

RECSM Working Paper Number 23

2011

The results of the MTMM experiments¹ in round 2

Willem Saris
Irmtraud Gallhofer

In the second round again 6 MTMM experiments have been done to evaluate the quality of the questions. In this report we discuss in detail the differences in quality of the responses in the different countries. But before we are going to discuss these results we will first indicate the quality criteria that we use.

The quality criteria

In Figure 1 we show the basic response model we are using as our starting point. For details of this approach we refer to our earlier papers and to Saris and Gallhofer (2007).

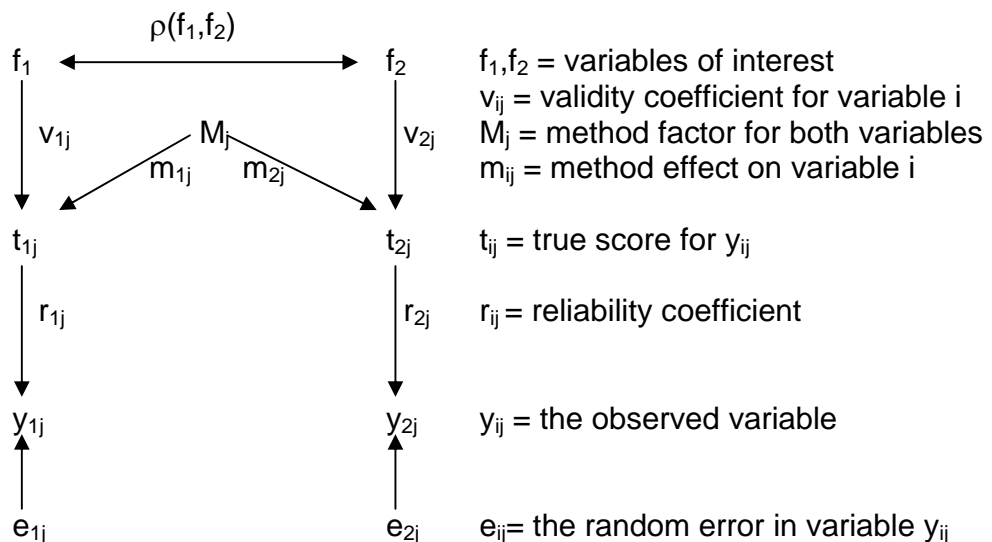


Figure 1 The response model used in the MTMM experiments

The difference between the observed response (y) and the so called “true score” (t) is random measurement error (e). So the coefficient r represents the reliability coefficient and r^2 is the reliability.

The difference between the true score and the concept by intuition (f_i) is systematic effects due to the method (m). So the coefficient v represents the true score validity coefficient and v^2 is the true score validity.

The quality of a measure (q^2) is defined as $q^2 = r^2 \cdot v^2$ and q is the quality coefficient.

¹ We are very grateful for the work that has been done by Irene Jonkers and Joyce Neys in preparation of this analysis

The correlation between the variables of interest is denoted by $\rho(f_1, f_2)$
Several remarks should be made. The first is that the correlation (r_{ij}) between two observed variables is:

$$R_{ij} = q_i \rho(f_1, f_2) q_j \quad (1)$$

A second point is that this means that this correlation between the observed variables can never be larger than the product of the quality coefficients.

A third point is that one can not compare correlation across countries without correction for measurement error if the measurement quality coefficients are very different across countries. This follows directly from the above equation (1).

In this paper we concentrate on the variation in measurement quality across different types of questions and across countries as far as this can be studied on the basis of the MTMM experiments of Round 2 of the ESS.

Experiment 1 How to ask numeric questions

In earlier experiments we have seen that the frequency and amount questions had very bad quality. Therefore, the format of such questions has been tested in the pilot of round 2 and a new version has been included in the main questionnaire. An experiment with alternative forms is done in order to see how these different versions work in the different countries. In Appendix 1 the formulations are presented.

The three questions in the main questionnaire are : G22/ G23 and G24. In the supplementary questionnaire the questions are IS2 / IS3 and IS4 and IS15 / IS 16 and IS17.

In order to get an impression of the quality of the questions we have compared the mean quality of measurement across the three questions and across countries for the three forms. The result was that the mean quality in the main questionnaire was .79. The two alternatives had a mean quality of respectively .81 and .88. These means are very high compared with the quality of other topics. The most successful one was the version where first hours were asked and after that the percentage of work done by each of the partners was presented on a 100 point scale.

