ESS-DACE

The European Social Survey - Data for a Changing Europe

Contract Number: 262208

Deliverable 4.6

Evaluation of the experiments in the supplementary questionnaire of the Round 5 of the ESS

Start date of project: July 2010 Duration: 48 months

Organization name of lead contractor for this deliverable: UPF

Dissemination level: PU (PUBLIC)

Due date of deliverable: June 2012

Actual submission date: September 2012

Evaluation of the experiments in the supplementary questionnaire of the Round 5 of the ESS

Melanie Revilla

Willem Saris

RECSM, UPF / Barcelona

Introduction

In Round 5 three different issues have been evaluated in the supplementary questionnaire:

1. Measurement of media use including the use of the new media
2. Three MTMM experiments
3. A new approach to measure Internal and External Political efficacy

These three topics will be discussed in sequence in this report.

1. Media use

We have mentioned in an earlier report that we thought that longer scales are necessary and that the Internet has to be taken into account. In group 1 and group 3 we specify alternatives for the present form of the questions about media use. The version in group 3 is a bit longer because we want to have an estimate of how much time is used for the TV, Radio, newspapers and other activities on the Internet. This can help in later decisions about the measurement of media use. For the questions we refer to the appendix. Here we will evaluate the results of these questions.

1.1 Impact of specifying that use of Internet to watch television, listen to the radio or read the news should be included

The impact of specifying explicitly "please include any time spent watching TV (listening radio, reading newspapers) using the internet" can be studied by comparing the answers of the respondents in the main questionnaire and in the supplementary questionnaire for respondents getting the first version of the supplementary questionnaire (i.e. "group 1").

Focusing only on respondents of group 1, we have the same respondents answering the two forms of the question (with and without the sentence "please include…internet" mentioned just before). This allows us to look at cross-tables and correlations between a question of the main questionnaire and the repetition in the supplementary questionnaire. Also, potential differences due to sampling are avoided in this case such that differences in answers should really be caused by the extra sentence in the second form.

First, we look at the distributions of the media variables in the main and supplementary questionnaire. By specifying that Internet should be included, we expect people to tell a longer time or the same (if they do not use Internet or already included it when it was not explicit).

So we expect the categories at the beginning of the scale to have a lower frequency in the supplementary questionnaire than in the main. On the other side, we expect the categories at the end of the scale to have a higher frequency in the supplementary than in the main.

The results are presented in Table 1.

Table 1: distributions of the media variables in the main questionnaire and the supplementary questionnaire fro group 1

Group 1	main	suppl	main	suppl	main	suppl
	TVTOT	testd1	RDTOT	testd2	NWSPTOT	testd3
No time	**674**	**1275**	**4738**	**4681**	**5773**	**5470**
<,5	**837**	**885**	**2361**	**2689**	**4785**	**4224**
,5-1	2094	1969	2480	2689	3942	3952
1-1,5	2159	2011	1234	1212	1118	1356
1,5-2	2665	2536	1055	930	492	640
2-2,5	2041	2000	621	568	174	233
2,5-3	**1987**	**1898**	**595**	**608**	**80**	**135**
>3	**4001**	**3718**	**3327**	**2772**	**96**	**176**
DK	29	47	70	106	26	77
No answer	8	131	14	220	9	211
NA		8		8		8
Refusal		17		12		13
Total	16495	16495	16495	16495	16495	16495

Table 1 shows that for the newspaper, the distributions are as expected (e.g. numbers in blue). On the contrary, for television, it is not at all as expected (e.g. numbers in red). For radio, it is in the middle.

But since we have the same respondents answering twice, it is more informative to look at the cross-table to see if really the respondents keep the same categories or switch from a lower to a higher or equivalent category in the scale (what we would expect by specifying explicitly to include the time using internet), or if something different is happening.

Therefore, we secondly look at the cross-table between the response in the main questionnaire and the one in the supplementary questionnaire for respondents of group 1.

Since the variable in the supplementary is the horizontal one, we expect to have high number on the diagonal or in the upper triangular matrix, and 0 (or very low numbers) below the diagonal. The results are presented in Table 2, Table 3 and Table 4.

Table 2: cross-table for the time spend watching television

```
tab tvtot testd1 if spltadmd==1 | spltadmd==4
```

TV watching, total time on average weekday	Tv watching, total time on average weekday						
	No time a	Less than	½ hour to	More than	More than	More than	More than
No time at all	556	47	21	10	8	6	4
Less than 0,5 hour	64	576	123	31	12	8	3
0,5 hour to 1 hour	92	148	1,374	279	116	25	13
More than 1 hour, up	69	35	274	1,233	351	98	39
More than 1,5 hours,	116	35	82	304	1,610	351	82
More than 2 hours, up	84	13	30	74	279	1,169	282
More than 2,5 hours,	90	11	22	30	82	262	1,187
More than 3 hours	199	20	39	50	75	77	287
Don't know	2	0	3	0	2	4	1
No answer	3	0	1	0	1	0	0
Total	1,275	885	1,969	2,011	2,536	2,000	1,898

TV watching, total time on average weekday	Tv watching, total time on average weekday					
	More than	Not appli	Refusal	Don't kno	No answer	Total
No time at all	9	0	1	4	8	674
Less than 0,5 hour	10	1	0	3	6	837
0,5 hour to 1 hour	21	0	0	6	20	2,094
More than 1 hour, up	37	1	2	5	15	2,159
More than 1,5 hours,	52	1	5	4	23	2,665
More than 2 hours, up	92	1	1	2	14	2,041
More than 2,5 hours,	283	1	3	5	11	1,987
More than 3 hours	3,208	3	5	6	32	4,001
Don't know	5	0	0	12	0	29
No answer	1	0	0	0	2	8
Total	3,718	8	17	47	131	16,495

Looking at Table 2 and the results for the television, we clearly see that we have some quite high numbers below the diagonal too, on the contrary of what we expected: for instance, 199 respondents said in the main questionnaire that they were watching TV more than 3 hours a day AND said they were not watching at all television when the question was repeated with the specification "Please include any time spent watching TV using the Internet".

The same pattern is found with radio, as shown in Table 3.

Table 3: cross-table for the time spend listening to the radio

```
. tab rdtot testd2 if spltadmd==1 | spltadmd==4
```

Radio listening, total time on average weekday	Radio listening, total time on average weekday						
	No time a	Less than	½ hour to	More than	More than	More than	More than
No time at all	3,893	369	214	73	39	14	8
Less than 0,5 hour	220	1,633	354	55	19	5	6
0,5 hour to 1 hour	160	374	1,500	257	81	14	22
More than 1 hour, up	69	97	284	552	117	46	16
More than 1,5 hours,	62	67	128	138	473	102	40
More than 2 hours, up	38	31	44	41	106	258	54
More than 2,5 hours,	39	23	29	22	29	65	289
More than 3 hours	181	87	129	72	65	61	173
Don't know	12	8	6	2	1	3	0
No answer	7	0	1	0	0	0	0
Total	4,681	2,689	2,689	1,212	930	568	608

Radio listening, total time on average weekday	Radio listening, total time on average weekday					
	More than	Not appli	Refusal	Don't kno	No answer	Total
No time at all	32	5	4	28	59	4,738
Less than 0,5 hour	18	1	3	7	40	2,361
0,5 hour to 1 hour	34	1	2	10	25	2,480
More than 1 hour, up	30	0	0	7	16	1,234
More than 1,5 hours,	29	0	1	6	9	1,055
More than 2 hours, up	42	0	0	3	4	621
More than 2,5 hours,	91	0	1	2	5	595
More than 3 hours	2,491	1	1	11	55	3,327
Don't know	4	0	0	32	2	70
No answer	1	0	0	0	5	14
Total	2,772	8	12	106	220	16,495

How can we explain this very counterintuitive finding? One potential idea to explain it is that these respondents did not understand the question of the supplementary questionnaire well and answered this second question thinking that ONLY the time spent watching TV using the Internet was asked, EXCLUDING any other time watching TV in a traditional way (which in fact was question 32 asked to respondents in split-ballot group 3).

