

Report about the MTMM experiments in Round 3

Willem Saris
Irmtraud Gallhofer
Lluís Coromina
Daniel Oberski

ESADE

There were two reasons for the proposals for MTMM proposals. The first was that some experiments done in the pilot study of round 3 were not conclusive and the second was that some basic questions, used in the core, were not evaluated with respect to their quality. Given these two reasons the MTMM experiments for the Round 3 data collection has been formulated. We will discuss the proposed designs in sequence starting with the proposals based on the pilot experiments.

1.1 Proposals based on the pilot experiments

In experiment 1 of the Pilot we tried to see if the number of categories increases the quality of the question. This was indeed true for the 11 point scale compared with the 5 point scale. However the results for the 7 point scale were different from the expectations. The reversal of the numbers at the categories may be the cause.

Given that it is important to know if these results hold for all countries involved, we suggested to repeat this experiment for all countries but with the corrections suggested above and limiting the experiment to the 3 positive items of the set of items

The second experiment concerned the hypothesis that the effect of variation in the use of scales will reduce the method effects. Although the hypothesis sounded plausible the effects were not found in the pilot study because the method effects were not significant. A possible reason for the lack of method effects in this case is that the items in this scale represented positive and negative points of view and a positive item was always followed by a negative one and vice versa. This means that respondent has to think about the use of the scales anyway. If one answers the questions a bit attentively one has to switch from agree to disagree all the time. This seems to have happened here.

The conclusion was that this experiment cannot be done as it has been done. We suggested to do it in a different way. Our suggestion was to select only the three positive items from the items 45-50 and use them in the experiment. In doing so we can also see if balancing the scale reduces the method effect because in the main questionnaire a balanced scale will be used.

1.2. Proposals based on Core questions

Given the above specified experiments there was still space for more experiments in the supplementary questionnaire. We suggested to experiment with two topics of the core questionnaire that have not been evaluated yet. That are the immigration questions and the consequences of more immigrants.

The first set is measured with a 4 point agree/disagree format in the main questionnaire. We suggested in group 1 to repeat the 5 point scale, in group 2 to use a 4 point scale and in group 3 a 7 point scale. In this way the variation of the scales experiment can be extended and we get more information about the effect of the number of categories.

The consequences of immigration is asked using an 11 point scale with anchored end points. This approach is arbitrary. One could also have used statements with an agree/disagree format. In groups we suggested to use a 5 point agree/disagree scale and

11 point agree/disagree scale in group 2 . Finally we suggested to use in group 3 a 7 point scale.

These experiments were included in the supplementary questionnaire of the pilot study (with a few adjustments) and will be evaluated when the data are available.

1.3. Summary of the proposal

This leads to the design of the experiments as indicated in Table 1.

Table 1 The combination of scales in the main and supplementary questionnaires

Questions	Main Questionnaire all	Scales used in the supplementary questionnaires of the different groups		
		group 1	group2	group 3
Immigration B35-B37	4 TS	5 A/D	4 TS	7 D/A
Consequences B38-B40	11 TS	5 A/D	11 D/A	7 D/A
Satisf. work E26 –E28	5 A/D	5 A/D	11 TS	11 D/A
Satisfy contacts E40, E43,E45	5 A/D	5 A/D	5 A/D	7 D/A

TS = trait specific question A/D = Agree /disagree scale D/A = Disagree/agree scale

In this design each group gets 12 extra questions and we can evaluate:

- 1 the quality of the measure of immigration
- 2 the quality of the measure of the consequences of more immigration
- 3 The effect of the number of answer categories
4. the quality of the trait specific scales can be compared with the quality of different agree/disagree scales
5. the effect of the similarity of scales after 2 sets and 3 sets of the same scales.
6. the cross cultural comparability of the 11, 7 and 5 point scale can be compared

2. Results of the MTMM experiments

The results of the tests are based on the analysis of 14 countries which delivered their data in time to be available in October 2007. The Scandinavian countries have been omitted for the moment because in these countries another mode of data collection has been used. Furthermore, Cyprus and Rumania have been omitted because the data were not collected in accordance with the rules specified. This complicated the analysis too much at this moment. The data of these countries will be analyzed together with the data from the countries which were not available at the moment that the analysis has been started.

