Journal of Official Statistics, Vol. 21, No. 3, 2005, pp. 367–387

Mixed Mode Designs: Finding the Balance Between Nonresponse Bias and Mode Effects

Robert J.J. Voogt[1] *and Willem E. Saris*[2]

Mixed mode designs in survey research can be used to lower nonresponse rates and prevent nonresponse bias. The main disadvantage is that the use of different modes of data collection can bias the results. By making a distinction between selection processes and transformation processes, it is possible to model the effects of a mixed mode design on the quality of survey data. Using data gathered on the occasion of the 1998 Dutch national elections, we show that mixed mode surveys and adding the possibility for nonrespondents to answer so called central questions, leads to higher response rates and to less nonresponse bias. Although we do find a higher level of response bias for the face-to-face mode as compared to the telephone and mail modes, the inclusion of the respondents that answered the face-to-face interview still leads to better estimates of the true population values than when these people are left out.

Key words: Mixed mode data collection; election research; survey research.

1. Introduction

In times of rising nonresponse rates in surveys, researchers are searching for methods that will help to keep nonresponse rates as low as possible. One possibility is to use different modes of data collection for different subgroups (Day et al. 1995; Groves and Kahn 1979; Hochstim and Athanasopoulos 1970; Shettle and Mooney 1999; Sudman and Bradburn 1982). The question is whether a higher response rate, reached by mixing different modes of data collection, will also lead to less biased data, as not only nonresponse bias, but also response bias can be a serious threat to the quality of survey data (Biemer and Lyberg 2003; Groves 1989). Using different modes of data collection might lead to higher levels of response bias, so it is even possible that, despite the higher response rate, the data will only get more biased.

To be able to better understand the possible advantages and disadvantages of using a mixed mode data collection design, we will use the distinction between selection processes and transformation processes, as explained in Saris (1997).

In selection processes, certain groups are "selected" (for example sampled or responding) with a smaller probability than other groups. Nonresponse is an example of a selection process. Using matrix algebra, if we denote the distribution of a certain

[1] University of Amsterdam, ASCoR, Klein Muiden 38, 1393 RL Nigtevecht, The Netherlands. Email: robertvoogt@gmail.com
[2] University of Amsterdam, Department of Sociology, Oudezijds Achterburgwal 237, 1012 DL Amsterdam, The Netherlands. Email: saris@pscw.uva.nl

dichotomous item $\boldsymbol{a}$ as $\mathbf{f}_{\boldsymbol{a}}$, and the distribution of item $\boldsymbol{a}$ among the final respondents as $\mathbf{f}_{\boldsymbol{ar}}$, we can map the nonresponse process as:

$$\begin{vmatrix} f_{ar}(1) \\ f_{ar}(2) \end{vmatrix} = \begin{vmatrix} \pi_{ar1} & 0 \\ 0 & \pi_{ar2} \end{vmatrix} \begin{vmatrix} f_{a}(1) \\ f_{a}(2) \end{vmatrix} \quad or \ \mathbf{f}_{\boldsymbol{ar}} = \mathbf{S}_{\boldsymbol{ar}} \cdot \mathbf{f}_{\boldsymbol{a}} \tag{1.1}$$

where π_{ark} is the probability that a sample element that has value k on item $\boldsymbol{a}$ will respond. In this notation, it is simple to see that nonresponse bias in item $\boldsymbol{a}$ will be absent if $\pi_{ar1} = \pi_{ar2}$, i.e., the probability of responding is equal for a sample element that has value 1 on item $\boldsymbol{a}$ and for a sample element that has value 2 on item $\boldsymbol{a}$. If $\pi_{ar1} \neq \pi_{ar2}$, nonresponse bias will be present.

What is characteristic for selection processes is that the off-diagonal elements in the selection matrices are always zero. So a difference in the distribution of a variable in the population and among respondents can only be caused by the fact that the probabilities of being sampled, i.e., of responding, are not equal for people that have different values on item $\boldsymbol{a}$. The advantage of mixed mode data collection is that nonrespondents can be reapproached with a different mode. If the selection processes are different for different modes, this can lead to less nonresponse bias. In this case, also the total bias will be lower, unless there are mode effects in the transformation processes, which may lead to higher levels of response bias.

In transformation processes, a transformation takes place from one kind of behavior or attitude to another. Answering a question (i.e., responding) is an example of a transformation process – behavior or attitudes are transformed into an answer to a question. The response process can be mapped, using matrix algebra, as

$$\begin{vmatrix} f_{avr}(1) \\ f_{avr}(2) \end{vmatrix} = \begin{vmatrix} \pi_{avr11} & \pi_{avr12} \\ \pi_{avr21} & \pi_{avr22} \end{vmatrix} \begin{vmatrix} f_{ar}(1) \\ f_{ar}(2) \end{vmatrix} \quad \text{or } \mathbf{f}_{\boldsymbol{avr}} = \mathbf{M}_{\boldsymbol{avr}} \cdot \mathbf{f}_{\boldsymbol{ar}} \tag{1.2}$$

If there is no response bias, $\pi_{avr11} = \pi_{avr22} = 1$, which automatically makes $\pi_{avr12} = \pi_{avr21} = 0$. This means each respondent reports the value of item $\boldsymbol{a}$ correctly. If each respondent does not report the value of the item correctly, $\pi_{av11} \neq \pi_{av22} \neq 1$ and $\pi_{av12} \neq \pi_{av21} \neq 0$, and response error does exist.