If it will be decided to use a question with an extra specification about the use Internet, it would be important to be very careful about the formulation in the different languages and be sure that no confusion will occur as shown above. Indeed, the results here suggests that the formulation of the question specifying to include also the time using internet to watch TV or listen to the radio such as used in round 5 is not clear and lead some respondents to misunderstanding.

On the other hand, for a majority of respondents, the expected pattern is found: much more respondents are on or above the diagonal than below. But still, a non negligible part is below the diagonal for TV and radio.

For newspapers, the picture looks more as expected, as can be seen in Table 4.

Table 4: cross-table for the time spend reading the newspapers

```
. tab nwsptot testd3 if spltadmd==1 | spltadmd==4
```

Newspaper reading, total time on average weekday	Newspaper reading, total time on average weekday						
	No time a	Less than	½ hour to	More than	More than	More than	More than
No time at all	4,673	530	256	98	51	13	15
Less than 0,5 hour	414	3,219	873	120	38	14	4
0,5 hour to 1 hour	233	398	2,529	504	128	36	17
More than 1 hour, up	85	39	221	538	151	36	12
More than 1,5 hours,	32	18	50	74	226	57	19
More than 2 hours, up	10	5	14	12	32	60	26
More than 2,5 hours,	7	2	3	6	5	15	31
More than 3 hours	5	3	3	4	9	2	10
Don't know	8	7	2	0	0	0	1
No answer	3	3	1	0	0	0	0
Total	5,470	4,224	3,952	1,356	640	233	135

Newspaper reading, total time on average weekday	Newspaper reading, total time on average weekday					
	More than	Not appli	Refusal	Don't kno	No answer	Total
No time at all	31	4	1	35	66	5,773
Less than 0,5 hour	16	1	6	15	65	4,785
0,5 hour to 1 hour	33	3	3	12	46	3,942
More than 1 hour, up	13	0	1	5	17	1,118
More than 1,5 hours,	9	0	1	0	6	492
More than 2 hours, up	10	0	1	1	3	174
More than 2,5 hours,	9	0	0	0	2	80
More than 3 hours	55	0	0	2	3	96
Don't know	0	0	0	7	1	26
No answer	0	0	0	0	2	9
Total	176	8	13	77	211	16,495

This is also what was found looking at the simple distributions. More research would be needed to explain why, but one guess is that this is due to the fact that most people do not read at all newspaper and that the ones that are reading them are probably more educated and so are less susceptible to misunderstand the question in its second form.

Finally, we also compare the correlations (excluding all missing values) between the answers in the main and in the supplementary questionnaire.

For television, this correlation is **0.7864**. For Radio, it is **0.8338** and for newspapers **0.7148**. These correlations are quite high, knowing that we also do not expect people to pick up exactly the same category depending on how much time they are using Internet for the different purposes asked. They are quite similar also for the three media.

Conclusion:

So overall, it seems that by specifying explicitly that the use of Internet should be included, we get for one part of the respondents an increase in the time reported, which suggest that these respondents did not "naturally" thought about including the time spent on the different media via Internet. However, the difference is not so large, and for a majority of respondents, there is no change in answer between the 2 forms, which can have different meanings:
- the respondents already included the time via internet in the first question
- the respondents are not using Internet for these different activities (this will be checked later, cf. section 1.3)
- the respondents are using internet but very little and since the answer categories are relatively large, at least for newspaper, adding the time spend via internet is not enough to make them switch from one category to the next.
Some respondents then switch to a lower category: this switch may be a result of a misunderstanding of the second form of the question.

1.2 Impact of having more scale categories and different cuts

The impact of having more categories and different cuts can be studied by comparing the answers of respondents in group 1 (first split-ballot group for the supplementary questionnaire) and group 3 (third split-ballot group for the supplementary questionnaire). It is important to notice that now the different forms of the questions are

proposed to different samples: this is not anymore a design with repeated questions for the same respondents, so it is not possible to do a cross-table here, neither to look at the correlation as we did in the section 1.1.

It is on the contrary a classic split-ballot experiment. We have two different samples. Since these two samples are drawn randomly, we do not expect systematic differences between them, but still some differences can appear just by chance. Therefore, the strength of the results is a bit more limited than in the previous analyses. We should also mention that, in both groups, the question specifies that they should include the time spend using the three media via Internet.

We have to make a distinction between television and radio on the one hand, and newspapers on the other hand. People used to spend very different kinds of amounts of time on these different medias and so the idea is that maybe the same scale is not equally good for all three, but that one scale should be used for television and radio, where people can spend really a lot more hours than reading the newspaper, where people usually spend a short amount of time or no time at all. Therefore in group 3, we tested a scale with more categories for the three media, but for the two first (TV, radio) the scale was increased by adding more categories at the end (till "more than 6 hours"), whereas for newspapers, the scale was increased by cutting down the size of the first categories of the scale (e.g. the first one is "less than 15 minutes" instead of "less than 30 minutes").

In order to compare easily the distributions of the variables in the two split-ballot groups, we regroup the categories in the form of group 3 such that they would correspond to the categories of the question in group 1.

For television and radio, what we expect is that when the highest category of the scale is "more than 3 hours", the respondents tend to think that if this is the extreme of the scale, this is a lot of time. This gives them a reference point which is lower than when the highest category is "more than 6 hours" and this leads them to select a lower total time. So we propose the following hypothesis (see Schwarz and Hippler):
When the scale is longer and the higher category is "more than 6 hours" instead of "more than 3 hours", people will tend to say that they use the media more time.

So when comparing the distributions, we expect higher frequencies in the upper part of the scale and in particular in the one "more than 3 hours" (after having grouped all the categories of the second method that are higher than 3 hours).

For newspaper, by cutting the first categories into smaller intervals, we expect higher frequencies at the beginning of the scale, once again because the scale suggests a different reference point of what is "normal" or "little" time for this kind of activity. Table 5 gives the results.

For television, as expected, when the scale is longer and go till more than 6h, the number of respondents telling they watch more than 2,5 hour is increased (numbers in blue). For radio also the results go in the expected direction: more people are telling in group 3 that they are listening radio 1.5 hour or more on an average day.

But also more people are choosing the "no time" category. This is a bit surprising. It may be in part due to random sampling fluctuations (since we do not have the same respondents here) but the difference seems too large to be due just to chance. Another idea is that it can be related to the length of the scale: seeing so many categories, some respondents get "afraid" and decide to satisfice by picking up the first category instead of thinking really about the question.

Table 5: distribution for the media variables using both scales (grouping categories higher than 3 hours in the second scale)

	tv testd1 gp1	tv regroup26 gp3	radio testd2 gp1	radio regroup28 gp3	nwsp testd3 gp1	nwsp regroup30 gp3
no time	1275	1134	4682	**5032**	5474	**5616**
<,5	887	754	2691	2167	4227	**4674**
,5-1	1973	1722	2695	2372	3955	3546
1-1,5	2014	1930	1214	1147	1359	1142
1,5-2	2537	2465	930	**1047**	640	552
2-2,5	2000	1886	568	**625**	233	256
2,5-3	1898	**2221**	609	**724**	135	96
>3	3721	**3946**	2774	**2899**	176	122
DK	47	74	106	104	77	94
no answer	4831	4792	4920	4804	4911	4825
na	29584	29844	29584	29844	29585	29844
refusal	19	18	13	21	14	19
total	50786	50786	50786	50786	50786	50786

For newspapers, the scale is different as mentioned before. Now what we expect is to have more responses at the beginning of the scale by splitting up the first categories in smallest ones. This is indeed what is happening.

In order to see if all these differences observed are significant, we did a chi-square test for equality of distributions. Our null hypothesis H0 is that all samples have the same frequency distributions (we consider the variables for group 3 once the categories have been grouped). The chi-square values for the 3 media (TV, radio and newspaper) is larger (72, 112, 87) than what can be expected if H0 holds for the number of degree of freedom we have (7). So we have to reject H0: the differences in distributions are significant across group 1 and 3 for the 3 media.