In all 14 countries 4 MTMM experiments were done according to the SB-MTMM design (Sarlis et al 2004). With this approach the reliability, validity, method effect can be estimated and the total quality (reliability * validity) of each question can be determined. These quality indices have been estimated, put together in a data base and on the basis of these data the results in Table 2 have been calculated. Table 2 presents the mean quality of the questions for each topic and each form.

Table 2 The mean quality of scales in the main and supplementary questionnaires

Questions	Main Questionnaire all	Scales used in the supplementary questionnaires of the different groups		
		group 1	group2	group 3
Immigration B35-B37	.75 (4 TS)	.59 (5 A/D)	.91 (4 TS)	.59 (7 D/A)
Consequences B38-B40	.68 (11 TS)	.51 (5 A/D)	.32 (11 D/A)	.36 (7 D/A)
Satisfaction with work E26 –E28	.56 (5 A/D)	.75 (5 A/D)	.65 (11 TS)	.62 (11 D/A)
Satisfaction with contacts E40, E43,E45	.65 (5 A/D)	.61 (5 A/D)	.56 (5 A/D)	.45 (7 D/A)

TS = trait specific question A/D = Agree /disagree scale D/A = Disagree/agree scale

Based on this table several of the questions raised above can be answered. We start with the quality of the two sets of questions from the core questionnaire

1. The quality of the measure of immigration

From the first row of Table 2 it is quite clear that the form chosen for the main questionnaire, using a “trait specific” scale, is far better than the two agree / disagree scales even though the number of categories is less and irrespective of the direction of the scale from agree to disagree or from disagree to agree. The quality of the questions in group 2 are better but these items were exactly the same as those of the main questionnaire. Therefore we think that the difference must be caused by the repetition of these questions which are very much debated in all European countries so that people not easily forget their previously given answer.

2. The quality of the measure of the consequences of more immigration

The same phenomenon, namely that the Trait specific question are much better than the Agree/disagree question with batteries, can also be seen for the topic Consequences of immigration. Here the trait specific questions are again much better irrespective of the number of response categories. This also holds for the 7 and 11 point scales which both have a very low quality.

3. The effect of the number of answer categories

This brings us to the evaluation of the effect of the number of categories. The above results show very clearly that an increase in the number of categories not automatically means that one gets better data. More important is the use trait specific scales. For the topic immigration a trait specific 4 point scale is better than a 5 point and 7 point A/D scale. For the topic Consequences of immigration the trait specific 11 point scale is even twice as good as an 11 point A/D scale. For the work satisfaction topic the 11 point trait specific scale is also better than the 11 point D/A scale but the difference is not so big¹. We also see that the 7 point scale gives rather bad results. This requires further research because this scale is each time used in the D/A form. This result has also been found for

¹ A deviating result has been obtained in group 1 for this topic but we suppose that this is again a memory effect because the questions for this topic in this group were exactly the same as those in the main questionnaire.

the same experiment in the pilot of round 3 Saris and Gallhofer (2007). This seems to suggest that there is a special problem with the 7 point scale.

4. The quality of the trait specific scales compared with agree/disagree scales

From the above results it is very clear that the use of trait specific scales has important advantages with respect to the quality of the questions. In all cases where a trait specific scale has been used the quality is considerably better than for any form of an A/D scale. This point has been made by Saris and Gallhofer (2007). The reason is that the answer categories are much clearer for the respondents and that the cognitive effort for them is also much less (Krosnick, forthcoming).

5. The effect of the similarity of scales on the quality

For the third and the fourth topic the questions in the main questionnaire and in the first form of the supplementary questionnaire (here denoted as group 1) are exactly the same. The only difference between these questions is that in group1 more questions with the same format were asked before these specific questions. For the topic satisfaction with work it looks as if the repetition of the same format has a positive effect on the quality of the questions because the quality in group 1 is much higher than in the main questionnaire.. However, if that is so, this should be even more so for the topic satisfaction with contacts because there in group 1 the same format is once more repeated. But now the quality of the questions in the main questionnaire is better than those for group 1. Given these contradictory results we can not say that similarity of scales or variation of the scales in the questionnaire improves or reduces the quality of the questions.

6. The cross cultural comparability of the 11, 7 and 5 point scale

The cross-cultural comparability of the different scale formats can be evaluated in more or less detail. We have decided to evaluate it on a high aggregation level because it is very unlikely that the comparability is better on the lower levels of aggregation if it is not at a higher level of aggregation. So we have estimated the standard deviation across countries of all scales and topics specified in Table 1. The results are summarized in Table 3.