Mode effects can be the result of both selection processes and transformation processes. Mode effects as a result of selection processes exist when the probability that certain subgroups will respond is not equal for different interview modes. For example, when nonvoters are more inclined to participate in a face-to-face survey than in a telephone survey while voters are not, selection bias will result as a consequence of this mode effect. That these probabilities are indeed different is known from the survey literature (Biemer 2001; Day et al. 1995; De Leeuw and Van der Zouwen 1988; Goyder 1985; Hox and De Leeuw 1994). The mixed mode design is in fact developed from the knowledge that different modes lead to different levels of nonresponse for different groups. It is precisely these unequal response probabilities that make mixed mode data collection attractive.

Mode effects as the result of transformation processes exist when the probability that a respondent will report the true value of a certain item in the questionnaire is different for different modes of interview. When this happens, it is possible that differences in the data

are the consequences of the use of different modes of data collection and do not reflect true differences in the population.

That mode effects exist has been clearly shown in the survey literature (see, for example Biemer 2001; De Leeuw et al. 1996; Epstein et al. 2001; Fowler et al. 1998; Kalfs 1993; Saris and Hagenaars 1997; Schwarz et al. 1991; Sykes and Collins 1988). But it has also been shown that mode effects do not always exist (see, for example De Leeuw 1992; De Leeuw and Van der Zouwen 1988; Galobardes et al. 1998; Greenfield et al. 2000; Groves and Kahn 1979; Hochstim 1967; Reubrand and Blasius 1996). It has been argued that mode effects are strongest in surveys that include questions with so-called socially desirable answer categories (De Leeuw 1992; Presser and Stinson 1998; Tourangeau et al. 2000) or in the case of sensitive subjects (Rasinski et al. 1994; Sudman and Bradburn 1974; Tourangeau et al. 2000). In previous studies it has for example been found that in face-to-face surveys people tend to overreport behavior that is considered socially desirable (Presser and Stinson 1998; Sykes and Collins 1988). This is caused by the presence of an interviewer during the interview and the tendency respondents have to present themselves in a favorable way to this interviewer.

As part of a larger study of nonresponse bias in election research, we have carried out a survey in which three modes of data collection are combined: telephone interviewing, mail and face-to-face interviewing. The data set contained both reported voter turnout and validated voter turnout at the 1998 Dutch national elections. Validated voter turnout was available for both respondents and nonrespondents. This means that it is possible to distinguish between response and nonresponse bias in this data set. As we have made several conversion attempts, using different modes of interview, the data give us the opportunity to investigate whether higher response levels actually lead to less bias in the data, or whether mode effects neutralize the lower level of bias gained by diminishing the level of nonresponse.

Our study did not have an experimental design in which subgroups of respondents had been randomly divided over different interview modes. This means that it cannot be considered a pure mode comparison study. Although this can be seen as a shortcoming, the fact that we did not use an experimental design is also the strong point of the study. In this way, we were able to investigate, in a "real-life situation" (in which different subgroups of respondents are reached by different modes of interview), the consequences of using a mixed mode of data collection for both the level of nonresponse and response bias in the data set.

2. Sampling and Data Collection

2.1. The sample

Our data have been gathered as part of a study of nonresponse bias in election surveys. One of the central goals of this study was to explain the overestimation of voter turnout. This overestimation could be caused by nonresponse bias or by overreporting of voting by nonvoters or even by a stimulus effect. To be able to gain an insight in what exactly was causing this overestimation of voter turnout, it was necessary to know the validated voter turnout of the voters that were sampled.

This validated voter turnout can be obtained from election lists. On these lists, it is recorded when a voter turns up to cast a vote. In several studies these election lists have

been used to validate the reported voter turnout, for example in election studies in Sweden (Andersson and Granberg 1997; Granberg and Holmberg 1991), the United Kingdom (Collins and Sykes 1987) and the United States (Abramson and Claggett 1989; Belli et al. 2001), but not in The Netherlands. In The Netherlands, the local authorities (municipalities) are responsible for organizing the elections. It is the boards of the municipalities that should be asked for permission if a researcher wants to use these election lists. As there are almost 500 municipalities in Holland, it means that for a national election study, a very large number of municipal boards have to be contacted. As Holland has a very strict privacy legislation, it can be expected that many of these municipalities will not grant permission to use the election lists. Because of this, and because it was not necessary for our study to use a national sample, we decided to focus on one municipality. We chose the municipality of Zaanstad. Zaanstad has approximately 135,000 inhabitants, and consists of one big city (Zaandam), and six small villages.