It turned out that the categorical questions, used in the main questionnaire, required a considerable amount of work because one has to transform the categories again to numeric scores which is not obvious besides that one creates extra correlation because in this process the same approach has been used for both questions. Given these complications it seems that the direct question with respect to hours would have been better (form 2) while the questions in percentages on a 100 point scale would even have been better (form 3). Given the choice for the categorical approach in the main questionnaire we compare now the quality of the three questions in the different countries. The results have been summarized in Table 1. In the cells the quality of the measures is indicated for the different countries. Although the mean quality is relatively high the quality is not equally high in all countries. In fact the quality is extremely low for all three questions in Sweden and rather low in Finland while the quality is very high in Estonia and Switzerland. In most other countries the quality was rather good, between .8 and .9.

Table 1 The quality (q^2) obtained in the different countries²

country	G22	G23	G24
aus	,7396	,7921	,7396
bel	,8281	,7569	,7396
cze	,7953	,7621	,7453
est	,9216	,8649	,8649
fin	,6561	,4900	,5776
ger	,8649	,8100	,8281
gre	,8464	,8836	,9409
lux	,8100	,8464	,8281
pol	,9025	,9025	,9025
por	,8493	,8143	,7642
slo	,8477	,8477	,7418
sp	,9413	,9413	,8314
swe	,4706	,3931	,4207
swi	,9025	,9025	,9025
Total	,8126	,7862	,7734

One should realize that given the low quality in Finland and Sweden one can not expect high correlations with other variables without correction for measurement error.

Experiment 2 different position of the items on the scale

In survey research very often batteries of statements are used, where within the statements an arbitrary position on the underlying dimension is specified. For example, it is said that something is “usually” done or “seldom” done. This choice is arbitrary but may have consequences for the results. One can also avoid such arbitrary choices and ask people to specify how frequently the activity happens on a scale from never to always. This experiment has been done with items about activities of doctors. The questionnaire is presented in Appendix 2. The questions in the main questionnaire are D25 / D26 and D27 the second and third forms are in the supplementary questionnaire respectively IS5 /IS6 and IS7 and IS28 / IS29 and IS30.

First we make again the comparison between the three forms based on the mean quality. The result was that the mean quality in the main questionnaire was .57 , the mean quality of the second form was .19 and of the third form .41. It will be clear that the second form using a battery with agree/disagree statements using the term “usually” can not be used. The third form, also an agree/disagree battery but with the term “seldom “ is better but the form in the main questionnaire, directly asking to scale the frequency, is even better. However compared with the previous experiment we see that the quality in this case is much lower. So it makes sense to look if that is due to a specific question or specific countries. This can be seen in Table 2.

² Norway is missing although in the end proper estimates for Norway were obtained. These estimates were clearly the lowest of all. So the further analysis does not change by this finding.

Table 2 The quality of the questions in experiment 2 for the different countries³

country	D25	D26	D27
aus	,3844	,7396	,6724
bel	,1849	,7056	,7396
cze	,0625	,8100	,8281
den	,0484	,7396	,7744
est	,4225	,8464	,8281
fin	,0144	,6084	,6400
ger	,3136	,5625	,7056
gre	,3025	,8100	,8100
lux	,0961	,9801	,8281
pol	,1521	,6561	,7744
por	,2601	,9801	,6561
slo	,2916	,6724	,7056
swe	,1936	,7225	,7569
uk	,1849	,8100	,6724
Total	,2080	,7602	,7423

Table 2 shows indeed that question D25 is far below an acceptable quality level while the other two are in average rather good. Question D25 asks how often “Doctors keep the whole truth from their patients” ? It seems that people have no clear opinion about this issue and their answer is thus random or guided by the wording of the question.

There are also countries that score rather low on the other two questions, especially Finland again and this time Germany, while there are again high scores on quality, especially Estonia again and Luxembourg. The difference is more than .2 which is very much and can create a considerable difference between the correlations with other variables.

Experiment 3: Item specific categories or batteries

In this experiment a comparison has been made between a standard battery and a set of separate questions with item specific response categories. The second form had 4 categories like the battery, the third form had 11 categories. The questions concern characteristics of a job. The questions in the main questionnaire are G64 / G65 and G66, in the supplementary questionnaire respectively IS19/IS20 and IS 21 and IS32 / IS33 and IS34. The questions are presented in Appendix 3. As expected the mean quality of the second and third form were far better than the first form using a battery. The form of the main questionnaire had an averaged quality of .61 while the second and third form had a quality of respectively .80 and .79.