Table 3 The variation (standard deviation) in quality of scales across countries in the main and supplementary questionnaires

Questions	Main Questionnaire all	Scales used in the supplementary questionnaires of the different groups		
		group 1	group2	group 3
Immigration				
B35-B37	.06 (4 TS)	.09 (5 A/D)	.07 (4 TS)	.08 (7 D/A)
Consequences				
B38-B40	.11 (11 TS)	.14 (5 A/D)	.15 (11 D/A)	.11 (7 D/A)
Satisfaction with work				
E26 –E28	.08 (5 A/D)	.08 (5 A/D)	.10 (11 TS)	.12 (11 D/A)
Satisfaction with contacts				
E40, E43,E45	.06 (5 A/D)	.09 (5 A/D)	.11 (5 A/D)	.10 (7 D/A)

TS = trait specific question A/D = Agree /disagree scale D/A = Disagree/agree scale

The table shows that for the first two topics the trait specific scales have the smallest variation in quality but for the first topic this is a 4 point scale and for the second it is an 11 point scale. For third topic the 5 point A/D scale has the smallest variation and this is also true for the fourth topic.

These results suggest that there is no clear indication in this experiment that one scale is more cross culturally comparable than the other. Certainly the 7 point scale is not a good candidate because this scale has most of the time rather bad quality and a relative high variation in quality across countries.

To illustrate the variation of quality across countries we present in Table 4 the difference in quality for each country across all topics asked in the round 3 MTMM experiments. It should be clear that this list is not an indication of the quality of all questions in the questionnaires in the different countries. We are giving this list only to illustrate the considerable difference in quality across countries.

Table 4 The mean quality across questions for all countries analyzed

Country	quality
Spain	.78
Bulgaria	.72
Germany	.71
UK	.69
Portugal	.68
Estonia	.67
Denmark	.67
Poland	.66
Belgium	.65
Switzerland	.64
France	.61
Slovenia	.60
Slovakia	.59
Russia	.58

In Spain the mean quality is .78 which means that 78% of the observed variance in the observed variables is real variance and only 22% comes from errors. For Russia we get for these specific questions only 58% real variance and 42% error variance.

If the correlation between two latent variables would be the same in these two countries and equal to .6, let us say, then the correlations for the observed variables with these qualities would be .47 in Spain and .35 in Russia. This difference has no substantial meaning but is only due to measurement error. This illustrates once more the importance of correction for measurement error in comparative research. But this requires information about the quality of all different scales used because there is no indication that one of the scales is the most robust in cross cultural comparison.

Appendix 1 Questions from the main questionnaire

The immigration questions

Now some questions about people from other countries coming to live in [country].

B35 CARD 14 Now, using this card, to what extent do you think [country] should² allow people of the same race or ethnic group as most [country's] people to come and live here³?

- | | |
|----------------------------------|---|
| Allow many to come and live here | 1 |
| Allow some | 2 |
| Allow a few | 3 |
| Allow none | 4 |
| (Don't know) | 8 |

B36 STILL CARD 14 How about people of a different race or ethnic group from most [country] people? Still use this card.

- | | |
|----------------------------------|---|
| Allow many to come and live here | 1 |
| Allow some | 2 |
| Allow a few | 3 |
| Allow none | 4 |
| (Don't know) | 8 |

B37 STILL CARD 14 How about people from the poorer countries outside Europe? Use the same card.

- | | |
|----------------------------------|---|
| Allow many to come and live here | 1 |
| Allow some | 2 |
| Allow a few | 3 |
| Allow none | 4 |
| (Don't know) | 8 |

² "Should" in the sense of 'ought to'; not in the sense of 'must'.

³ "Here" = country throughout these questions.

The question about consequences of immigration

B38 CARD 15 Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries? Please use this card.

Bad for the economy											Good for the economy	(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88	

B39 CARD 16 And, using this card, would you say that [country]'s cultural life is generally undermined or enriched by people coming to live here from other countries?

Cultural life undermined											Cultural life enriched	(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88	

B40 CARD 17 Is [country] made a worse or a better place to live by people coming to live here from other countries? Please use this card.