We used a two-stage stratified sampling procedure. In the first stage we sampled ten out of the 63 districts that Zaanstad is divided into. To be sure the number of districts sampled from the main city of Zaandam, in which half of the inhabitants of Zaanstad are living, would be the same as the number of districts sampled from the six villages, we defined two strata, one containing all districts of the city, the other containing all districts of the villages. From each stratum, five districts were sampled. In the second stage we sampled in each district 100 people from the electoral register. (In one of the ten districts, only 95 people were sampled.) As the electoral register is part of the register of the population, we had a wide range of background characteristics of the sampled people at our disposal (for example date of birth, gender, address, household composition, and marital status).

2.2. *Data collection*

We used a mixed mode design to gather the data. We first distinguished between the sample elements with a registered telephone number ($N = 810$) and those without a telephone, with a disconnected telephone or with a unregistered telephone number ($N = 185$). The sample elements with a registered telephone number were contacted by telephone to be interviewed. Those for whom we did not have a telephone number were sent a mail questionnaire.

As surveys done by government institutions and universities are known to have higher response rates than surveys by other organizations (Fox et al. 1988; Heberlein and Baumgartner 1978; Hox and de Leeuw 1994), the interviewer in the telephone interview introduced him- or herself as an employee of the University of Amsterdam. He or she explained the purpose of the study and asked if the respondent was willing to answer a short five-minute questionnaire. When confronted with a refusal, the interviewer tried to persuade the respondent to change his or her mind, and when the respondent claimed to have no time to answer the interview, offered to call back at a more suitable moment. When the respondent stuck to his or her refusal, he or she was called back a few days later by a different interviewer. This interviewer introduced him- or herself as calling on behalf of the research management and again stressed the importance of the participation of the respondent in the survey. When the respondent still could not be persuaded to answer the questionnaire, the interviewer asked if he or she was willing to answer at least two short

questions, crucial to the study. These questions were whether the respondent had voted at the last national elections and whether the respondent was interested in politics.

Around the same time, the sample elements without a (known) telephone number were sent a mail questionnaire in a University of Amsterdam envelope, together with a letter signed by hand (personalized letters on official stationery are known to raise response rates: see Dillman 1978; Harvey 1987). In the signed letter, the purpose of the survey was explained, and the respondent was asked to help the researchers by filling in the questionnaire. A prepaid University of Amsterdam reply envelope was included to send back the questionnaire.

After approximately a month, both those not reached and the refusers from the telephone group and those who had been sent the mail questionnaire but did not return it, were sent a second mail questionnaire. In the accompanying letter the importance of the participation of the respondent was again stressed, and those who were not willing to fill in the questionnaire were asked to at least answer the two central questions that were printed on a strip at the end of the letter.

In an ultimate effort to raise the response level, interviewers visited all those who still had not responded. It was not inconceivable that people would get irritated at finding an interviewer on their doorstep, after having twice refused to participate by telephone and then having been bothered by a letter, that again asking for their cooperation in the same survey. So we decided to first wait a few months before sending out the interviewers. We expected that most people would have forgotten being asked earlier to participate in the survey, and those who still remembered would probably be less irritated than if they had been recontacted by an interviewer at an earlier time.

We instructed the interviewers to explain to the respondents the importance of their participation in the study, and to ask the people that kept refusing to at least answer the two central questions. When a respondent was not found at home, the interviewers paid two more visits, at different times of the day.

Table 1 shows the final response results.

3. Results

3.1. The selection process

The selection processes can be represented in a tree diagram, which is shown in Figure 1. In the diagram, we see that each node leads to two possible branches. The two end branches of each bough represent the long interview or the central questions.

It is possible to draw some general conclusions from this tree diagram.

Table 1. Final response results

	Interview	Central questions	Nonresponse	Total
Telephone	49.3 (491)	8.5 (85)		57.9 (576)
Mail	10.2 (101)	2.7 (27)		12.9 (128)
Face-to-face	18.8 (187)	3.7 (37)		22.5 (224)
Nonresponse			6.7 (67)	6.7 (67)
Total	78.3 (779)	15.0 (149)	6.7 (67)	995

