Mode effects in the standard Eurobarometer questions

Saris, Willem E.; Hagenaars, Jacques A.

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

CHAPTER 6

MODE EFFECTS IN THE STANDARD EUROBAROMETER QUESTIONS

WILLEM E. SARIS AND JACQUES A. HAGENAARS

6.1 Introduction

In many studies it is suggested that a change from personal to telephone interviewing does not make much of a difference (Groves and Kahn, 1979; De Leeuw and Van der Zouwen, 1988; De Leeuw, 1992). There are, however, also studies which indicate quite large mode effects. Silberstein and Scott (1991) have discovered large mode effects in family expenditure research. Kalfs (1994) has shown for time budget research quite large differences for media use and transport between telephone interviewing and self administrated interviews. Scherpenzeel (1995) has also found large mode effects between telephone interviews and personal interviews for two topics.

These contradictory results seem to suggest that in controlled experiments small differences are found while in real life data collections where the procedures for the different modes are in some sense optimised differences can occur.

With respect to the reasons for differences in univariate distributions the research group suggested that three aspects had to be studied. The first obvious reason for differences was that the results of the telephone and personal interviews will be different due to the differences in penetration of telephones in the different countries (coverage error). A second issue concerns the effect of the organisational characteristics of the fieldwork. These activities lead to more or less participation and nonresponse.

A third component is the pure mode effect, an effect of the medium which is used in the data collection: a direct face to face interview or an interview using a mediating instrument like a telephone.

The aspect of sampling has been discussed in the first part of the book. In the previous chapter we have seen that the difference between telephone and face to face interviewing can be decomposed into three components of which the coverage error is the smallest and the organisational component the largest. It was also shown that the pure mode effects could be considerable. Therefore, in this second part of the book the main emphasis will be on mode effects.

In this chapter the mode effects for the standard Eurobarometer questions will be analysed. In chapter 9, several other questions, less frequently used in the Eurobarometers, will be studied which allows the evaluation of the mode effects and cross cultural differences for different types of questions. In between, in chapters 7 and 8, attention will be given to two special types of questions. First the effect of the mode of data collection on two open ended questions will be discussed. In chapter 8 the effect of the mode of data collection and the adjustment to the different modes of the formulation for the very commonly used Left Right scale will be scrutinised. In this chapter, a model will be formulated which can be used to describe and test mode effects for questions with precoded answers.

6.2 The latent class model applied for mode effects

In any data collection measurement errors are made. As long as these errors are similar except for random fluctuations, it is no problem to switch from one mode of data collection to another. If, however, different modes produce systematically different errors this switch is not so simply made. It has already been shown that for several questions systematic bias due to the mode of data collection has been observed. Therefore, an explanatory model for responses in surveys is presented.

For this purpose the formalisation of the latent class model developed by Lazarsfeld (1950a; 1950b) is used. Imagine the simplest case of a variable (x) with two categories, for example people who think that the own country has benefited of the EU (x=1) or not (x=2). The percentage of people of the population in each category π^x_1 and π^x_2 is by definition unknown. The only information which can be obtained is the percentage of people answering a question positively or negatively in a sample. But these questions can be formulated in many different ways, the data can be collected in many different ways, and each approach can lead to a different response distribution.

In this specific project the question remained the same, but it is asked either in a face to face interview or in a telephone interview. Using a slightly different notation of Goodman (1974a; 1974b) and Hagenaars (1990), the conditional probabilities to react positively or negatively, given the score on the variable x, can be presented in a matrix as follows

variable latent	x=1	x=2	marginal
observed			
1	π^f_{11}	π^f_{12}	π^f_1
2	π^f_{21}	π^f_{22}	π^f_2
	1.0	1.0	1.0

variable latent	x=1	x=2	marginal
observed			
1	π^t_{11}	π^t_{12}	π^t_1
2	π^t_{21}	π^t_{22}	π^t_2
	1.0	1.0	1.0