Worse place to live											Better place to live	(Don't know)
00	01	02	03	04	05	06	07	08	09	10	88	

Question on satisfaction with work

CARD 35 Using this card, please tell me to what extent you agree or disagree with each of the following statements. **READ OUT EACH STATEMENT AND CODE IN GRID**

		Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don' know)
E26	I love learning new things.	1	2	3	4	5	8
E27	Most days I feel a sense of accomplishment from what I do.	1	2	3	4	5	8
E28	I like planning and preparing for the future.	1	2	3	4	5	8

Questions on satisfaction with contacts

CARD 39 Using this card, please say to what extent you agree or disagree with each of the following statements.

		Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
E40	I generally feel that what I do in my life is valuable and worthwhile ⁴	1	2	3	4	5	8
E43	There are people in my life who really care about me	1	2	3	4	5	8
E45	I feel close to ⁵ the people in my local area	1	2	3	4	5	8

⁴ "Worthwhile" means 'beneficial to others'.

⁵ "Close to" means 'identify with', 'feel attached to'

Appendix 2 Supplementary questionnaires

Proposal 15.5.06

(Supplementary Questionnaire Version 1)

INTERVIEWER: RECORD SERIAL NUMBER OF RESPONDENT:

INTERVIEWER: ENTER VERSION NUMBER OF QUESTIONNAIRE:

Version 1	1
Version 2	2
Version 3	3

To help us improve our questions in the future, here are just a few final questions which are similar to previous ones. Please treat them as if they were completely new questions.

CARD 1 Using this card, please say how much you agree or disagree with each of these statements.

		Agree strongly				Disagree strongly (Don't know)	
S301.	[Country] should allow more people of the same race or ethnic group as most [country's] people to come and live here	04	03	02	01	00	88
S302.	[Country] should allow more people of different race or ethnic group as most [country's] people to come and live here	04	03	02	01	00	88
S303.	[Country] should allow more people from the poorer countries outside Europe to come and live here	04	03	02	01	00	88
S304.	It is generally bad for [country's] economy that people come to live here from other countries	04	03	02	01	00	88
S305.	[Country's] cultural life is generally undermined by people coming to live here from other countries	04	03	02	01	00	88
S306.	[Country] is made a worse place to live by people coming to live here from other countries	04	03	02	01	00	88

		Agree Strongly				Disagree strongly (Don't know)	
S307.	I love learning new things	04	03	02	01	00	88
S308.	Most days I feel a sense of accomplishment from what I do	04	03	02	01	00	88
S309.	I like planning and preparing for the future	04	03	02	01	00	88
S310	I generally feel that what I do in my life is valuable and worthwhile.	04	03	02	01	00	88
S311	There are people in my life who really care about me.	04	03	02	01	00	88
S312	I feel close to the people in my local area	04	03	02	01	00	88

(Supplementary Questionnaire Version 2)

INTERVIEWER: RECORD SERIAL NUMBER OF RESPONDENT: _____

INTERVIEWER: ENTER VERSION NUMBER OF QUESTIONNAIRE:

Version 1	1
Version 2	2
Version 3	3

To help us improve our questions in the future, here are just a few final questions which are similar to previous ones. Please treat them as if they were completely new questions.

S201 CARD 14 Now, using this card, to what extent do you think [country] should⁶ allow people of the same race or ethnic group as most [country's] people to come and live here⁷?

Allow many to come and live here	1
Allow some	2
Allow a few	3
Allow none	4
(Don't know)	8

S202 STILL CARD 14 How about people of a different race or ethnic group from most [country] people? Still use this card.

Allow many to come and live here	1
Allow some	2
Allow a few	3
Allow none	4
(Don't know)	8

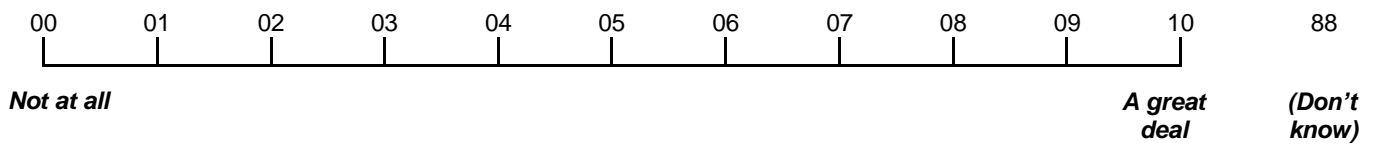
S203 STILL CARD 14 How about people from the poorer countries outside Europe? Use the same card.