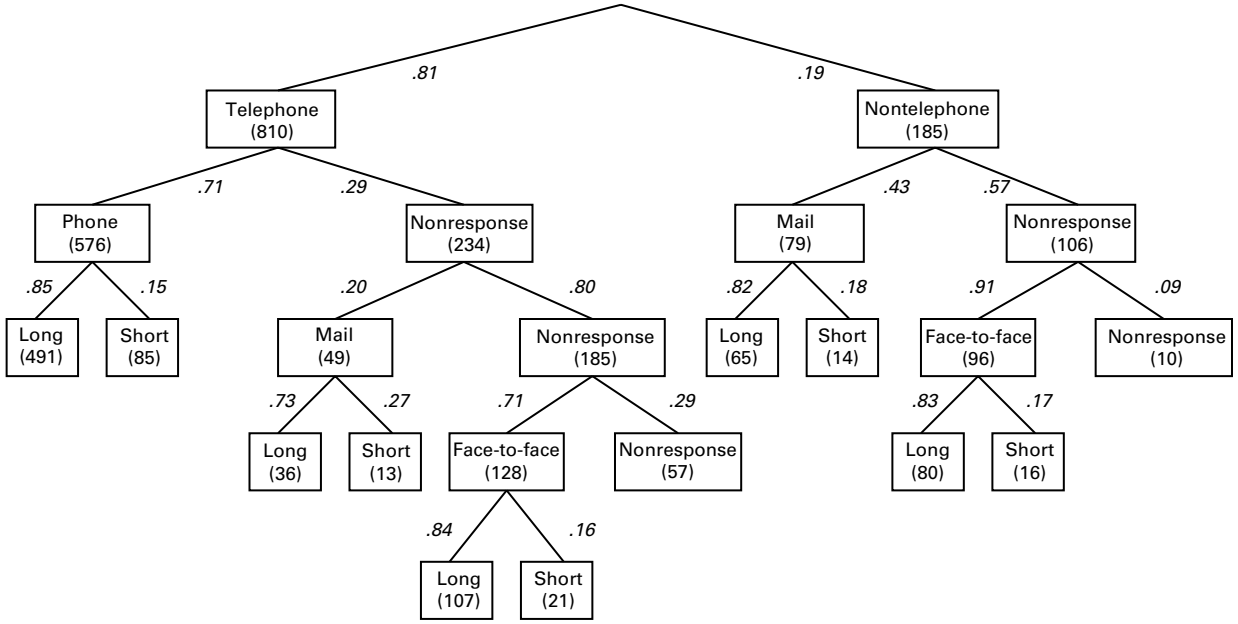

Fig. 1. A tree diagram of the selection processes that have been caused by the nonresponse mechanism

First, with each new effort to get the nonrespondents to respond, a considerable proportion of the nonrespondents do respond. This means that recontacting respondents through a different interview mode really pays off by substantially raising the response level.

Second, as the group of nonrespondents is getting smaller with each new node, the further the node is from the root, the more the group of nonrespondents will represent the hardcore of the nonrespondents. So it can be expected, ignoring the effect of the mode of interview, that the response rate will be lower at each new node of the tree diagram. From this diagram, we see that this is not true for all nodes. For those that are pictured on the left side of the diagram (the people with a listed telephone number), the response rate is 71% for those interviewed by phone. The response level is much lower, only 20%, for those that were mailed a questionnaire and again 71% for the face-to-face interview. For the group pictured on the right side of the tree diagram (those who could not be phoned), the response level is 43% for the mail questionnaire and much larger, 91%, for the face-to-face interview. So it seems that of the three interview modes, other things being equal, the face-to-face interview will lead to the highest response rate, followed by the telephone interview, while the mail interview will have the lowest response rate. This is in line with results reported in previous research (de Leeuw and van der Zouwen 1988; Hox and de Leeuw 1994; Sykes and Collins 1988).

A third striking result is that, irrespective of the group or the interview mode, the proportion of the respondents that are willing to answer the long questionnaire is always between .82 and .85, with only one, rather small, exception of .73 for those with a listed telephone number who responded to the mail questionnaire. In other words, giving the nonrespondents the possibility of answering only two central questions instead of the longer questionnaire will lead to extra information about a group of nonrespondents that is as large as 20% of the respondents.

We know now that using three different modes of interview and giving nonrespondents the opportunity to answer two central questions will lead to a lower level of nonresponse. The next question we have to answer is whether less nonresponse also means less nonresponse bias. In Table 2 we have pictured the selection matrices for the three modes of interview we have used, while we also have distinguished between the respondents who answered the long questionnaire and those who only answered the two central questions – for both the group that could be contacted by phone and those who could not be contacted by phone.

The results in Table 2 show that there is indeed a difference in response behavior of voters and nonvoters. The selection matrices show that, with the exception of the face-to-face mode for the group that could not be reached by phone, the voters have a larger probability of answering the long questionnaire, while the nonvoters have a larger probability of answering the central questions. This means that adding the possibility of answering central questions to the data collection design will lead to less nonresponse bias. It is also clear from the table that with each new data collection method that is used, one delves deeper into the group of nonvoters, as the voter turnout is highest among the people that answered the telephone interview, and lowest among those that answered the face-to-face interview. So it seems that using different modes of data collection leads to less nonresponse bias in voter turnout.

