_ID is different in samples: nationalism in Flanders and regionalism in Wallonia (in line of world knowledge)
- It is possible to model a style factor even in very unbalanced sets (condition is enough reversed wordings over the sets)
- When is it not longer possible: see next pages on failures

4. (partial) failure WHY?



Example 4: four concepts on welfare state ESS round 4 (2008)

- At occasion of publications of Meuleman et al. on the multidimensionality of welfare state legitimacy (Meuleman, JSW 2011; Meuleman & Van Oorschot, IJSW, 2006)
- Proposed a model with STYLE factor (footnote in coming publication) with ESS data 2008.
- Reflections on this model:
 - is the style factor ARS?
 - is it possible to model ARS with these items
 - why not?
 - how to solve in principle?

Example 4. the items for four dimensions (concepts)

EQUAL: income equality

- The government should take measures to reduce differences in income levels (+)
- Large differences in people's incomes are acceptable to properly reward differences to obey authority (-)
- For a society to be fair, differences in people's standard of living should be small (+)

MORAL_CO: moral consequences of WS

- Social benefits and services make people lazy
- Social benefits and services make people less willing to care for one another
- Social benefits and services make people less willing to look after themselves and their family

Example 4. the items...

ECO_CONS: economic consequences of WS

- Social benefits place too great a strain on the economy
- Social benefits cost business too much in taxes and charges

SOC_CONS: social consequences of WS

- Social benefits prevent widespread poverty
- Social benefits lead to more equal society
- Social benefits make it easier for people to combine work and family life

Attention: items of ECON_CONS might be in contrast with items of SOC_CONS but these are all in same direction within the concepts

Comparison between ME model with and model without a STYLE factor in two samples (Flemish, Walloon)

ME model is full metric and scalar invariant

Models	Chi- square	Df	RMSEA	p-value of close fit	NFI
Model a: factorial invariant: no Style factor	233.17	90	0.043	0.945	0.918
Model b: factorial invariant; Style factor	148.99	85	0.031	1.00	0.946

correlations of four concepts with STYLE are fixed in model 2

The model

Common Metric Completely Standardized Solution

LAMBDA-Y

	EQUALTY	ECO_CONS	SOC	C_CONS	MORAL_CO	STYLE
b30_b	0.629					0.261
d1_b	-0.600					0.262
d4_b	0.654					0.261
d21_b		0.661				0.261
d25_b		0.597				0.261
d22_b	-3-1			0.473		0.261
d23_b				0.953		0.261
d26_b				0.327		0.261
d27_b					0.780	0.261
d28_b					0.723	0.260
d29_b	_				0.818	0.261

Structural relations (Flemish sample)

	EQUALTY	ECO_CONS	SOC_CONS	MORAL_CO	STYLE
EQUALTY	0.454 (0.041) 11.083				
ECO_CONS	-0.129	0.419 (0.050) 8.409			
SOC_CONS	(0.019)	-0.053 (0.021) -2.510			
MORAL_CO	(0.026)	0.263 (0.029) 8.999	(0.019)	(0.036)	
STYLE					0.039 (0.009) 4.425
Correlat	ion Matrix	of ETA			
	EQUALTY	ECO_CONS	SOC_CONS	MORAL_CO	STYLE
EQUALTY ECO_CONS SOC_CONS	0.238	1.000			
MORAL_CO STYLE	-0.185 	0.537 	-0.168 	1.000	1.000

Structural relations (Walloon sample)

	EQUALTY	ECO_CONS	SOC_CONS	MORAL_CO	STYLE
EQUALTY	0.312				
LQ011L1	(0.040)				
	7.833				
	,,,,,				
ECO_CONS	-0.108	0.464			
	(0.029)	(0.074)			
	-3.727	6.300	Jan		
SOC_CONS	-0.014	-0.094	0.238		
	(0.017)	(0.027)	(0.043)	- <i>)</i>	
	-0.871	-3.511	5.490		
MORAL CO	-0.084	0.293	-0.074	0.668	
MORAL_CO	(0.028)	(0.046)		(0.056)	
	-3.035	6.434	-3.171	12.004	
	3.033	0.454	3.171	12.004	
STYLE				-0.057	0.113
				(0.026)	(0.018)
				-2.180	6.178

	EQUALTY	ECO_CONS	SOC_CONS	MORAL_CO	STYLE
E0113 T E17	1 000				
EQUALTY ECO_CONS	1.000 -0.284	1.000			
SOC CONS	-0.254	-0.282	1.000		
MORAL CO	-0.183	0.526	-0.185	1.000	\ \
MORAL_CO STYLE	-0.103	0.526	-0.103	-0.209	1.000
SIILE				-0.209	1.000
OHIO CONTRACTOR				*****	

discussion

- Is it response style?
- Rather strong correlation with # agree"
- not ARS!
 Why? No mix of positive and negative items within each concept (dimension)
- Confusion with content (see Walloon sample)
 see approach A1 (p. 7)
- Possible to combine ARS wit MTMM?
 - if 4 measures per threat each with 3 response different scales
 - all agree disagree items
 - one of these reversed wording (1 neg and 1 positive) varying over scales