If we denote the proportion of people who have score 1 on variable x as π^x_1 and those with score 2 on variable x as π^x_2 than we can formulate that the proportion of people who say yes in the face to face study is equal to

$$\pi^f_1 = \pi^f_{11}\,\pi^x_1 + \pi^f_{12}\,\pi^x_2$$

and the proportion of people who say no is equal to

$$\pi^f_2 = \pi^f_{21}\,\pi^x_1 + \pi^f_{22}\,\pi^x_2$$

In the same way we can formulate for the telephone interviewing that the proportion of people who say yes is

$$\pi^t_1 = \pi^t_{11}\,\pi^x_1 + \pi^t_{12}\,\pi^x_2$$

and the proportion who say no is

$$\pi^t_2 = \pi^t_{21}\,\pi^x_1 + \pi^t_{22}\,\pi^x_2$$

In matrix algebra this can be simplified to

$$\pi^f = \Pi^f \pi^x \qquad (1)$$

and

$$\pi^t = \Pi^t \pi^x \qquad (2)$$

where π^f is the vector with the marginal distribution obtained by face to face interview

π^t is the vector with the marginal distribution obtained by telephone interview

π^x is the vector with the marginal distribution of x

Π^f is the response probability matrix in face to face interview given the score on x

Π^t is the response probability matrix in telephone interview given the score on x

If the probabilities that one will answer a question positively in a face to face interview (π^f_{11}, π^f_{12}) and in telephone interview (π^t_{11}, π^t_{12}), given the score on the latent variable x, are the same then the distribution of the observed distributions (π^f_1, π^f_2) and (π^t_1, π^t_2) will also be the same except for random fluctuations. If, however, the probabilities are unequal than the distribution for the different variables will also be different.

This point can be illustrated with a simple example. Imagine that π^x_1 =.9 and π^x_2 = .1 while the response probabilities are as given in table 6.1.

Table 6.1 Response probabilities

Face to face	x=1	x=2	**Telephone**	x=1	x=2
1	.8	.1	1	.9	.4
2	.2	.9	2	.1	.6

Due to this difference in tendency to say yes to the same question in personal and telephone interviews the distributions for the two variables will become different. The distribution on the face to face would be π_1 = .73 which is .8 x .9 + .1x.1 and π_2 = .27 while for telephone interviewing will be π^t_1 = .85 and π^t_2 =.15. This difference in distribution would, of course, not have occurred if the response probabilities would have been the same.

Such tendencies to prefer certain categories more in one mode than in another can, for example occur if in personal interviewing show cards are used and on the telephone the response categories are read by the interviewer in a fixed order. This is only one reason why such differences will be found. In the literature many other reasons can be found (Groves, 1989; De Leeuw, 1992).

In practice one does not know the distribution of x and the response probabilities but only the distributions for the two response variables. Schuman and Presser (1981), Billiet et al. (1986) and many others have shown how these differences in distributions can be tested using a research design with independent samples from the same population. Only under extreme experimental conditions one can use these so called “split ballot experiments” for a test on the effect of the mode of data collection. In general, also other differences such as coverage and nonresponse errors will play a role as was shown before. Another problem is that the response probabilities can not be estimated from such a design.

In case of panel studies this is possible by using the “turnover table”. This table presents the relationship between the responses collected with the different independent modes. In our example this table would look like table 6.2.

Table 6.2 The relationship between the observed variables if π_1^x = .9 and π_2^x = .1 and the response probabilities of table 6.1.

Face to face	**Telephone**		
	1	2	Total
1	.652	.078	.730
2	.198	.072	.270
Total	.85	.15	1.0

This table shows the distributions of the two variables in the marginals while the combinations of the response variables can be found in the cells of the matrix.