Allow many to come and live here	1
Allow some	2
Allow a few	3
Allow none	4
(Don't know)	8

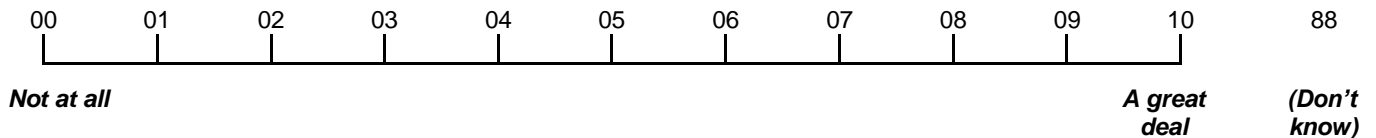
S204. CARD 1 To what extent do you think it is generally bad for [Country] 's economy that people come to live here from other countries?

⁶ "Should" in the sense of 'ought to'; not in the sense of 'must'.

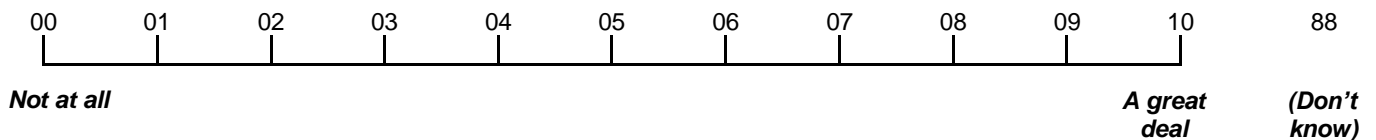
⁷ "Here" = country throughout these questions.



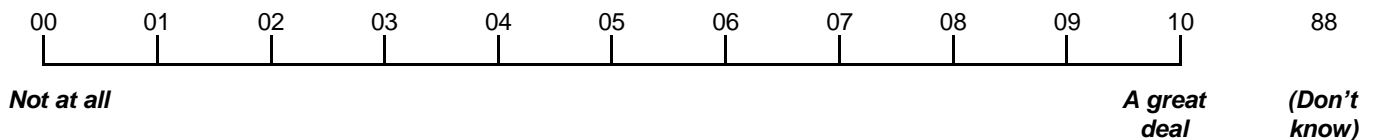
S205. STILL CARD 1 To what extent do you think that [Country] 's cultural life is generally undermined by people coming to live here from other countries?



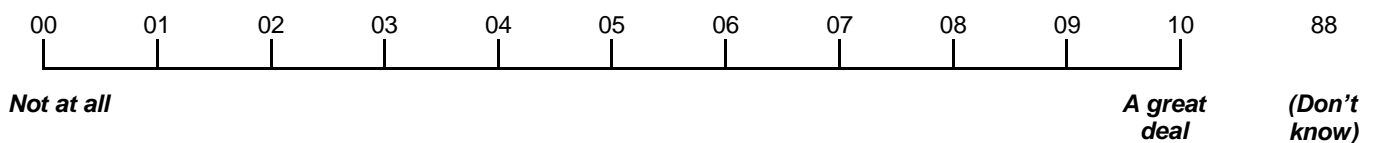
S206. STILL CARD 1 To what extent do you think [Country] is a worse place to live by people coming here from other countries



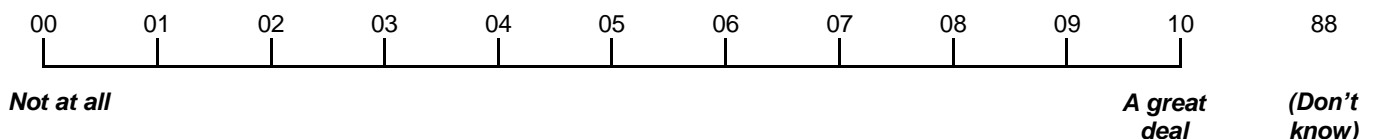
S207. CARD 1 To what extent do you love learning new things?



S208. STILL CARD 1 To what extent do you most days feel a sense of accomplishment from what I do?



S209. STILL CARD 1 To what extent do you like planning and preparing for the future?



CARD 2 Using this card, please say how much you agree or disagree with each of these statements.

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
S210. I generally feel that what I do in my life is valuable and worthwhile.	01	02	03	04	05	08
S211. There are people in my life who really care about me	01	02	03	04	05	08
S212. I feel close to people in my local area	01	02	03	04	05	08