Conclusion:

The expected result is found: by adding categories at the end of the scale for television and radio, we get more respondents telling they are using these media for a higher time, and on the contrary by cutting down the first categories for newspapers, we get more respondents at the beginning of the scale. We think that the distributions when using the second form (longer one) is closer to the true one and that it should be preferred. Using a different scale for television and radio on a one hand and newspaper on the other hand also seems to make sense since these media are very different. Finally, by having longer scales, more variations can be seen. Here we did not report the distributions of the longer scale (without grouping) but they obviously offer more variations that the reduced scale.

1.3 Is Internet used a lot for watching TV, listening radio, reading newspapers and other activities?

The questions in the third split-ballot group also allow us to look if Internet is used a lot for watching TV, listening radio, reading newspapers and for other activities. We have information about the total time spent on Internet for these different purposes, and about the time spent on politics and current affairs more specifically.

Looking at the time of use of Internet will give us some information related to our first analyses (impact and usefulness of specifying explicitly that Internet should be

included). Internet became a central media so it is also interesting to have more information about its use. Finally in the perspective of a change from a face-to-face interview to a mixed-mode data collection, it is useful to have some information about the penetration of internet and the familiarity of the ESS respondents with Internet. Table 6 gives for this purpose the distribution of the variables related to the use of internet for the different activities with respect to the time in general and the time spend on politics and current affairs. The categories have been grouped again in order to show the distribution according to the classic scale (the one of the main questionnaire).

Table 6: distribution of the variables (categories grouped) in group 3

	Time in general				Time for politics			
	TV	radio	newsp	other	TV	radio	newsp	other
no time	**12561**	**13997**	**11188**	**7353**	**1521**	**1169**	**1075**	**4025**
<,5	1326	876	**3338**	**1496**	1451	881	**3363**	**3192**
,5-1	814	413	**1069**	**2219**	553	254	593	852
1-1,5	427	219	212	**1396**	166	84	77	232
1,5-2	300	150	95	**1025**	95	60	29	97
2-2,5	174	78	45	641	38	28	8	38
2,5-3	149	60	22	544	20	9	6	17
>3	256	207	29	1302	28	23	16	40
66	29846	29846	29847	29848	41803	43112	40572	37584
77	19	18	18	18	15	15	16	17
88	104	78	95	131	83	83	92	143
99	4810	4844	4828	4813	5013	5068	4939	4549
total	50786	50786	50786	50786	50786	50786	50786	50786

Looking at table 6, the more striking point is that in total not so much time is spent using the Internet: a big majority of respondents do not spend any time using the Internet for the different purposes, neither watching television nor listening to the radio, nor reading the newspapers. There are however differences across these three media. For newspapers, there are around 2800 respondents more than for radio that are using Internet but overall they use it for short periods (less than one hour).

There are more people using Internet for other activities than television, radio and newspapers. Indeed, there are less people answering “no time”. But still there is a high part of the sample that do not use Internet (close to half).

Besides that, out of the ones that are using Internet for these purposes, many are not spending any time for politics and current affairs issues. And when some time is spent on politics, it is usually less than 30 minutes, and mainly via online newspapers or other things (little TV, even less radio).

Conclusion:

Overall, a lot of respondents are not using Internet, which can explain why in section 1.1 we found that a majority of respondents had chosen the same category when it was specified to include Internet and when it was not. Moreover, it seems that when people are using Internet it is more for “other things” than for TV, radio or newspapers. Still some people are using it and for these respondents, it is better to clarify the question in order to be sure that they all interpret the question in the same way.

2. Evaluation of three MTMM experiments

The scales for the MTMM experiments are chosen in order to introduce variation in the forms we have evaluated. So far, we have always specified horizontal, partial labelled scales if we have 11 point scales. This means that we cannot separate the effect of the number of categories, partial labelling and horizontal or vertical scales. So in round 5 we have specified 11 point fully labelled scales based on the research on distances between category labels summarized in Krosnick and Fabrigar ("the book that may never be published"). We have also specified horizontal partial labelled scales with 4 and 5 points based on the same information.

Three topics have been chosen for the experiments:
1. The general evaluation of the police
2. The treatment by the police
3. Likelihood to be caught by the police

In each experiment, we have three traits measured by three methods (one in the main questionnaire, one in split-ballot group 1 and one in split-ballot group 2). For the details of the questions we refer to the appendix. Here we discuss the results of the experiments in the given sequence.

2.1. The general evaluation of the police

In this experiment a question was formulated about the success of the police in preventing crimes (trait 1 = T1), a question was formulated about the success of the police in catching criminals (T2) and one question about the speed with the police arrives on a place when there is a problem (T3).

The answer categories of all the questions were formulated as "item specific scales". All three scales were horizontal. The difference was that in the main questionnaire an 11 points partially labelled scale was used. In the first subgroup a 7 points fully labelled scale was used and in the second group a 5 point fully labelled scale was used. In the first two scales the end points of the scales were fixed reference point "extremely unsuccessful" (or slowly) and "extremely successful" (or quickly), while in the third method the end points were just "very unsuccessful" (or slowly) and "very successful" (or quickly).

On the basis of previous research, we expected that the 11 point scales is better than the 7 and 5 point scale, and that the scales with fixed reference points are better than the one without fixed reference point. So we expect to have the highest quality for method 1, then for method 2 and then for method 3.

Table 7 presents the mean reliability, validity and quality of the different questions for the different methods.

Table 7. Experiment 1: Mean estimates over all countries from the MTMM analyses of the general evaluation of the police for the 3 traits (T_i) and the 3 methods (M_i)

Expt1	$r^2 T_1$	$r^2 T_2$	$r^2 T_3$	$v^2 T_1$	$v^2 T_2$	$v^2 T_3$	$q^2 T_1$	$q^2 T_2$	$q^2 T_3$
M1	0.78	0.85	0.89	0.91	0.96	0.95	0.72	0.82	0.85
M2	0.71	0.74	0.74	0.75	0.78	0.86	0.53	0.58	0.63
M3	0.68	0.70	0.72	0.77	0.79	0.87	0.53	0.55	0.63

Note: T=trait, M=method, r^2=reliability, v^2=validity and q^2=quality

The table shows that the first method is the best with respect to reliability, validity and quality (i.e. the product of the reliability and the validity). This is what we expected. However, the two other scales (method 2 and 3) are with respect to quality approximately the same for all three traits, even if the one of method 2 has fixed reference points whereas the one of method 3 has not.

2.2. The treatment by the police

In the second experiment the questions used asked whether the police treat the people with respect (T1), make fair and impartial decisions (T2) and generally explain its decisions (T3). In this case, the answer categories chosen for the main questionnaire:

not at all often,	1
not very often,	2
often,	3
or, very often?	4
(Don't know)	8

The second method was a fully labelled 11 point vertical scale with fixed reference points on the end points "never" and "always".
The third method was a 11 point scale partially labelled vertical scale with the end points "almost never" and "almost always".

The categories of the first scale are a bit strange. We expected the first method to have the lowest quality since we expected confusion of the respondents about the first two categories in this first method.

In the second scale, some of the categories of the fully labelled scale are a bit complicated and we also expect them to be confusing (in particular, the category "more often than not") and to lead to a lower quality of that scale than of the partially labelled scale. On the other end, the 11 point scale of the second has fixed reference points at the end points, whereas the 11 point scale of the third method does not: we expect that to lower the quality of the third scale with respect to the one of the second. These two things taken together and since they go in opposite directions, we all in all expect methods 2 and 3 to have relatively similar quality.

The results are summarized in Table 8.

Table 8. Experiment 2: Mean estimates over all countries from the MTMM analyses of the treatment by the police for the 3 traits (T_i) and the 3 methods (M_i)

Expt 2	$r^2 T_1$	$r^2 T_2$	$r^2 T_3$	$v^2 T_1$	$v^2 T_2$	$v^2 T_3$	$q^2 T_1$	$q^2 T_2$	$q^2 T_3$
M1	0.64	0.64	0.62	0.65	0.61	0.72	0.43	0.39	0.45
M2	0.86	0.81	0.90	0.99	0.99	0.99	0.84	0.80	0.89
M3	0.89	0.85	0.92	0.98	0.98	0.98	0.87	0.83	0.91

Note: T=trait, M=method, r^2=reliability, v^2=validity and q^2=quality

Table 8 shows indeed that the first method led to a lot or errors and therefore the quality of this scale for all three questions is much worse than the quality of the other two methods: the quality of the first scale is only around half of the quality of the 2 other scales. This is really huge and has a big impact on the observed relationships between these variables that can easily lead to wrong conclusions.