A problem in this experiment was that the third item was hardly correlated with the other two variables and due to that in many countries the model could not be estimated (Saris et al. 2004). Therefore, we have only the results for a limited number of countries presented in

³ Norway is missing because one set of questions has not been asked in the supplementary questionnaire

Table 3.

Table 3 The quality of the questions in experiment 3 for the different countries⁴

country	G64	G65	G66
aus	,5752	,5898	,5752
bel	,8836	,8836	,9216
cze	,5929	,4356	,5184
den	,6889	,6724	,7396
fin	,5929	,6241	,4096
ger	,6724	,7396	,6724
lux	,6241	,5184	,4624
slo	,6147	,2122	,5547
sp	,5329	,5329	,5625
swe	,6084	,6241	,5625
Total	,6386	,5833	,5979

For those countries where estimates could be obtained the difference in mean quality between the three questions is not significant. However, it is remarkable that Finland again has a relative bad result now together with the Czech Republic while in Belgium far better results are obtained than in any other country.

Experiment 4: The use of different labels and positions of items

In this experiment we test the effect of the positions of the items on the underlying dimension as in experiment 2 but also the effect of scale with long labels at both sides of the scale. The number of categories was each time the same. The positions of the items were changed by changing the item from positive to negative while in the last form the positive and negative statements were placed at the end points of the scale. The topic was the role of men and women in society. The questions in the main questionnaire were G6 / G7 and G8, in the supplementary questionnaire respectively IS8 / IS9 and IS10 and IS22 / IS23 and IS24. For details we refer to appendix 4.

The analysis of the mean quality shows that all three forms were not very good but the second form where the more stereotype statements of form 1 were negated was clearly the worst with a mean quality of .33. The other two forms had approximately equal quality : the first form had a mean of .56 and the third one a mean of .54 but both were not very good. Looking at Table 4 one can see that especially question G7 was responsible for the negative result. This question had a mean quality of .39 while question G6 had a quality of .63 and question G8 of .68. The weak result was not due to some countries. Table 4 shows that most countries had a very low quality for this item. The only country with a good score for even all three questions is Portugal. The UK score is rather high on the first and third question while in this case Estonia scores rather low and also Norway. Note that Norway was not present so far, because, so far the data from Norway could not produce an acceptable result. In this case there is an acceptable result but the result is very bad.

⁴ Norway is missing because the correlations between the factors was zero so that improper solutions were obtained

Table 4 The quality of the questions in experiment 4 for the different countries

country	G6	G7	G8
aus	,5398	,4058	,5941
bel	,5843	,1019	,6916
cze	,6400	,3600	,6400
den	,8100	,5184	,7056
est	,4489	,7569	,5476
fin	,7569	,1849	,6400
ger	,6561	,5041	,7396
gre	,7225	,6724	,7225
lux	,6301	,3617	,7590
nor	,4174	,1935	,4413
pol	,5625	,2704	,6889
por	,7396	,7056	,8281
slo	,5367	,2544	,6752
sp	,5476	,4225	,6561
swe	,6889	,2704	,6241
swi	,5776	,4225	,6889
uk	,7921	,2025	,9025
Total	,6265	,3887	,6791

Experiment 5: The use of Fixed reference points

In the ESS we usually use what has been called fixed reference points, i.e., labels that have a fixed position on the underlying opinion scale for example “extremely satisfied”. That is definitely the end point of the satisfaction scale. A non fixed reference point could be “very satisfied”. Some people will see it as the end point of the scale others don’t. This difference of perception can cause differences in responses that have nothing to do with the substantive opinions. Therefore, fixed reference points have advantages. This experiment should show if this is indeed the case in all countries and also if 3 fixed reference points are better than 2. The topic for the experiments were the satisfaction questions belonging to the Core of the ESS. In the main questionnaire the standard form was used B25 / B26 and B27. In the supplementary questionnaire these questions are respectively IS11 / IS12 and IS13 and IS35 / IS36 and IS37. The exact formulation can be found in Appendix 5. Looking at the mean quality across the forms we see that the standard form is quite a bit better (.77) than the form without fixed reference points (.70) and the one with three reference points (.70).

Also in this case we ask ourselves the question whether the quality varies much across the questions and across the countries. Table 5 shows what we have found for the three questions in the main questionnaire. We see first of all that question 1 has a considerable lower quality than the other two. This is not due to a few countries because in all countries the quality of the first question is lower than of the other two questions. In this case the quality is the lowest in Spain and the Czech Republic and the highest in Estonia and Portugal. One should note that the differences are very big , more than .2. This will give

considerable differences in correlations with other variables.