Table 2. Selection matrices mapping the nonresponse process, for three different modes of interview (telephone, mail and face-to-face) and two different questionnaires (long interview and only central questions) for both the group that could be reached by telephone (because of having a listed telephone number) and the group that could not be reached by telephone (because of having an unlisted telephone number or no telephone), using the variable "voted at the 1998 national elections." In the last column the validated number of voters and nonvoters in each of the subgroups can be found. Below the selection matrices, the voter turnout in the subgroup is shown

		Long interview	Central questions	*N*
Telephone	Voters	$\begin{vmatrix} 0.879 & 0 \\ 0 & 0.724 \end{vmatrix}$	$\begin{vmatrix} 0.121 & 0 \\ 0 & 0.276 \end{vmatrix}$	478
	Nonvoters			98
		(85.5)	(68.2)	
Mail (listed numbers)	Voters	$\begin{vmatrix} 0.794 & 0 \\ 0 & 0.600 \end{vmatrix}$	$\begin{vmatrix} 0.206 & 0 \\ 0 & 0.400 \end{vmatrix}$	34
	Nonvoters			15
		(75.0)	(53.8)	
Mail (unlisted/no telephone)	Voters	$\begin{vmatrix} 0.827 & 0 \\ 0 & 0.815 \end{vmatrix}$	$\begin{vmatrix} 0.173 & 0 \\ 0 & 0.185 \end{vmatrix}$	52
	Nonvoters			27
		(66.2)	(64.3)	
Face-to-face (listed)	Voters	$\begin{vmatrix} 0.882 & 0 \\ 0 & 0.769 \end{vmatrix}$	$\begin{vmatrix} 0.118 & 0 \\ 0 & 0.231 \end{vmatrix}$	76
	Nonvoters			52
		(62.6)	(42.9)	
Face-to-face (unlisted/no telephone)	Voters	$\begin{vmatrix} 0.776 & 0 \\ 0 & 0.894 \end{vmatrix}$	$\begin{vmatrix} 0.224 & 0 \\ 0 & 0.106 \end{vmatrix}$	49
	Nonvoters			47
		(47.5)	(68.8)	

Table 3. Fit of four loglinear models including mode of interview (M), length of questionnaire (L) and voter turnout (V)

Model	*df*	*X2*	*Prob.*
M*L + V	5	78.4701	0.00
M*L + L*V	3	67.6136	0.00
M*L + M*V	3	18.3922	0.00
M*L + M*V + L*V	2	2.2980	0.32

To test whether the table represents significant differences, we did a loglinear analysis, using the program LEM (Vermunt 1997). We included three variables in our model: mode of interview (M), length of the questionnaire (L) and validated voter turnout, as registered on the election lists (V). The relation between mode of interview and length of the questionnaire on the one hand and validated voter turnout on the other hand is our main interest, so we will take the correlation between mode of interview and length of the questionnaire (M∗L) for granted. If our hypothesis, that using different modes of interview and giving nonrespondents the possibility of answering a short questionnaire will lead to less nonresponse bias, is correct, we expect that a model with the inclusion of the factors M∗V (the correlation between mode of interview and validated voter turnout) and L∗V (the correlation between length of the questionnaire and validated voter turnout) will fit, and models without these factors will not.

In Table 3, we have fitted four possible models, and we see that it is indeed the case that inclusion of the factors M∗V and L∗V is necessary to arrive at a model with a satisfactory fit. In Table 4, the parameters of the model with the most satisfactory fit are shown.[2] We have expressed each effect in terms of deviations from the average effect and we used the restriction that each parameter summed over any of its subscripts equals zero, as is also common in analysis of variance (Hagenaars 1993).

The results in Table 4 indicate that the odds of finding a voter among the people that respond to the telephone questionnaire are greater than the odds of finding a voter among the people that respond to the mail questionnaire, while the odds of finding a voter among the people that respond to the mail questionnaire is greater than the odds of finding a voter among the people that respond to the face-to-face questionnaire. These differences are statistically significant at the 5% level. Table 4 shows that voters more often answer the long questionnaire than nonvoters do: the partial odds of finding a voter among the people that responded to the long questionnaire are (1.2153/.8228)/(.8228/1.2153) = 2.18 times greater than finding a voter among the people that have answered the central questions. Also this difference is statistically significant.

3.2. The transformation process

To be able to find out whether this higher response rate reached by combining various modes of interview will also lead to a better estimation of the true level of voter turnout, we need to take both selection and transformation processes into account. If using different

[2]We also repeated this analysis, excluding the group that could not be reached by phone. This did not lead to different results.

*Table 4. Parameter values for the model M*L + M*V + L*V*

	λ	s_λ	τ	*Waldorf*	*Degrees of freedom*	*Probability*
*Method * Length*						
Telephone – short	0.0074	0.0669	1.0074			
Telephone – long	−0.0074		0.9927			
Mail – short	0.1718	0.0836	1.1875			
Mail – long	−0.1718		0.8421			
Face-to-face – short	−0.1792	0.0836	0.8359			
Face-to-face – long	0.1792		1.1963	5.25	2	0.07
*Method * Voter turnout*						
Telephone – voted	0.3703	0.0545	1.4482			
Telephone – not voted	−0.3703		0.6905			
Mail – voted	−0.0411	0.0706	0.9598			
Mail – not voted	0.0411		1.0419			
Face-to-face – voted	−0.3292	0.0706	0.7195			
Face-to-face – not voted	0.3292		1.3899	61.06	2	0.00
*Length * Voter turnout*						
Short – voted	−0.1950	0.0513	0.8228			
Short – not voted	0.1950		1.2153			
Long – voted	0.1950		1.2153			
Long – not voted	−0.1950		0.8228	14.44	1	0.00

modes of interview and differences in the length of the questionnaire do not cause any differences in the transformation processes, we can conclude that using a mixed mode of data collection leads to less selection bias without any negative consequences for the comparability of the answers of the respondents interviewed by different methods. If there are differences, we need to find out whether the final results are better or worse when the data gathered with different modes are included in the analysis, because the profit of less nonresponse bias gained using a mixed mode design could be cancelled out by the loss in response quality as a result in differences in transformation processes caused by mode effects.