The difference between the two other scales is not really present. The partially labelled 11 point scale (M3) is a little bit better than the fully labelled scale (M2) but the difference is quite small, which is in line with our hypothesis.

2.3. The likelihood to be caught by the police

In the third experiment the questions asked concerned the likelihood that you will be caught and punished by the police if you made a false insurance declaration (T1), if you bought something that you thought that might have been stolen (T2) or were involved in a traffic offence (T3).

The first method used is a 4 points horizontal fully labelled scale with categories: "not at all likely", "not very likely", "likely", "very likely". We expected this scale to lead to confusion because it is not clear, at least in other languages than the English language, what is the order of the categories. In principle with "very likely" and "not very likely" one covers the whole continuum but then the question is: what is the position of "likely"? Besides, the questions are part of a battery, which often also leads to lower quality.

Method 2 is also 4 point numeric horizontal scale with the same labels as method 1, but now it is not a battery but separate questions. We also expect it to have a low quality but not as low as the one of method 1 which is in a battery.

Method 3 is a vertical 4 points scale with only the end points labelled "very unlikely" and "very likely". The questions are treated separately and not in a battery and the labels are quite clear. So our expectation is that method 3 would be the best, then method 2 and finally method 1.

The results of the MTMM experiments are presented in Table 9.

Table 9. Experiment 3: Mean estimates over all countries from the MTMM analyses of the likelihood to be caught by the police for the 3 traits (T_i) and the 3 methods (M_i)

Expt 3	$r^2 T_1$	$r^2 T_2$	$r^2 T_3$	$v^2 T_1$	$v^2 T_2$	$v^2 T_3$	$q^2 T_1$	$q^2 T_2$	$q^2 T_3$
M1	0.76	0.78	0.70	0.59	0.57	0.62	0.45	0.45	0.43
M2	0.93	0.93	0.90	0.89	0.89	0.87	0.83	0.83	0.78
M3	0.86	0.89	0.79	0.90	0.90	0.88	0.78	0.81	0.70

Note: T=trait, M=method, r^2=reliability, v^2=validity and q^2=quality

The table shows indeed that Method 1 has the lowest reliability, validity and also quality. But surprisingly method 2 is the best, and we have a big difference between method 1 and 2 whereas the scales are using the same labels and are quite similar Method 3 is quite similar in term of quality to method 2 and with a relatively high quality as we could expect. The surprising result here is really the high quality of method 2. More research would be needed to confirm that point that is counterintuitive and not in line with previous research.

3. A new approach to measure Internal and External Political Efficacy

In the supplementary questionnaire for group 2, we have 4 variables about political efficacy. All the analyses in this section are therefore done on the sub-sample of the respondents of group 2 ("spltadmd" = 2 or 5). The 4 questions have been asked in order to test if these 4 variables are a good instrument to measure 2 different latent concepts:

- external validity → 2 reflective indicators testd22 and testd24
- internal validity → 2 reflective indicators testd23 and tesd25

The 2 are supposed to be correlated, but not too high, since we assume they are really two different concepts that also relates with very different strength to other variables as political participation, political trust or satisfaction with political institutions.

In order to test if these 4 variables are a "good" instrument, we are going first to test for the measurement equivalence of these 4 questions (configural, metric and scalar invariance). Then, we will look at the quality of the composite scores that can be made using these 4 variables. Finally, we will look at the relationship of the two composite scores of external and internal political efficacy such as measured by our 4 variables with three other composite scores: political participation, political trust and satisfaction.

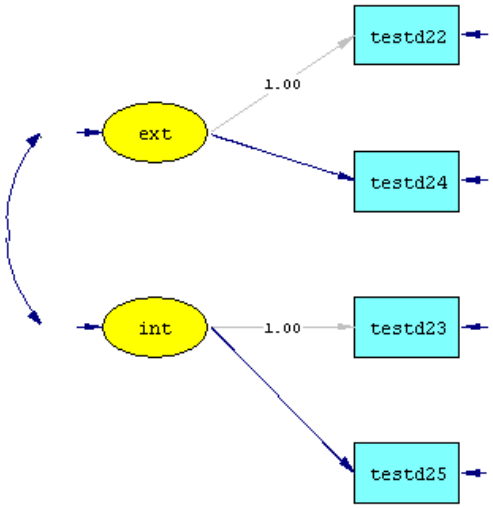
3.1 Test for measurement equivalence in LISREL

The number of observations in each group, the Pearson correlation matrices, the standard deviation and means are obtained in Stata.

Then, the testing for measurement equivalence is done in LISREL, using a multiple group analysis with 24 countries (Hungary is omitted since they do not complete the supplementary questionnaire and Ukraine is omitted also since the correlation matrix suggested there were a problem with the data).

The first step is to look at configural invariance, i.e. does the same model holds for all the 24 countries? The model that we are interested in has been mentioned at the beginning of section 1.3 and is represented in Figure 1.

Figure 1: the model tested for measurement equivalence across 24 countries



Results for configural invariance test

We get a chi-squared of 51.21 with df=24. The RMSEA is .044. According to this fit indices, the model does not have to be rejected. Because of the limits of the chi-square and fit indices that test at the global level, we also consider measures of local fit: Modification indices and Expected parameter changes, as well as the power, by using the JRule software. No relevant misspecifications are detected by the program.

Conclusion:

The different way of looking at the fit of the model all lead to the same conclusion that configural invariance holds over the 24 countries for the model tested. We can therefore go to the next level of invariance and test for metric equivalence.

Results for metric invariance test:

We get a chi-square of 258.9 with df=70. The RMSEA=.068. The increase in RMSEA is lower than .03, which is often use as an indication that equivalence holds. However, when looking at JRule, we can see that the second loading for internal political efficacy in Finland is really deviant. We therefore allow it to be free in this country, which leads to a new chi-square of 224.71 with df=69 and a new RMSEA of .062. JRule does not suggest any other relevant deviation.

Conclusion:

Metric invariance hols in all countries except in Finland: here the internal validity concept is deviant. The external political efficacy one is on the contrary metric invariant also in Finland. Therefore, it is possible to compare unstandardised relationships in all 24 countries for external political efficacy. For internal political efficacy, it is possible to compare them in 23 countries, excluding Finland. We can go on this the next step and test scalar invariance. For Finland, we let the intercept of the non-equivalent item free from the beginning.

Results for scalar invariance test:

The initial model is not converging. By starting freeing some intercepts that were not invariant convergence is achieved. We again use both indications from global and local fit measures in order to decide of the corrections that have to be made. We consider that a deviation in intercept wants to be detected if is around .4 (on a 5 point scale).

We finally:

- Had to free in Bulgaria and Estonia the intercept for the second item of external political efficacy
- Had to free in the Netherlands and Norway the second item of internal political efficacy.

The chi-square of this final model is 778.58 with 110 degrees of freedom. The RMSEA is 0.102. Both are still higher than what is usually used as criteria for accepting a model. But looking at the local fit, there are no more misspecifications larger than .4 that could be introduced. The high values of the classical global fit measures indicates that the test is very sensitive but on a substantive point of view, including more freedom on the parameters will not lead to relevant variations. So we stick with this final model.

Conclusion:

There is no scalar invariance for external political efficacy in Bulgaria and Estonia. So the means cannot be compared for these two countries for external political efficacy, but they can be compared in all the other countries.
There is no scalar invariance for internal political efficacy in the Netherlands and Norway. So the means cannot be compared for these two countries for internal political efficacy, but they can be compared in all the other countries.