Table 5 The quality of the questions in experiment 5 for the different countries

country	B25	B26	B27
aus	,7921	,8836	,8649
bel	,7056	,8464	,8649
cze	,5997	,6555	,6561
den	,6889	,9025	,8100
est	,7921	,9025	,8649
fin	,6724	,8281	,8281
ger	,5219	,7792	,8129
gre	,7632	,7964	,8138
lux	,7225	,7921	,9801
nor	,7396	,9801	,7569
pol	,7569	,9025	,8836
por	,8100	,8464	,8281
slo	,5858	,7162	,6416
sp	,5675	,6688	,6688
swe	,6235	,7474	,6521
swi	,7396	,8100	,9409
uk	,7744	,8836	,8100
Total	,6974	,8201	,8046

Experiment 6: The effect of repetition for different items

These MTMM experiments are not possible without repeated observations. In our SB-MTMM design the number of repetitions has been reduced to 1 for all respondents but this can nevertheless have a positive (memory) or a negative (inaccurateness) effect on the quality of the data. This can be seen in an experiment, where exactly the same questions are repeated in the different parts of the data collection. This experiment is done with questions with respect to “trust in political institutions”. The standard questions are B4/B5 and B6 while the same questions are repeated as IS25 / IS 26 and IS27 and for the other group in IS38 / IS39 and IS40. Looking at the mean quality we see that the quality goes down by repetition because the questions in the supplementary questionnaire have a bit less quality, resp .78 and .75 than the questions in the main questionnaire (.82).

Also in this experiment we look at differences in quality between questions and between countries. This can be done on the basis of Table 6. This table shows that the quality of the questions is not much different, but, between the countries there are quite large differences in quality of measurement. In this case Belgium got relatively low scores while Estonia and Portugal are at the top and the differences are again more than .2 in quality which will have a large effect on the observed correlations.

Table 6 The quality of the questions in experiment 6 for the different countries⁵

country	d11qal	d12qal	d13qal
aus	,7971	,7971	,7632
bel	,6349	,7621	,6831
cze	,9025	,8836	,8281
den	,7607	,8129	,7953
est	,9025	,9216	,8649
ger	,7621	,7621	,7792
gre	,9025	,9025	,8836
lux	,9025	,8836	,8464
pol	,9025	,8836	,8649
por	,9216	,9409	,8649
slo	,8281	,7921	,8281
sp	,7569	,8649	,8281
swe	,6559	,6413	,7321
swi	,8100	,8836	,8464
uk	,8100	,7225	,8100
Total	,8167	,8303	,8145

Comparison of the methods

Finally, we are going to make a comparison of the different methods that have been used in the MTMM experiments. Table 7 presents the results with respect to the different methods used in the different experiments. The questions 1-3 were asked in the main questionnaire and used method 1, questions 4-6 represent the same questions but now asked using method 2 in the supplementary questionnaire and questions 7-9 are the same questions but now asked using method 3 in the supplementary questionnaire.

In experiment 1 method 3 provided the best results. In that form the amount of time for home work was asked directly and after that the proportions of the work done by the two partners was asked. The first question was in all forms the same but in method 1 a category scale was used for the second and third question which worked less good and also the direct questions in hours (second method) were less good.

In experiment 2 method 1 was clearly the best. In this form no frequency of activities was specified while in the other two forms that was done in opposite ways. It seems that one can better ask the people to provide the frequency than specifying a frequency in the statement and asking people to agree or disagree with the statement.

In experiment 3 a battery of agree/disagree statements was used as the first method while in the other two methods item specific scales were used with different number of categories. This experiment shows once again the big advantage in quality of item specific questions above batteries of agree disagree items.