From the model specified before it follows that the table denoted by $\mathbf{T}^{ft}$ can be written as a function of the matrices with the response probabilities and the values of the latent variable if it can be assumed that the modes are independent of each other given the value of x and that x is stable over time. In order to do so, we first create a diagonal matrix **X** which contains on the diagonal the values of the latent variable in our example, thus the number or proportion of people in the classes x_1 and x_2:

$$\mathbf{X} = \begin{vmatrix} .9 & .0 \\ .0 & .1 \end{vmatrix} \tag{3}$$

Using this matrix the $\mathbf{T}^{ft}$ can be shown to be:

$$\mathbf{T}^{ft} = \Pi^{f}.X.\ \Pi^{t'} \tag{4}$$

So if the matrix with the proportions of people in the latent classes is pre- and post-multiplied by the two matrices representing the response probabilities, one gets table $\mathbf{T}^{ft}$. This formulation is attractive because it makes the connection between the table obtained from the panel study and the model characteristics one is really interested in. As one does not know the values of the probabilities in the two matrices $\Pi^{f}.\ \Pi^{t}$ and the matrix (**X**) the estimation of these values is the task to be done.

In this chapter, these response probabilities will be estimated in order to see whether they are different for the different modes of data collection. If they are different, one can expect differences in the distribution of the observed variables. If they are the same, no mode effect can be found given that X is the same for the two modes.

6.3 Research design

In order to test the equality of the response probabilities, data have to be collected from the same people in two different ways so that a turnover table can be constructed as indicated above. Such a panel study has been done as a continuation of the Eurobarometer study 41. In the personal interview the interviewer noted whether the people had a telephone. The households with a telephone have been contacted again within a period of one to two weeks for a second interview, this time by telephone, with a set of the most important questions of the Eurobarometer. This panel experiment has been done only in France, Belgium and Spain. These countries were selected because they had large differences in telephone penetration. In France approximately 350 people have completed a personal as well as a telephone interview, in Belgium approximately 250 and in Spain 320 people (see chapter 2 for details). Although these samples are much smaller than the original samples, it has been found that for most variables the distribution of the responses of the respondents did not deviate significantly from the responses of the original samples. This result suggests that the people who dropped out the study at the occasion of the second interview did not hold different opinions on the issues covered as the people who did not drop out. To continue this analysis, an important assumption was made:

The people who continued in this research did not differ in their response behaviour from the people who dropped out after the personal interview

While the research group believes that this assumption is not very strong and most likely true, the data nevertheless do not allow a test for this assumption.

In order to be able to test the equality of the response probabilities for the different response modes, it should be possible to estimate these parameters from the turnover tables. This can be done with the ML estimation procedure (Haberman, 1979) using the EM algorithm (Goodman, 1974a; 1974b; Hagenaars, 1993). The program LEM used in this study has been written by Vermunt (1995). The program uses turnover tables like the ones seen before as input. The user has to specify some mild restrictions on the probability matrices because otherwise the models of interest are not identified due to the fact that the number of unknown parameters is larger than the number of independent cells in the table.

The program (LEM) provides also a goodness of fit test for the whole model. The procedure will be illustrated below by an example.

The questions for which the analysis is done are a number of standard questions of the Eurobarometer which have been asked many times before and for which mode effects would be very troublesome. Therefore the following questions have been used:

1. Evaluation of membership of the EU

1.a Membership

Generally speaking, do you think that (our country's) membership of the EU is

a good thing / bad thing / good nor bad / DK/No answer

1.b Benefit

Taking everything into consideration, would you say that (our country) has on balance benefited or not from being a member of the (EU/EC) ?

benefited / not / DK/No answer

2. Satisfaction

2.a Life satisfaction

On the whole, are you very satisfied / fairly satisfied / not very satisfied / not at all satisfied with the life you lead ? Would you say you are ?

very satisfied / fairly satisfied / not very satisfied / not at all satisfied / DK/No answer

2.b Satisfaction with the way democracy works in (our country)

On the whole, are you very satisfied / fairly satisfied / not very satisfied / not at all satisfied with the way democracy works in (your country) ? Would you say you are ?

very satisfied/ fairly satisfied / not very satisfied / not at all satisfied / DK/No answer

3. Political interest

3.a Political discussion

When you get together with friends, would you say you discuss political matters frequently, occasionally, or never ?

frequently / occasionally / never / DK/No answer

3.b persuade others

When you hold a strong opinion, do you find yourself persuading your friends, relatives or fellow workers to share your views ? Does this happen ?

frequently / occasionally / never / DK/No answer

4. Media involvement

4.a Read newspapers

About how often do you read the news in daily newspapers ?