In Table 5, the transformation matrices are pictured for the three modes of interview we have used, for the respondents who answered the long questionnaire and those who only answered the two central questions for both the group with a listed telephone number and those with an unlisted number or without a telephone. We see that the six transformation matrices that describe the transformation process of both the long questionnaires and the central questions for the telephone and mail mode do not differ that much from each other. Almost all voters report correctly that they have voted, while (with one exception) around 20% of the nonvoters incorrectly claim they have voted. The transformation matrices of the face-to-face mode clearly deviate from the matrices of the other two modes. The diagonal probabilities are much smaller and the off-diagonal probabilities are much larger for the face-to-face mode than for the other two modes, for the transformation matrices of the central questions even more strongly so than for those of the long questionnaire. So it seems that the response bias is larger for the face-to-face mode than for the other two modes.

To test whether this really is the case, we have added the reported voter turnout (R) to the loglinear model we have tested in the previous paragraph. The transformation process can be expressed as V*R, i.e., the transformation from the true voter turnout to the reported voter turnout. This factor will be included in all models that we will test, as a dependent factor. We start by including the M*L*V factor – as we are interested in the transformation process, we take the selection process for granted in all models – and we add the V*R factor. From the first row of Table 6, it is very clear that this two-factor model does not fit.

In the second, third and fourth rows of Table 6, we have added the factors M*R (second row), L*R (third row) and both (fourth row) to the initial model. We did not include the validated voter turnout V in those factors. In other words: the mode of the interview and the length of the questionnaire are expected to have an effect on the reported voter turnout (R), but not on the relation between the reported and validated voter turnout (V*R), so not on the transformation process. These three models do not fit either, which leaves us no other choice than to include a factor in the model that contains mode (M) or length (L) and the factor of reported and validated voter turnout (V*R). This means that the hypothesis that the mode or length of the interview does not have an effect on the transformation process has to be rejected. In the fifth row we have included a factor with mode of interview, validated voter turnout and reported voted turnout (M*V*R), and in the sixth row we have included a factor with length of the interview, validated voter turnout and reported voter turnout (L*V*R). Neither model fitted, but we see that the inclusion of the M*V*R factor leads to a large improvement in fit, while the inclusion of the L*V*R factor

Table 5. Transformation matrices mapping the response process, for three different modes of interview (telephone, mail and face-to-face) and two different questionnaires (long interview and only central questions) for both the group that could be reached by telephone (because of having a listed telephone number) and the group that could not be reached by telephone (because of having an unlisted telephone number or no telephone), using the variable "voted at the 1998 national elections." In the last column the reported number of voters and nonvoters in each of the subgroups can be found

		Long interview		Central questions		*N*
Telephone	Voters	0.998	0.211	0.964	0.185	493
	Nonvoters	0.002	0.789	0.036	0.815	81
Mail (listed numbers)	Voters	0.963	0.222	1	0	35
	Nonvoters	0.037	0.778	0	1	14
Mail (unlisted/no telephone)	Voters	1	0.227	1	0.200	58
	Nonvoters	0	0.773	0	0.800	21
Face-to-face (listed)	Voters	0.758	0.432	0.200	0.182	66
	Nonvoters	0.242	0.568	0.800	0.818	49
Face-to-face (unlisted/no telephone)	Voters	0.658	0.436	0.286	0	44
	Nonvoters	0.342	0.564	0.714	1	43

Table 6. Fit of six loglinear models including mode of interview (M), length of questionnaire (L), validated voter turnout (V) and reported voter turnout (R)

Model	*df*	*X2*	*Prob.*
M*L*V + V*R	10	228.9775	0.00
M*L*V + M*R + V*R	8	141.5740	0.00
M*L*V + L*R + V*R	9	188.7034	0.00
M*L*V + M*R + L*R + V*R	7	112.9595	0.00
M*L*V + M*V*R	6	24.9592	0.00
M*L*V + L*V*R	6	188.7805	0.00
M*L*V + L*R + M*V*R	5	7.8837	0.16
M*L*V + L*V*R + M*V*R	4	4.1973	0.38

does not. In the seventh row, we kept the M∗V∗R factor and added the L∗R factor. This model does fit very well. Replacing the factor L∗R by the factor L∗V∗R, as is done in the eight row, does not lead to a significantly better fit (the improvement in the Pearson chi-square is 3.6864; with 1 degree of freedom this means that there is no significant improvement of the fit of the model), so we can leave the factor L∗V∗R out. The parameters of the model in the seventh row of Table 6 are shown in Table 7.[3]