⁵ Norway is missing because one questions has not been asked in the supplementary questionnaire

Table 7 The comparison of the different method with respect to quality

Report				
experiment		m1to3qal	m4to6qal	m7to9qal
1	Mean	,7907	,8120	,8769
	N	14	14	14
	Std. Deviation	,13609	,21162	,11852
2	Mean	,5702	,1918	,4058
	N	14	14	14
	Std. Deviation	,06797	,12836	,10526
3	Mean	,6066	,8001	,7944
	N	10	10	10
	Std. Deviation	,12658	,13018	,11330
4	Mean	,5648	,3248	,5363
	N	17	17	17
	Std. Deviation	,09755	,15238	,14082
5	Mean	,7740	,7039	,7025
	N	17	16	17
	Std. Deviation	,07990	,10278	,14190
6	Mean	,8205	,7796	,7501
	N	15	15	15
	Std. Deviation	,07236	,06719	,09717
Total	Mean	,6918	,5876	,6691
	N	87	86	87
	Std. Deviation	,14576	,28358	,19777

In experiment 4 the often used statements (method 1) or the bipolar scales with long labels (methods 3) give much better quality than the items in method 2 which were the opposite of the items of method 1.

In experiment 5 method 1 using 2 fixed reference points is working better than method 3 with three fixed reference points or method 2 with only one fixed reference point.

The last experiment was done to see what the effect was of the repetition of the questions in the supplementary questionnaire. This experiment shows that the quality goes down a little bit although we did not check if this was a consequence of a change in data collection mode (face to face, interviewer administered to self completion).

Conclusion and discussion

In the next two figures we give a summary of the results we have seen above. Figure 2 shows the mean quality of the questions in the main questionnaire across countries for each experiment.

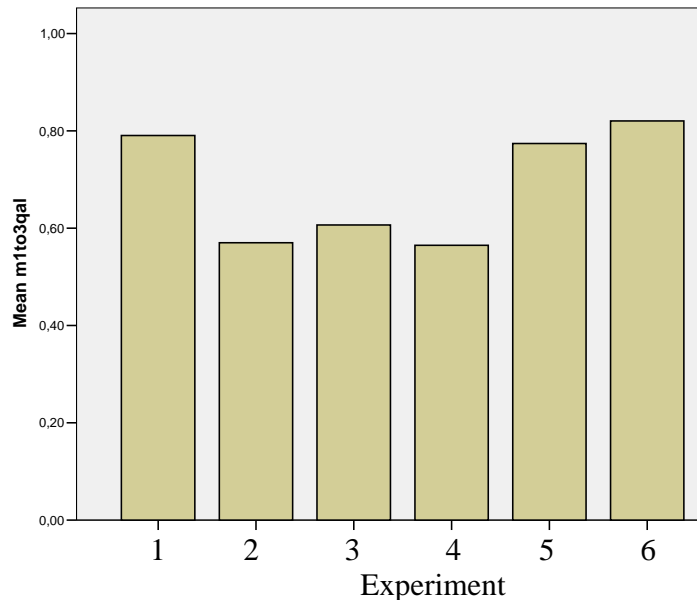


Figure 2 The quality of the questions in the different experiments

The figure shows very clearly that the data in experiment 1, 5 and 6 are far better than those in experiment 2, 3 and 4. This is also what we have seen above.

- The quality in experiment 2 is mainly less good because of the first questions which people seem to have difficulty with.
- The reason for the low quality in experiment 3 is not due to a single question; all questions have a low quality.
- Finally the low quality in the fourth experiment is due to the second question which has a much lower quality than the other two.

Figure 3 shows the mean quality of the questions in the main questionnaire across the experiments for the different countries.