General conclusion about measurement equivalence:

- all countries are configural invariant.
- all countries are metric invariant for external political efficacy, so we can compare unstandardized relationships for external political efficacy in all countries.
- 1 country (Finland) is not metric invariant for internal political efficacy whereas the 23 other countries are metric invariant. So we can compare unstandardized relationships for internal political efficacy in these 23 countries.
- 22 countries are scalar invariant for external political efficacy, so we can compare the means for external political efficacy in these 22 countries (exclude Bulgaria and Estonia)
- 21 countries are scalar invariant for internal political efficacy, so we can compare the means for internal political efficacy in these 21 countries (exclude Finland, the Netherlands and Norway)
- 19 countries are scalar invariant for both concepts, so we can compare the means in these 19 countries for both concepts (exclude Finland, Bulgaria, Estonia, the Netherlands and Norway).

3.2 Quality of the composite scores

Researchers often are not looking at the latent variables behind the observed answer but directly use these answers to combine them in a more general construct also called composite score. But these composite scores are not free of errors. The quality of the composite score can be defined in the same way as the one of single items: it is the strength of the relationship between the variable we are really interested in (latent) and the observed variable (composite score constructed directly from the observed variables). It can be computed using the following formula (see for instance Saris and Gallhofer, 2007):

$$q^2_{CS} = \rho^2(LV, CS) = \left(\frac{\sum_{i=1}^{n} q_i w_i}{\sqrt{var(CS)}} \right)^2 \quad (3)$$

Where: $$\sqrt{var(CS)} = \sum_i w_i^2 var(item_i) + 2\sum_{i,j} w_i w_j cov(item_i, item_j) \quad (4)$$

We compute the quality of the composite score for external political efficacy and the quality of the composite score for internal political efficacy. We use weights of .5 for each indicator in order to create the composite scores ("unweighted composite scores"). The results can be found in Table 10 (quality q^2).

Table 10: the quality q^2 of the composite scores for external ("ext") and internal ("int") political efficacy in the different countries

	BE	BG	CH	CY	CZ	DE	DK	EE	ES	FI	FR	GB
ext	.77	.56	.80	.89	.80	.77	.74	.70	.76	.76	.68	.76
int	.75	.81	.76	.90	.68	.70	.79	.68	.80	.74	.72	.69

	GR	HR	IE	IL	NL	NO	PL	PT	RU	SE	SI	SK
ext	.78	.74	.79	.85	.76	.75	.83	.82	.79	.71	.80	.68
int	.80	.78	.73	.80	.76	.70	.80	.80	.80	.66	.81	.69

Table 10 shows that there are indeed errors in the composite scores. The highest quality is indeed .89 for external and .90 for internal political efficacy (in Cyprus) meaning that at most 90% of the variance in the composite score is explained by the underlying latent variable. This is not perfect but this is however quite good. The problem is that there are variations in quality over countries and that in some countries the quality is much lower than that: for external political efficacy, in Bulgaria, the quality is only .56; for internal political efficacy in Sweden it is only .66. Therefore it is not possible to compare standardised relationships across countries without first correcting for measurement errors by taking the quality of the composite scores into account. Nevertheless, this does not mean that our instrument is not a good measure for external and internal political efficacy. There are always some measurement errors and as long as they can be corrected, this is not really problematic. The quality estimates here are overall quite good if we compare with what it used to be for other concepts in the ESS.

3.3 External validity

In order to see if the measure of external and internal political efficacy is working well, we look finally at the impact of internal and external political efficacy on:
- political participation
- political trust
- satisfaction with politics

The composite score for political participation is based on the variables B13 (contacted a politician) to B19 (boycotted certain products) of the main ESS questionnaire. These variables are dummies that we recoded: 0=no, 1=yes. We add them in order to create the composite score (Stata). We call it "pp".

For political trust, we create the composite score using the 3 items about trust in the parliament, the politicians and the political parties. We take the mean of these 3 variables as value for the composite score. We call it "pt".

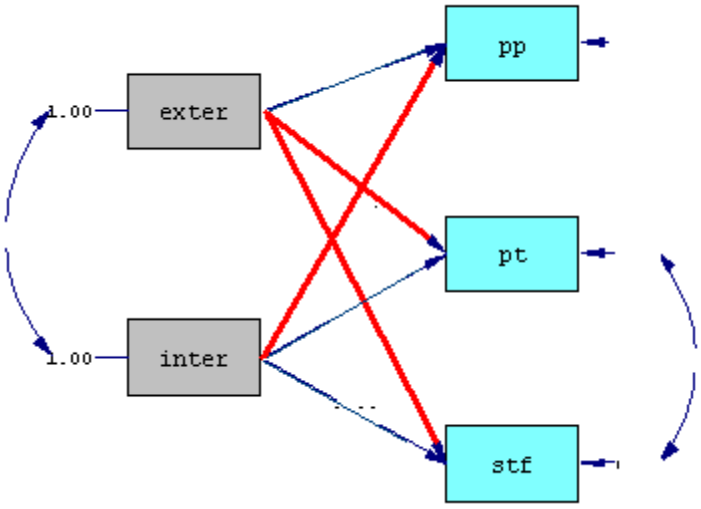
For satisfaction, we create the composite score using 2 items: satisfaction with the democracy and with the government and taking the mean. We call it "stf".

We look at these relationships because in the literature it has been argued that these three other concepts should be related with external and internal political efficacy at different levels. More precisely the hypotheses are that:

1) External validity loads higher on political trust and satisfaction and lower on political participation
2) Internal validity loads higher on political participation and lower on political trust and satisfaction.

So we analyse the model of Figure 2 in LISREL. The arrows in red are the one where we expect the high loadings. The observed variables here are the composite scores that we created as just explained.

Figure 2: the model estimated in all the 24 countries.

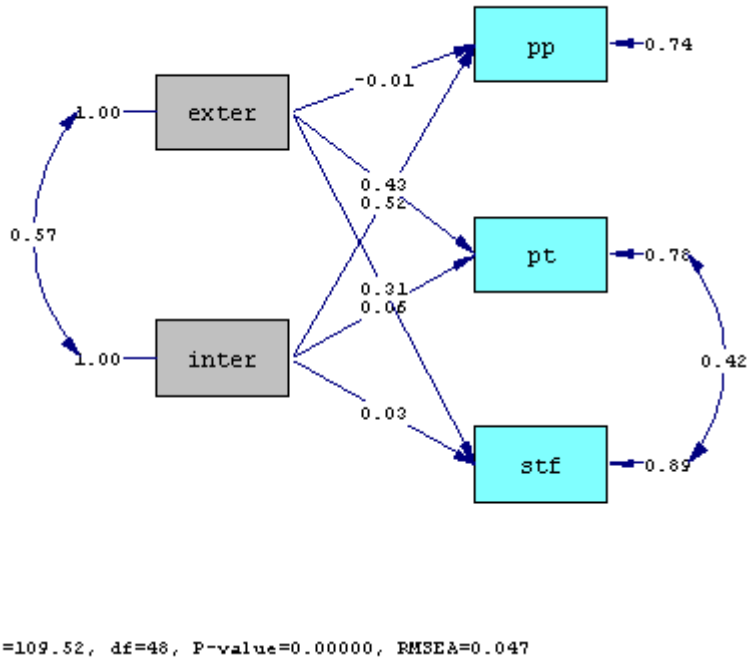


We correct for measurement errors for the 2 composite scores about political efficacy, by putting the quality estimates obtained in the previous analyses on the diagonal of the correlation matrices (reduction of variance approach). For the 3 others, we assume that the quality is perfect, even if this is unrealistic.

We should note that we had to allow political trust and satisfaction to be correlated otherwise the fit of the model was extremely bad. This is done in all countries.

The 24 countries are analysed together but with no cross-countries constraints. The results (once introduced the correlation between political trust and satisfaction) lead to an acceptable model. The estimates for Belgium are presented in Figure 3. They are in agreement with our hypotheses.