Every day / several times a week / once or twice a week / less often / never / DK/No answer

4.b Listen to radio

About how often do you listen to the news on the radio ?

Every day / several times a week / once or twice a week / less often / never / DK/No answer

4.c Watch TV

About how often do you watch the news on television ?

Every day / several times a week / once or twice a week / less often / never / DK/No answer

6.4 Results

The procedure which has been used for all questions will first be illustrated for one particular question. For this illustration the question concerning the frequency with which people try to persuade friends of political issues, was chosen. In table 6.3 the observed table from the French sample is given.[25]

Table 6.3 The table for the French sample with in the cells the frequencies of (absolute numbers) the answers for the different modes and within brackets the estimated frequencies on the basis of the latent class model with equal response probabilities.

Face to face	**Telephone**				
	Often	From time to time	Rarely	Never	Total
Often	13	14	1	1	29
	(13)	(12.5)	(2.0)	(1.5)	
From time to time	11	106	25	8	150
	(12.5)	(106)	(23.6)	(6.5)	
Rarely	3	22	24	23	72
	(2.0)	(23.5)	(24)	(18.0)	
Never	2	5	13	63	83
	(1.5)	(6.5)	(18.0)	(63)	
Total	29	147	63	95	334

On the basis of this table the response probabilities have to be estimated. This can be done with many different restrictions. The most interesting one in this case is the assumption that the response probabilities for the two modes are identical. In that case, using the model with equations (1) and (2) would mean that also the marginal distributions for the two modes have to be the same. The assumption of equal response probabilities can be specified in the model as

$$\Pi^f = \Pi^t \tag{5}$$

This assumption can in this case not be tested without further restrictions on the probabilities for identification reasons. The extra assumption made was that

$$\pi^f_{i,i+2}=\pi^f_{i+2,i} \text{ and } \pi^f_{i,i+3}=\pi^f_{i+3,i} \tag{6}$$

25 In this analysis we ignore the DK/No answer category because there are only a few cases and it complicates the analysis too much.

This assumption should also holds for the probabilities in telephone interviewing. These constraints concern probabilities which are very close to each other and also close to zero and therefore will have little effect on the fit of the model but help in the identification of the parameters.

With these restrictions and the assumptions in (5) the response probabilities have been estimated which were for both modes equal to the values in table 6.4.

Table 6.4 The estimated values for the response probabilities in France for the "Persuade" question using model (1) and (2) and assumptions (5) and (6).

Observed variable	**Latent variable**			
	Often	From time to time	Rarely	Never
Observed				
Often	.63	.04	.02	.01
From time to time	.34	.79	.10	.02
Rarely	.02	.15	.47	.01
Never	.01	.02	.41	.96
p^x	.10	.48	.27	.15

Applying equations (1) and (2) and using the results in table 6.4, one can compute the expected marginal distributions for the two modes. Using (4), the expected frequencies for table 6.3 can be obtained. The results are presented in brackets in the table. They show that in most cells the observed frequencies and expected frequencies do not deviate very much. Only in the cells (3,4) and (4,3) a larger difference emerges. This suggests a rather good fit of the model. As a formal test the likelihood ratio test is used for this purpose which gives in this case a value of $L^2 = 5.45$. With 4 degrees of freedom this test indicates that the model with the equality assumption (5) cannot be rejected. This result is rather remarkable because in the test with independent samples (chapter 5) it was found that there was a significant mode effect in France for this question. This test with panel data now suggests that the response probabilities might be equal and therefore the distributions of the observed variables will not differ more than by chance. This result is also surprising as the test, based on dependent samples, has more power than the test based on independent samples (Hagenaars, 1990). In chapter 5 this result is explained in the sense that a large part of the difference which was

detected before was due to the different characteristics of the fieldwork in the different studies and, only to a smaller extent, to the mode of data collection.

Table 6.5 The turnover table for the Belgian and Spanish sample with the cell frequencies (absolute numbers) of the answers for the different modes and within brackets the estimated frequencies on the basis of the latent class model with equal response probabilities.