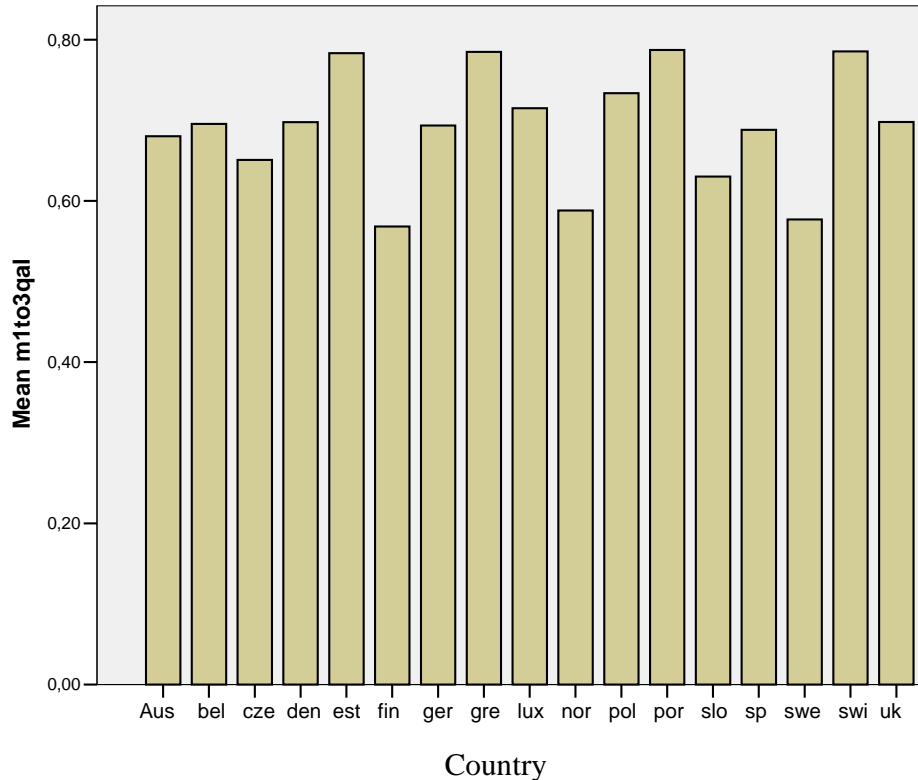


Figure 3 The quality of the questions in the different countries

A remarkable phenomenon one can see in this figure is that the Scandinavian countries have the lowest quality of all while the highest quality has been obtained in Estonia, Greece, Portugal and Switzerland.

The difference is also quite considerable. The highest mean quality is .79 in Portugal while the lowest is .57 in Finland. If the correlation between the constructs of interest is .6 in both countries and the measures for these variables have the above quality then the observed correlation in Portugal would be .474 while the observed correlation in Finland would be .342. Normally people would say that this is a large difference in correlations which requires a substantive explanation. However this difference can be expected on the basis of differences in data quality and has no substantive meaning at all.

This shows that it is absolutely necessary to correct for data quality before one compares correlations across countries. These corrections can be done using the information presented in the tables provided above. Let us give an example. We look at the observed correlation between item 2 and 3 for the second experiment. In Estonia the correlation between these two variables is .519 and in Finland .328. This is quite a difference. But before we decide that this is substantively relevant we should correct for measurement errors because, as we have seen, the quality is much better in Estonia than in Finland. Table 2 shows for these variables in Estonia respectively a quality of .846 and .828 while the quality in Finland is respectively .608 and .640. The correction for measurement

error can be done by dividing the observed correlation by the product of square root of the quality estimates for the two variables⁶ or

$$\rho_{23} = r_{23} / \sqrt{q_2 q_3} \quad (2)$$

This follows directly from equation (1). Using this formula we get for Estonia a correlation corrected for measurement error of .62 and for Finland .53. These two correlations are both larger than the observed correlations, but the correction also made the difference between the two much smaller. This is due to the fact that the quality of the measures was much lower in Finland than in Estonia. Therefore the correction for the Finnish correlation was bigger.

This example shows how the obtained quality information can be used directly to correct for measurement error. Why the quality varies so much for different countries remains an open question. It can have to do with the translation, the formulation of the questions but also with the amount of interviews in the countries or cultural differences. This point requires further research.

⁶ Note that the quality estimates in the tables are q^2 . Therefore one has to get the square root of the quality estimates which are equal to the quality coefficients (q).

Appendix 1 Measurement of time spend on home work

The Form in the main questionnaire

G22 CARD 62 I'd now like to talk about housework, as described in this list. By housework, we mean anything done around the home, such as cooking, washing, cleaning, care of clothes, shopping, maintenance of property, but not including childcare and leisure activities. Thinking about the total amount of time people spend on housework for your home, about how many hours are spent **on a typical weekday?**

WRITE IN:

(Don't know) 88

G23 CARD 63 And about how much of this time do you spend yourself? Please use this card.

None or almost none	01
Up to a quarter of the time	02
More than a quarter, up to a half of the time	03
More than a half, up to three quarters of the time	04
More than three quarters, less than all of the time	05
All or nearly all of the time	06
(Don't know)	88

G24 STILL CARD 63 And about how much of this time does your husband/wife/partner spend?
Please use this card.

None or almost none	01
Up to a quarter of the time	02
More than a quarter, up to a half of the time	03
More than a half, up to three quarters of the time	04
More than three quarters, less than all of the time	05
All or nearly all of the time	06
(Don't know)	88

Second form

Only to be used if the respondent has a partner

I'd now like to talk about housework. By housework, we mean anything done around the home, such as cooking, washing, cleaning, care of clothes, shopping, maintenance of property, but not including childcare and leisure activities. Thinking about the total amount of time people spend on housework for your home, about how many hours are spent **on a typical weekday**?