The factor that is of most interest to us is the M∗V∗R factor. When we take a closer look at the parameters of this factor, it turns out that the transformation processes of the telephone or the mail mode do not differ that much from each other, especially when they are compared to the face-to-face mode. This is in line with what we already had concluded about the differences between the transformation matrices of the three modes of interview in Table 5. The conditional odds ratio of validated voter turnout and reported voter turnout for the respondents interviewed by telephone is only 1.5 times greater than for the respondents that answered the mail questionnaire – a difference that is not statistically significant. The conditional odds ratio of validated voter turnout and reported voter turnout for respondents that were interviewed face-to face is 132 times smaller than for those interviewed by phone and as much as 196 times smaller for those who answered the mail questionnaire – both very large and statistically significant differences. This leads to the conclusion that only the face-to-face interview mode leads to a deviating transformation process.[4]

3.3. The trade-off between nonresponse bias and response bias

Now that we have seen that interview mode has an effect on the transformation process, although only for the face-to-face mode, we will turn to the last question that has to be answered. Does using mixed mode data collection, including the possibility for nonrespondents to answer a short questionnaire, lead to less biased results, despite the fact

[3]We also re-did this analysis, excluding the group that could not be reached by phone. This did not lead to different results.

[4]We also tested a model that included only the respondents of the telephone and mail questionnaires, leaving out the respondents that answered the face-to-face questionnaire. Without having to include the M∗V∗R-effect, we found a model with a satisfactory fit. This result also clearly shows that there is no mode effect for the telephone and mail modes of interview.

Table 7. Parameter values for the model M∗L∗V + L∗R + M∗V∗R

	λ	s_λ	τ	*Waldorf*	*Degrees of freedom*	*Probability*
*Mode *Length * Validated voter turnout*						
Telephone – short – voted	−0.0428	0.0714	0.9581			
Telephone – short – not voted	0.0428		1.0437			
Telephone – long – voted	0.0428		1.0437			
Telephone – long – not voted	−0.0428		0.9581			
Mail – short – voted	0.1114	0.0886	1.1178			
Mail – short – not voted	−0.1114		0.8946			
Mail – long – voted	−0.1114		0.8946			
Mail – long – not voted	0.1114		1.1178			
Face-to-face – short – voted	−0.0686	0.0886	0.9337			
Face-to-face – short – not voted	0.0686		1.0710			
Face-to-face – long – voted	0.0686		1.0710			
Face-to-face – long – not voted	−0.0686		0.9337	1.61	2	0.45
*Length * Reported voter turnout*						
Short – voted	−0.3165	0.0898	0.7287			
Short –not voted	0.3165		1.3723			
Long – voted	0.3165		1.3723			
Long –not voted	−0.3165		0.7287	12.42	1	0.00
*Mode * Validated voter turnout * Reported voter turnout*						
Telephone – voted – voted	0.4829	0.1406	1.6208			
Telephone – voted – not voted	−0.4829		0.6170			
Telephone – not voted – voted	−0.4829		0.6170			
Telephone – not voted – not voted	0.4829		1.6208			
Mail - voted – voted	0.3499	0.1891	1.4190			
Mail – voted – not voted	−0.3499		0.7047			
Mail – not voted – voted	−0.3499		0.7047			
Mail – not voted – not voted	0.3499		1.4190			

Table 7. Continued

	λ	s_λ	τ	*Waldorf*	*Degrees of freedom*	*Probability*
Face-to-face – voted – voted	−0.8328	0.1891	0.4348			
Face-to-face – voted – not voted	0.8328		2.2998			
Face-to-face – not voted – voted	0.8328		2.2998			
Face-to-face – not voted – not voted	−0.8328		0.4348	68.64	2	0.00

that transformation processes differ as a result of mode effects? The results in Table 8 answer this question.

In Table 8 we see that using more than one mode of interview leads to a voter turnout that is closer to the true voter turnout in the sample (so in the population), and adding the possibility of answering the central questions leads to a further improvement. The improvement is mainly caused by the fact that with every new mode of interview that is added to the data collection design, the selection bias diminishes, while the (minor) increase of the transformation bias (which is strongest when all three methods are used, without the possibility for nonrespondents to answer the short questionnaire) is not strong enough to cancel out this improvement. The results in Table 8a also show that giving the nonrespondents the possibility of answering the central questions does not lead to a higher level of transformation bias, while it does lead to a substantially lower level of selection bias.

In Table 8b the consequences of using different modes of interview and the possibility of answering a shorter version of the questionnaire for the correlations between the voter turnout at the parliamentary elections and the validated voter turnout at the provincial elections can be seen. In the first cell of the second row of the table the correlation in the complete sample is pictured: .41.