			Belgium		
Face to face			**Telephone**		
	Often	From time to time	Rarely	Never	Total
Often	11	11	0	1	23
	(11.0)	(9.5)	(0.0)	(0.5)	
From time to time	8	66	17	11	102
	(9.5)	(66.0)	(19.5)	(7.0)	
Rarely	0	22	16	18	56
	(0.0)	(19.5)	(16.0)	(12.0)	
Never	0	3	6	25	34
	(0.5)	(7.0)	(12.0)	(25.0)	
Total	19	102	39	55	215

			Spain		
Face to face			**Telephone**		
	Often	From time to time	Rarely	Never	Total
Often	24	25	2	2	53
	(24.3)	(22.2)	(2.1)	(2.0)	
From time to time	20	59	26	8	113
	(22.2)	(60.1)	(22.1)	(5.6)	
Rarely	2	20	21	38	81
	(2.1)	(22.1)	(23.1)	(24.2)	
Never	2	3	13	38	56
	(2.0)	(5.6)	(24.2)	(39.2)	
Total	48	107	62	86	303

When the same model was also tested for Belgium and Spain, the likelihood ratio statistic was respectively L^2 = 13.6 and 16.9. With 4 degrees of freedom this means that the equality hypothesis has to be rejected for both countries. Table 6.5 presents the observed and expected frequencies for the two countries. Table 6.6 indicates again that the model fits the data rather well except, as in France, for the cells (3,4) and (4,3).

This finding seems to suggest that the probabilities for the categories 4 and 3 of the observed and latent variable have been constrained too much. Therefore the equality assumption (5) is now corrected by suggesting that the response probabilities π^{A}_{33} and π^{D}_{33} and π^{A}_{34} and π^{D}_{34} in each country do not have to be equal. This means that also π^{A}_{43} and π^{A}_{44} and also π^{D}_{43} and π^{B}_{44} can vary because the probabilities should add up to 1 for each column.

These two extra parameters were enough to obtain a very good fit for the model in each country. In France L^2 becomes 2.29, in Belgium 2.17 and in Spain 2.69. It follows that there are large differences between the modes in category 3 of the latent variable but not for category 4. This suggests that only the category 3 needs free parameters across modes. This turns out to be correct because the fit of the models does not change if the assumption is made that all probabilities in category 4 of the latent variable are the same for personal and telephone interviewing. The result of this analysis is therefore that people in category 3 of the latent variable behave differently when they get a personal interview or a telephone interview. The differences are indicated in table 6.6.

This table clearly indicates that there is quite a large change in response probability going from personal interviewing to telephone interviewing and that this change is in the same direction in all three countries: The probability to say “never” in a telephone interview increases considerably even though the probabilities are different in the different countries. Such a change in response probabilities can be an explanation for the significant differences which are found in the distributions of the responses in Belgium and Spain on this question for the different modes of data collection.

Table 6.6 The difference in reaction of respondents in category 3 of the latent variable in face to face (ftf) and telephone (tel.) interviews

Response categories	France ftf	France tel.	Belgium ftf	Belgium tel.	Spain ftf	Spain tel.
Often	.02	.02	.00	.00	.00	.00
From time to time	.14	.14	.37	.37	.11	.11
Rarely	.64	.46	.54	.32	.56	.19
Never	.20	.38	.10	.32	.34	.70

This analysis was then repeated for all questions: First the model with equal response probabilities is tested. If this model fits the data, the analysis stops. If the model does not fit, a less restricted model allowing in one column differences in probabilities between the modes is used. This approach is applied for one country, and the obtained model is then also tested for the other countries. If the corrected model does not work, a better and parsimonious alternative model is tried for the other country. Generally, obtaining an identical model for each country for the same questions was regarded as the most desirable solution, taking into

account that, however, this goal was given up in order to achieve the most parsimonious model. This means that a model with more parameters was not accepted if a model with less parameters turned out to be equally good, even if it did not hold up for all countries.