WRITE IN:

(Don't know) 88

25. And about how many of these hours do you spend yourself?

WRITE IN:

(Don't know) 88

26. And about how many does your husband/wife/partner spend?

WRITE IN:

(Don't know) 88

Third Form

Only to be asked if the respondent has a partner

I'd now like to talk about housework. By housework, we mean anything done around the home, such as cooking, washing, cleaning, care of clothes, shopping, maintenance of property, but not including childcare and leisure activities. Thinking about the total amount of time people spend on housework for your home, about how many hours are spent **on a typical weekday**?

WRITE IN:

(Don't know) 88

And about what proportion of this time do you spend yourself?
Chose a number between 0 and 100 indicating the proportion you do

And what proportion of this time does your husband/wife/partner spend?
Chose a number between 0 and 100 indicating the proportion you do

Appendix 2 The social distance between the doctor and patients
Form in the main questionnaire

CARD 30 Using this card, please indicate how often you think the following applies to doctors in general:

	Never or almost never	Some of the time	About half of the time	Most of the time	Always of almost always	(Don't know)
D25 Doctors keep the whole truth from their patients?	1	2	3	4	5	8
D26 GPs treat their patients as their equals?	1	2	3	4	5	8
D27 Before doctors decide on a treatment, they discuss it with their patient?	1	2	3	4	5	8

The Second form

Please say how much you agree or disagree with each of these statements:

CARD: *Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree*

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
...doctors usually keep the whole truth from their patients?	1	2	3	4	5	8
...GPs usually treat their patients as their equals?	1	2	3	4	5	8
Before doctors decide on a treatment, they usually discuss it with their patient	1	2	3	4	5	8

Third Form

Please say how much you agree or disagree with each of these statements:

CARD: *Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree*

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
...doctors seldom keep the whole truth from their patients?	1	2	3	4	5	8
...GPs seldom treat their patients as their equals?	1	2	3	4	5	8
Before doctors decide on a treatment, they seldom discuss it with their patient	1	2	3	4	5	8

Appendix 3 Question about work
Form in the main questionnaire

CARD 70 Using this card, please tell me how much you agree or disagree with each of the following statements 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
G65 My job is secure	1	2	3	4	8
G66 My health or safety is at risk because of my work	1	2	3	4	8

Second Form (Only if they have a job)

Please tell me what you think of your current job.

Is your work (Read out)

- 1 not at all varied
- 2 a little varied
- 3 quite varied
- 4 or very varied ?

Is your job

- 1 not at all secure
- 2 a little secure
- 3 quite secure
- 4 or very secure

Does your work put your health or safety

- 1 not at all at risk
- 2 a little at risk
- 3 quite a lot at risk
- 4 or very much at risk ?

Third Form (Only if they have a job)

Now we would like to ask what you think of your current job.

Please indicate by a number between 0 and 10 how varied your work is.

0 means not at all varied

10 means very varied

Please indicate by a number between 0 and 10 how secure your work is.
0 means not at all secure
10 means very secure

Please indicate by a number between 0 and 10 how your health or safety are at risk by your work .
0 means not at all at risk
10 means very much at risk

Appendix 4 The role of men and Women in society

Form in the main questionnaire

CARD Using this card, please tell me how much you agree or disagree with the following statements. Firstly...

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
PG6. A woman should be prepared to cut down on her paid work for the sake of her family.	1	2	3	4	5	8
PG7. Men should take as much responsibility as women for the home and children	1	2	3	4	5	8
PG8. When jobs are scarce, men should have more right to a job than women	1	2	3	4	5	8

Second Form

CARD Using this card, please tell me how much you agree or disagree with the following statements. Firstly...

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
PG6. A woman should not have to cut down on her paid work for the sake of her family.	1	2	3	4	5	8
PG7. Women should take more responsibility for the home and children	1	2	3	4	5	8
PG8. When jobs are scarce, women should have the same right to a job as men	1	2	3	4	5	8

Third Form

If you have to chose between the two following two opinions what position would you prefer ?
Chose a number between 1 and 5 that is closest to your opinion.

A woman should be prepared to cut down on her paid work for the sake of her family.	1	2	3	4	5	A woman should not have to cut down on her paid work for the sake of her family
---	---	---	---	---	---	---

Please do the same for the following two opinions ?
Chose a number between 1 and 5.

Men should take as much responsibility as women for the home and children	1	2	3	4	5	Women should take more responsibility for the home and children
---	---	---	---	---	---	---

Please do the same for the following two opinions ?
Chose a number between 1 and 5.