Figure 3: estimates for Belgium



Chi-Square=109.52, df=48, P-value=0.00000, RMSEA=0.047

The loadings for all countries are presented in Table 11. We can see that the same trend can be found in most countries. In 15 countries, the difference is really large between the effects that we expected to be lower or higher and the trend is really clear. In some other countries as Portugal, the differences are not so clear but still the pattern is

respected. In a few cases finally, there are some estimates that are not in line with our hypotheses: for instance in Ireland the effect of internal political efficacy on political participation is really low.

Table 11: estimates of the effects of external and internal political efficacy on political participation, political trust and satisfaction in the different countries

Country	Ext-pp	Ext-pt	Ext-stf	Int-pp	Int-pt	Int-stf
BE	-.01	.43	.31	.52	.06	.03
BG	.10	.20	.40	.34	.13	-.13
CH	.02	.36	.45	.45	-.17	-.17
CY	-.05	.26	.52	.38	.05	-.16
CZ	-.12	.40	.55	.42	-.01	-.14
DE	.03	.34	.42	.47	.02	-.10
DK	-.08	.54	.50	.45	-.02	*-.33*
EE	.01	.56	.63	.44	-.14	-.12
ES	.07	.44	.46	.49	-.02	-.11
FI	-.02	.38	.41	.39	.05	-.08
FR	-.09	.46	.45	.60	.10	-.11
GB	-.01	.45	.45	.45	.05	-.04
GR	-.15	.34	.41	.41	-.12	*-.25*
HR	-.03	.33	.31	.40	-.12	-.05
IE	*-.26*	.39	.32	.53	-.03	-.13
IL	.06	.28	.24	***.05***	-.01	-.09
NL	-.13	.60	.61	.46	-.10	*-.19*
NO	-.09	.42	.53	.50	.02	*-.21*
PL	.02	.28	.30	.39	.01	.02
PT	*-.20*	*-.09*	.22	.45	*.30*	.02
RU	*-.17*	.20	.30	.35	-.01	-.08
SE	-.11	.42	.43	.53	.04	-.02
SI	.12	.38	.38	.38	.04	-.01
SK	.09	.39	.26	.32	*-.32*	-.07
Expected	lower	higher	higher	higher	lower	lower

Conclusion:

Overall, our hypotheses get quite some support, so it seems that the two composite scores for political efficacy are working in the expected way and quite well.

General conclusion about the analyses of political efficacy:

All together with the results from the measurement equivalence testing and the computation of the quality, we can conclude that in most of the countries the new proposal for measuring political efficacy with 2 correlated concepts representing internal and external political efficacy, each measured with 2 items (testd22 and testd24 for external, testd23 and testd25 for internal) is working quite well. Most countries achieved the three levels of equivalence, the quality of the composite scores is quite high even if it varies across countries and should be taken therefore into account when doing the analyses. The relationships with three other concepts finally are in most countries as expected.

Appendix: Questions in the Main questionnaire involved in the MTMM experiments

1.Media use

A1 CARD 1 On an average weekday, how much time, in total, do you spend watching television? Please use this card to answer.

No time at all	00	**GO TO A3**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to1½ hours	03	
More than 1½ hours, up to 2 hours	04	**ASK A2**
More than 2 hours, up to 2½ hours	05	
More than 2½ hours, up to 3 hours	06	
More than 3 hours	07	
(Don't know)	88	

A2 STILL CARD 1 And again on an average weekday, how much of your time watching television is spent watching news or programmes about politics and current affairs[1]? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to 1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours	07
(Don't know)	88

[1] About "politics and current affairs": about issues to do with governance and public policy, and with the people connected with these affairs.

ASK ALL

A3 STILL CARD 1 On an average weekday, how much time, in total, do you spend listening to the radio? Use the same card.

No time at all	00	**GO TO A5**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to 1½ hours	03	
More than 1½ hours, up to 2 hours	04	
More than 2 hours, up to 2½ hours	05	**ASK A4**
More than 2½ hours, up to 3 hours	06	
More than 3 hours	07	
(Don't know)	88	

A4 STILL CARD 1 And again on an average weekday, how much of your time listening to the radio is spent listening to <u>news</u> or programmes about <u>politics and current affairs</u>? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to 1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours	07
(Don't know)	88

ASK ALL

A5 STILL CARD 1 On an average weekday, how much time, in total, do you spend reading the newspapers? Use this card again

No time at all	00	**GO TO A7**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to 1½ hours	03	
More than 1½ hours, up to 2 hours	04	
More than 2 hours, up to 2½ hours	05	**ASK A6**
More than 2½ hours, up to 3 hours	06	
More than 3 hours	07	
(Don't know)	88	

A6 STILL CARD 1 And how much of this time is spent reading about politics and current affairs? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to 1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours	07
(Don't know)	88

ASK ALL

A7 CARD 2 Now, using this card, how often do you use the internet, the World Wide Web or e-mail – whether at home or at work – for your personal[2] use?

No access at home or work	00
Never use	01
Less than once a month	02
Once a month	03
Several times a month	04
Once a week	05
Several times a week	06
Every day	07
(Don't know)	88

[2] "Personal use'" is private or recreational use that does not have to do with one's work or occupation.

2. General evaluation of the police

ASK ALL

B7 CARD X From what you have heard or experienced how successful do you think the police[3] are at preventing crimes in [country] where violence is used or threatened? Choose your answer from this card, where 0 means extremely unsuccessful and 10 means extremely successful.

Extremely Unsuccessful										**Extremely successful**	**(Don't know)**
00	01	02	03	04	05	06	07	08	09	10	88

B8 STILL CARD X And how successful do you think the police are at catching people who commit house burglaries[4] in [country]? Choose your answer from this card, where 0 means extremely unsuccessful and 10 means extremely successful.

Extremely Unsuccessful										**Extremely successful**	**(Don't know)**
00	01	02	03	04	05	06	07	08	09	10	88

B9 CARD X If a violent crime or house burglary were to occur near to where you live and the police were called[5], how quickly do you think they would arrive at the scene? Choose your answer from this card, where 0 means not at all quickly and 10 means extremely quickly.

Extremely slowly										**Extremely quickly**	**(Don't know)**
00	01	02	03	04	05	06	07	08	09	10	88

(Not possible for the police to arrive at the scene quickly near to where I live) 55

(Violent crimes and / or house burglaries never occur near to where I live) 56

[3] Again the generic name for all police in [country] should be used (see footnote 5). The question itself and those that follow provide specific cues to respondents that might limit the frame of reference to a specific group of police in some countries. However this should be achieved by the crime referred to and NOT by amending the name of the specific police referenced.

[4] House burglaries occur when someone uninvited breaks into a property with the intention of stealing.

[5] Called in the sense of telephoned.

3. Treatment by the police

Now some questions about how the police[6] who deal with crimes such as burglary and physical assault in [country] typically treat members of the public[7] they come into contact with. The next few questions do not have a show card.

B12 Based on what you have heard or your own experience, about[8] how often would you say that the police in [country] treat members of the public with respect? Would you say…**READ OUT…**

…not at all often,	1
not very often,	2
often,	3
or, very often?	4
(Don't know)	8

B13 About[9] how often would you say that the police [in country] make fair[10] decisions in the cases they deal with? Would you say…**READ OUT…**

…not at all often,	1
not very often,	2
often,	3
or, very often?	4
(Don't know)	8

B14 And when dealing with members of the public[11], how often would you say the police explain their decisions and actions when asked to do so? Would you say…**READ OUT…**

…not at all often,	1
not very often,	2
often,	3
or, very often?	4
(No one ever asks the police to explain their decisions and actions)	(5)
(Don't know)	(8)

---

[6] Note we do not suppose there is a police force that only deals with crimes such as burglary and physical assault in a particular country. Refer here to the police force or forcer that, (among other things), deal with such crimes.

[7] Members of the public' in the sense of the general public / everyone in society. References to 'people' should be avoided because this may lead respondents to think only of the people the police deal with most frequently when wider society is intended.

[8] 'About' – meaning 'approximately' or 'roughly'.

[9] See footnote 20.

[10] 'Fair decisions' in the sense of 'just decisions'.

[11] See footnote 19.

4. Likelihood to be caught

Now some questions about how likely it is that you would be caught and punished[12] if you did certain things in [country].