The results obtained with this approach are presented in table 6.7. This table shows that for several questions the model with equal probabilities did not fit. In these cases the mode of data collection had an effect on the response probabilities and consequently on the distribution of the answers in the different modes.

Table 6.7 The fitted models for 9 different questions

Question	Country	Equality model	Necessary parameter	Size of probability ftf		tel.	
Persuade others							
	France	accepted*	π^t_{33}, π^t_{43}	.64	.20	.46	.38
	Belgium	rejected	π^t_{33}, π^t_{43}	.53	.10	.32	.32
	Spain	rejected	π^t_{33}, π^t_{43}	.56	.34	.19	.70
Political discussion							
	France, Spain	accepted					
	Belgium	rejected	π^t_{23}, π^t_{33}	.24	.76	.00	.99
Benefit of country from EU membership							
	all	accepted					
Evaluation of EU membership							
	France	rejected	π^t_{12}, π^t_{22}	.34	.42	.19	.81
	Belgium	rejected	π^t_{12}, π^t_{22}	.36	.45	.18	.82
	Spain	rejected	π^t_{12}, π^t_{22}	.34	.42	.19	.81
Newspaper							
Radio/TV	all countries	accepted					
Satisfaction with democracy							
	France	rejected	π^t_{23}, π^t_{43}	.00	.30	.30	.00
	Belgium	rejected	π^t_{33}, π^t_{43}	.41	.51	.77	.15
	Spain	accepted					
Satisfaction with life							
	France	accepted					
	Belgium	rejected	π^t_{12}, π^t_{22}	.47	.53	.18	.82
	Spain	accepted					

* Although the model fitted , the correction made a significant improvement.

For the “persuade” question a systematic pattern was found that the people in the third class out of four latent classes had a tendency to say “never” more frequently in telephone interviews than in personal interviews. This pattern was found in all three countries, and the difference in response probabilities was quite large.

A similar phenomenon was discovered for the variable “membership”. In all countries the model with equal response probabilities was rejected. The reason seems to be the same in all three countries, namely that there is a tendency for the people with a middle position to express this middle position more frequently in telephone interviews than in personal interviews. These effects are very similar in all three countries and very large.

For the variables “political discussion” and “satisfaction with life” only in Belgium the model with equal probabilities had to be rejected and the necessary changes are also considerable.

For the variable “satisfaction with democracy” in France and Belgium significant differences in response probabilities have been found, but not in Spain, and the reason for these differences are also different.

Finally for the “benefit” question and the question about the frequency of looking or listening to the news, the model assuming equal response probabilities could not be rejected. So only for these 4 questions there is no problem of unequal distributions of the variables in the different data collection modes.

6.5 Conclusion

In this analysis the mode effect was studied by specifying a latent class model and testing whether the response probabilities for the respondents in a given latent class are the same for personal and for telephone interviewing. If that were the case, no mode effects should emerge. If differences occur, mode effects will be detected in form of differences between the responses in a telephone and in a personal interview.

This analysis has clearly indicated that for several standard Eurobarometer questions differences in the response probabilities occur at least in some countries for some questions. This suggests that at least a part of the total mode effects can be explained by this factor. For some questions these effects are the same in all countries, but for other questions these effects are different for different countries. This might have to do with the specific formulation and interpretation of the labels of the categories in the different countries.

An attractive feature of this methodological approach is that the response probabilities for the different classes give an impression whether the questions are interpreted in the same way in the various countries. If the response probabilities for the same question are very different one can doubt whether the questions have the same meaning for the respondents. It is at least questionable whether the responses can be compared because the people in the different countries interpret the questions apparently in a different way. Such differences can be seen in table 6.8 for the variables “persuade” and “satisfaction with democracy” where the response probabilities are very different for the different countries. This is, however, a different

problem than the mode problem we have dealt with before. In chapter 9 this problem will be discussed in greater detail.

Returning to the issue of this chapter, one can say that this analysis has given strong evidence that the mode of data collection can cause considerable differences in response distributions. This finding suggests that correction methods should be developed in order to make the results for the different interviewing modes comparable. This topic will be discussed in the last part of the book. In the next chapters, mode effects will be studied for other questions.