When jobs are scarce, men have more right to a job than women	1	2	3	4	5	When jobs are scarce, women should have the same right to a job as men
---	---	---	---	---	---	--

Appendix 5 Satisfaction questions
Form in the main questionnaire

B31 STILL CARD 13: On the whole how satisfied are you with the present state of the economy in [country]? Still use this card.

Extremely Dissatisfied														Extremely satisfied	(Don't know)
	00	01	02	03	04	05	06	07	08	09	10	88			

B32 STILL CARD 13 Now thinking about the [country] government, how satisfied are you with the way it is doing its job? Still use this card.

Extremely Dissatisfied														Extremely satisfied	(don't Know)
	00	01	02	03	04	05	06	07	08	09	10	88			

B33STILL CARD 13 And on the whole, how satisfied are you with the way democracy works in [country]? Still use this card.

Extremely Dissatisfied														Extremely satisfied	(don't Know)
	00	01	02	03	04	05	06	07	08	09	10	88			

Second Form

B31 STILL CARD 13: On the whole how satisfied are you with the present state of the economy in [country]? Still use this card.

Very Dissatisfied												Very satisfied
	00	01	02	03	04	05	06	07	08	09	10	

B32 STILL CARD 13 Now thinking about the [country] government, how satisfied are you with the way it is doing its job? Still use this card.

**Very
Dissatisfied**

**Very
satisfied**

00 01 02 03 04 05 06 07 08 09 10

B33STILL CARD 13 And on the whole, how satisfied are you with the way democracy works in [country]? Still use this card.

**Very
Dissatisfied**

**Very
satisfied**

00 01 02 03 04 05 06 07 08 09 10

Third Form

B31 STILL CARD 13: On the whole how satisfied are you with the present state of the economy in [country]? Still use this card.

**Extremely
Dissatisfied**

**neither
satisfied
nor
dissatisfied**

**Extremely
satisfied**

00 01 02 03 04 05 06 07 08 09 10

B32 STILL CARD 13 Now thinking about the [country] government, how satisfied are you with the way it is doing its job? Still use this card.

**Extremely
Dissatisfied**

**neither
satisfied
nor
dissatisfied**

**Extremely
satisfied**

00 01 02 03 04 05 06 07 08 09 10

B33STILL CARD 13 And on the whole, how satisfied are you with the way democracy works in [country]? Still use this card.

**Extremely
Dissatisfied**

**neither
satisfied
nor
dissatisfied**

**Extremely
satisfied**

00 01 02 03 04 05 06 07 08 09 10

Appendix 6 Political Trust
Form in the main questionnaire

Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out.
 0 means you do not trust an institution at all, and
 10 means you have complete trust. Firstly...

		No trust at all										Complete trust (Don't know)	
PB2	... [country]'s parliament?	00	01	02	03	04	05	06	07	08	09	10	88
PB3	... the legal system?	00	01	02	03	04	05	06	07	08	09	10	88
PB4.	... politicians?	00	01	02	03	04	05	06	07	08	09	10	88

Appendix 7 Differences in quality across countries

country	Mean	N	Std. Deviation
aus	,6853	6	,18070
bel	,6284	6	,17145
cze	,6173	6	,20204
den	,6093	5	,18700
est	,6803	5	,21679
fin	,5663	5	,18799
ger	,6873	6	,20511
gre	,7539	5	,21963
lux	,6460	6	,22120
nor	,5004	2	,16866
pol	,6349	5	,28311
por	,7528	5	,13231
slo	,6343	6	,16840
sp	,6814	5	,11475
swe	,5776	6	,18892
swi	,7281	3	,17825
uk	,6013	4	,21472
Total	,6493	86	,18524

Appendix 8 Difference in quality compared with Germany

country	Mean	N	Std. Deviation
aus	-,0020	6	,18070
bel	-,0589	6	,17145
cze	-,0700	6	,20204
den	-,0780	5	,18700
est	-,0070	5	,21679
fin	-,1210	5	,18799
ger	,0000	6	,20511
gre	,0666	5	,21963
lux	-,0413	6	,22120
nor	-,1869	2	,16866
pol	-,0524	5	,28311
por	,0655	5	,13231
slo	-,0530	6	,16840
sp	-,0059	5	,11475
swe	-,1097	6	,18892
swi	,0408	3	,17825
uk	-,0860	4	,21472
Total	-,0380	86	,18524