CARD X How likely is it that you would be caught and punished in [country] [13] if you…**READ OUT…**

		Not at all likely	Not very likely	Likely	Very likely	(Don't know)
B40	...made an exaggerated or false insurance claim[14]?	1	2	3	4	8
B41	...bought something you[15] thought might be stolen?	1	2	3	4	8
B42	...committed a traffic offence like speeding or crossing a red light?	1	2	3	4	8

[12] 'Punished' as in 'punished by the law'; this could be in the form of a prison sentence, fine or any other sentence.

[13] See footnote 37.

[14] The answer code itself is item E15 in ESS Round 2 but the question stem is different.

[15] 'You' as in 'the respondent personally'.

Supplementary group 1

The first few questions concern the amount of time you spend watching television, listening to the radio and reading newspapers.

A1CARD X On an average weekday, how much time, in total, do you spend watching television? Please include any time spent watching TV via the internet. Please use this card to answer

No time at all 00
Less than ½ hour 01
½ hour to 1 hour 02
More than 1 hour, up to1½ hours 03
More than 1½ hours, up to 2 hours 04
More than 2 hours, up to 2½ hours 05
More than 2½ hours, up to 3 hours 06
More than 3 hours 07
(Don't know) 88

A3CARD X On an average weekday, how much time, in total, do you spend listening to the radio? Please include any time spent listening to the radio via the internet. Use the same card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours	07
(Don't know)	88

A5 CARD X And on an average weekday, how much time, in total, do you spend reading the newspapers? Please include any time spent reading news papers via the internet. Use this card again.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to 1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours	07
(Don't know)	88

2. Evaluation of the police

B7 CARD X Based on what you have heard or experienced how successful do you think the police[16] are at preventing crimes in [country] where violence is used or threatened?

Extremely unsuccessful	Very unsuccessful	Rather unsuccessful	Neither unsuccessful nor successful	Rather successful	Very successful	Extremely successful
00	01	02	03	04	05	06
☐	☐	☐	☐	☐	☐	☐

B8STILL CARD X And how successful do you think the police are at catching people who commit house burglaries[17] in [country]?

Extremely unsuccessful	Very unsuccessful	Rather unsuccessful	Neither unsuccessful nor successful	Rather successful	Very successful	Extremely successful
00	01	02	03	04	05	06
☐	☐	☐	☐	☐	☐	☐

B9 CARD X If a violent crime or house burglary were to occur near to where you live and the police were called[18], how slowly or quickly do you think they would arrive at the scene?

Extremely slowly	Very slowly	Rather slowly	Neither slowly nor quickly	Rather quickly	Very quickly	Extremely quickly
00	01	02	03	04	05	06
☐	☐	☐	☐	☐	☐	☐

(Not possible for the police to arrive at the scene quickly near to where I live) 55
(Violent crimes and / or house burglaries never occur near to where I live) 56

3. Treatment by the Police

Now some questions about when the police deal with crimes like house burglary and physical assault.

B12 CARD X Based on what you have heard or your own experience, about[19] how often would you say that the police in [country] treat members of the public with respect?

never, 0
hardly never, 1
rarely, 2

[16] Again the generic name for all police in [country] should be used (see footnote 5). The question itself and those that follow provide specific cues to respondents that might limit the frame of reference to a specific group of police in some countries. However this should be achieved by the crime referred to and NOT by amending the name of the specific police referenced.

[17] House burglaries occur when someone uninvited breaks into a property with the intention of stealing.

[18] Called in the sense of telephoned.

[19] 'About' – meaning 'approximately' or 'roughly'.

occasionally, 3
sometimes 4
half of the time 5
more often than not 6
often, 7
very often, 8
almost always, 9
always 10
(Don't know) 88

B13 CARD X About[20] how often would you say that the police [in country] make fair, impartial decisions in the cases they deal with?

never, 0
hardly never, 1
rarely, 2
occasionally, 3
sometimes 4
half of the time 5
more often than not 6
often, 7
very often, 8
almost always, 9
always 10
(Don't know) 88

B14 **CARD X** And when dealing with people in [country], how often would you say the police explain their decisions and actions when asked to do so?

never, 0
hardly never, 1
rarely, 2
occasionally, 3
sometimes 4
half of the time 5
more often than not 6
often, 7
very often, 8
almost always, 9
always 10
(Don't know) 88
(No one ever asks the police to explain their decisions and actions 55)

[20] See footnote 20.

4. Likelihood to be caught

Now some questions about how likely it is that you would be caught and punished if you did certain things in [country].

B40 How likely is it that you would be caught and punished if you made an exaggerated or false insurance claim[21]?

Not at all likely	Not very likely	Likely	Very likely
1 ☐	2 ☐	3 ☐	4 ☐

B41 How likely is it that you would be caught and punished if you bought something you[22] thought might be stolen?

Not at all likely	Not very likely	Likely	Very likely
1 ☐	2 ☐	3 ☐	4 ☐

B42 How likely is it that you would be caught and punished if you committed a traffic offence like speeding or crossing a red light?

Not at all likely	Not very likely	Likely	Very likely
1 ☐	2 ☐	3 ☐	4 ☐

[21] The answer code itself is item E15 in ESS Round 2 but the question stem is different.

[22] 'You' as in 'the respondent personally'.

Supplementary questionnaire group 2

2. Evaluation of the police

B7 CARD X Based on what you have heard or experienced how unsuccessful or successful do you think the police[23] are at <u>preventing</u> crimes in [country] where violence is used or threatened?

Very unsuccessful	Rather successful	Neither unsuccessful nor successful	Rather successful	Very Successful	(Don't know)
0	1	2	3	4	8

B8 STILL CARD X And how unsuccessful or successful do you think the police are at <u>catching</u> people who commit house burglaries[24] in [country]?

Very unsuccessful	Rather successful	Neither unsuccessful nor successful	Rather successful	Very Successful	(Don't know)
1	2	3	4	5	8

B9 CARD X If a violent crime or house burglary were to occur near to where you live and the police were called[25], how slowly or quickly do you think they would arrive at the scene?

Very Slowly	Rather slowly	Neither slowly nor quickly	Rather quickly	Very Quickly	(Don't know)
1	2	3	4	5	8

(Not possible for the police to arrive at the scene quickly near to where I live) 55

(Violent crimes and / or house burglaries never occur near to where I live) 56

[23] Again the generic name for all police in [country] should be used (see footnote 5). The question itself and those that follow provide specific cues to respondents that might limit the frame of reference to a specific group of police in some countries. However this should be achieved by the crime referred to and NOT by amending the name of the specific police referenced.

[24] House burglaries occur when someone uninvited breaks into a property with the intention of stealing.

[25] Called in the sense of telephoned.

3. Treatment by the police

B12 CARD X Based on what you have heard or your own experience, about[26] how often would you say that the police in [country] treat members of the public with respect?

Almost never	0	☐
	1	☐
	2	☐
	3	☐
	4	☐
	5	☐
	6	☐
	7	☐
	8	☐
	9	☐
Almost always	10	☐

B13 CARD X About[27] how often would you say that the police make fair, impartial decisions in the cases they deal with?

Almost never	0	☐
	1	☐
	2	☐
	3	☐
	4	☐
	5	☐
	6	☐
	7	☐
	8	☐
	9	☐
Almost always	10	☐

B14 CARD X And when dealing with people in [country], how often would you say the police generally explain their decisions and actions when asked to do so?

Almost never	0	☐
	1	☐
	2	☐
	3	☐
	4	☐
	5	☐
	6	☐
	7	☐
	8	☐
	9	☐
Almost always	10	☐

(No one ever asks the police to explain their decisions and actions 55)

[26] 'About' – meaning 'approximately' or 'roughly'.

[27] See footnote 20.

4. Likelihood to be caught

Now some questions about how likely it is that you would be caught and punished[28] if you did certain things in [country].

B40 How likely is it that you would be caught and punished in [country] if you made an exaggerated or false insurance claim[29]?

Very unlikely	1	☐
	2	☐
	3	☐
Very likely	4	☐

B41 How likely is it that you would be caught and punished in [country] if you bought something you[30] thought might be stolen?

Very unlikely	1	☐
	2	☐
	3	☐
Very likely	4	☐

B42 How likely is it that you would be caught and punished in [country] if you committed a traffic offence like speeding or crossing a red light?

Very unlikely	1	☐
	2	☐
	3	☐
Very likely	4	☐

[28] 'Punished' as in 'punished by the law'; this could be in the form of a prison sentence, fine or any other sentence.

[29] The answer code itself is item E15 in ESS Round 2 but the question stem is different.

[30] 'You' as in 'the respondent personally'.

Direct measures of Internal efficacy or subjective competence

1. Do you think that you could take an active role in a group involved with political issues?

Definitely not	1
Probably not	2
Not sure either way	3
Probably	4
Definitely	5
(Don't know)	8

2. How confident are you in your own ability to participate in politics?

Not at all confident	1
A little confident	2
Quite confident	3
Very confident	4
Completely confident	5
(Don't know)	8

Direct measures for External efficacy or System responsiveness

1. how much would you say the political system in [country] allows people like you to have a say about what the government does.?

Not at all	1
Very little	2
Not much	3
Much	4
Very much	5
(Don't know)	8

2. And how much would you say that the political system in [country] allows people like you to have a direct influence on politics ? READ OUT…

Not at all	1
Very little	2
Not much	3
Much	4
Very much	5
(Don't know)	8

Supplementary group 3

The first few questions concern the amount of time you spend watching television, listening to the radio and reading newspapers.

A1 CARD X On an average weekday, how much time, in total, do you spend watching television? Please include the time you spend doing these activities using the internet . Please use this card to answer.

No time at all	00	**GO TO A3**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to1½ hours	03	
More than 1½ hours, up to 2 hours	04	
More than 2 hours, up to 2½ hours	05	**ASK A2**
More than 2½ hours, up to 3 hours	06	
More than 3 hours, up to 4 hours	07	
More than 4 hours, up to 5 hours	08	
More than 5 hours, up to 6 hours	09	
More than 6 hours	10	
(Don't know)	88	

A2 STILL CARD X And again on an average weekday, how much of your time watching television is spent watching **news** or programmes about **politics and current affairs**? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours, up to 4 hours	07
More than 4 hours, up to 5 hours	08
More than 5 hours, up to 6 hours	09
More than 6 hours	10
(Don't know)	88

ASK ALL

A3 STILL CARD X On an average weekday, how much time, in total, do you spend listening to the radio? Please include the time you spend doing these activities using the internet . Use the same card.

No time at all	00	**GO TO A5**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to1½ hours	03	
More than 1½ hours, up to 2 hours	04	**ASK A4**
More than 2 hours, up to 2½ hours	05	
More than 2½ hours, up to 3 hours	06	
More than 3 hours, up to 4 hours	07	
More than 4 hours, up to 5 hours	08	
More than 5 hours, up to 6 hours	09	
More than 6 hours	10	
(Don't know)	88	

A4 STILL CARD X And again on an average weekday, how much of your time listening to the radio is spent listening to **news** or programmes about **politics and current affairs**? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours, up to 4 hours	07
More than 4 hours, up to 5 hours	08
More than 5 hours, up to 6 hours	09
More than 6 hours	10
(Don't know)	88

ASK ALL

A5CARD X On an average weekday, how much time, in total, do you spend reading the newspapers? Please include the time you spend doing these activities using the internet. Use this card.

No time at all	00 **GO To A7**
Less than 15 minutes	01
15 minutes up to ½ hour	02
More than ½ hour up to 45 minutes	03
More than 45 minutes up to 1 hour	04**ASK A6**
More than 1 hour, up to 1½ hours	05
More than 1½ hours, up to 2 hours	06
More than 2 hours, up to 2½ hours	07
More than 2½ hours, up to 3 hours	08
More than 3 hours	09
(Don't know)	88

A6 STILL CARD X And how much of this time is spent reading about **politics and current affairs**? Still use this card.

No time at all	00
Less than 15 minutes	01
15 minutes up to ½ hour	02
More than ½ hour up to 45 minutes	03
More than 45 minutes up to 1 hour	04
More than 1 hour, up to 1½ hours	05
More than 1½ hours, up to 2 hours	06
More than 2 hours, up to 2½ hours	07
More than 2½ hours, up to 3 hours	08
More than 3 hours	09
(Don't know)	88

ASK ALL

We would now like to ask you questions about your use of the internet.

A7 CARD X On an average weekday, how much time, in total, do you spend using the internet to watch television programmes online. Please use this card.

No time at all	00	**GO TO A9**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to1½ hours	03	
More than 1½ hours, up to 2 hours	04	**ASK A8**
More than 2 hours, up to 2½ hours	05	
More than 2½ hours, up to 3 hours	06	
More than 3 hours, up to 4 hours	07	
More than 4 hours, up to 5 hours	08	
More than 5 hours, up to 6 hours	09	
More than 6 hours	10	
(Don't know)	88	

A8 CARD X And how much of this time is spent watching online programs about **politics and current affairs**? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours, up to 4 hours	07
More than 4 hours, up to 5 hours	08
More than 5 hours, up to 6 hours	09
More than 6 hours	10
(Don't know)	88

ASK ALL

A9 CARD X On an average weekday, how much time, in total, do you spend using the internet to listen the radio online Please use this card.

No time at all	00	**GO TO A11**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to1½ hours	03	
More than 1½ hours, up to 2 hours	04	**ASK A10**
More than 2 hours, up to 2½ hours	05	
More than 2½ hours, up to 3 hours	06	
More than 3 hours, up to 4 hours	07	
More than 4 hours, up to 5 hours	08	
More than 5 hours, up to 6 hours	09	
More than 6 hours	10	
(Don't know)	88	

A10 CARD X And how much of this time is spent listening radio programs online about **politics and current affairs**? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours, up to 4 hours	07
More than 4 hours, up to 5 hours	08
More than 5 hours, up to 6 hours	09
More than 6 hours	10
(Don't know)	88

ASK ALL

A11 CARD X On an average weekday, how much time, in total, do you spend using the internet to read the newspapers online. Please use this card.

No time at all	00 **GO toA13**
Less than 15 minutes	01
15 minutes up to ½ hour	02
More than ½ hour up to 45 minutes	03
More than 45 minutes up to 1 hour	04 **Ask A12**
More than 1 hour, up to 1½ hours	05
More than 1½ hours, up to 2 hours	06
More than 2 hours, up to 2½ hours	07
More than 2½ hours, up to 3 hours	08
More than 3 hours	09
(Don't know)	88

A12 CARD X And how much of this time is spent reading online newspapers about **politics and current affairs**? Still use this card.

No time at all	00
Less than 15 minutes	01
15 minutes up to ½ hour	02
More than ½ hour up to 45 minutes	03
More than 45 minutes up to 1 hour	04
More than 1 hour, up to 1½ hours	05
More than 1½ hours, up to 2 hours	06
More than 2 hours, up to 2½ hours	07
More than 2½ hours, up to 3 hours	08
More than 3 hours	09
(Don't know)	88

ASK ALL

A13 CARD X On an average weekday, how much time, in total, do you spend using the internet for other activities than watching television, listening radio and reading newspapers. Please use this card.

No time at all	00	**GO TO B7**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to1½ hours	03	
More than 1½ hours, up to 2 hours	04	**ASK A14**
More than 2 hours, up to 2½ hours	05	
More than 2½ hours, up to 3 hours	06	
More than 3 hours, up to 4 hours	07	
More than 4 hours, up to 5 hours	08	
More than 5 hours, up to 6 hours	09	
More than 6 hours	10	
(Don't know)	88	

A14 CARD X And how much of this time is spent on activities about **politics and current affairs** (excluding watching television, listening radio and reading online newspapers)? Still use this card.

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours, up to 4 hours	07
More than 4 hours, up to 5 hours	08
More than 5 hours, up to 6 hours	09
More than 6 hours	10
(Don't know)	88