The results in Table 8b show that including the respondents who answered the central questions diminishes the selection bias. Second, adding the possibility of answering the central questions does not have any substantial effect on the amount of transformation bias, as can be read from the comparison of the figures in the third and sixth columns in the table. Third, the transformation bias is largest when all three modes of interview are combined. Fourth, the correlation between voter turnout at the national elections and at the provincial elections is underestimated for all scenarios. This means that the correlation between voter turnout at the national elections and at the provincial elections gets weaker as a consequence of the measurement errors in voter turnout at the national elections. This explains why the correlation between the two variables becomes smaller when the face-to-face respondents are added, as the transformation bias (i.e., the measurement error) is largest for this group of respondents. This finding is in line with the result of the loglinear analysis in the previous paragraph.

Summarizing the results in Table 8 we conclude that, to get as good an estimate as possible of the distribution of voter turnout in the population, we need to include all three modes of interview and the possibility for the respondents to answer the central questions. If we had been mainly interested in the relationship between voter turnout and other variables of interest, a telephone interview – possibly followed up by sending out a mail questionnaire – combined with the possibility for nonrespondents to answer a short questionnaire, would have been sufficient.

4. Discussion

In this article we set out to answer the question what the consequences are of using a combination of different modes of data collection in survey research as a means to maximize the response level and thus improve the quality of the data. Our goal was to find out whether mixed mode data collection really leads to less nonresponse bias in the data

Table 8. Bias in voter turnout and bias in the correlation of voter turnout and voter turnout at the provincial elections and political interest for each of six scenarios

8a. Selection and response bias in the voter turnout at the national elections in each of six scenarios

Validated voter turnout at the national elections in the total *sample*	*Reported* voter turnout at the national elections *among the respondents*					
72.9	*only the long questionnaire*			*both the long and the short questionnaire*		
	selection bias	*response bias*	*reported turnout*	*selection bias*	*response bias*	*reported turnout*
Mode of data collection:						
Only telephone	+12.6	+2.9	88.4	+10.1	+2.9	85.9
Telephone and mail	+10.7	+2.5	86.1	+7.2	+3.4	83.5
Telephone, mail and face-to-face	+3.5	+3.7	80.1	+1.3	+2.8	77.0

8b. Selection and response bias in the correlation of voter turnout at the national elections and voter turnout at the provincial elections for each of six scenarios

Correlation of *validated voter* turnout at the national elections and *validated voter* turnout at the provincial elections *in the total sample*	Correlation of *reported* voter turnout at the national elections and *validated voter* turnout at the provincial elections *among the respondents*					
0.41	*only the long questionnaire*			*both the long and the short questionnaire*		
	selection bias	*response bias*	*reported correlation*	*selection bias*	*response bias*	*reported correlation*
Mode of data collection:						
Only telephone	−0.06	−0.05	0.30	−0.02	−0.05	0.34
Telephone and mail	−0.06	−0.04	0.31	−0.02	−0.04	0.35
Telephone, mail and face-to-face	−0.04	−0.07	0.30	−0.01	−0.06	0.34

and, if so, if this advantage is not neutralized by larger levels of response bias caused by mode effects as a consequence of combining different interview modes. We were able to answer these questions, as we had a data set at our disposal that made it possible to gain more insight into the consequences of the data collection design with regard to the quality of the data than one usually can get.

We started out by demonstrating that using different modes of data collection leads to lower levels of nonresponse. We showed that subgroups that responded on different modes of interview differed with respect to the main subject of the study – in our case, voter turnout. We also demonstrated how this nonresponse bias was further diminished by giving nonrespondents the possibility of answering two central questions (about their voter turnout and political interest).

Our results showed that there was response bias in our data. This response bias led to a further overestimation of the reported voter turnout. We also found a difference in the level of response bias in the face-to-face interview on the one hand and the telephone and mail interviews on the other hand. So far, we have not discussed what could have caused this difference, and whether this difference is really the result of a mode effect.

Before we are able to answer this question, it is first important to realize what kind of bias one would expect in our variable of interest, voter turnout, as the result of using a face-to-face mode of data collection. It is known that in a democracy voting is, in general, seen as some kind of a civic duty, so as a socially desirable act (Belli et al. 2001; Bernstein et al. 2001; Harbaugh 1996; Presser 1990). As we have noted above, people tend to overreport social desirable behavior in face-to-face surveys. This means that in our study, we would expect that in a face-to-face survey more nonvoters will claim to have voted and fewer voters will claim to not to have voted, i.e., that the odds ratio for a voter to report to have not voted versus a nonvoter to have voted, should be greater for respondents that participated in the face-to-face interview than for those that answered the telephone or the mail questionnaire. However, this is not supported at all by our data. This odds ratio among the face-to-face respondents is only 1.9, while the odds ratio among the telephone respondents is 24.0 and among the mail respondents as much as 113.6. This means that the nonvoters interviewed face-to-face are *less* often reporting having voted than the voters reporting not having voted.

So it is unlikely that the different transformation process for the face-to-face interview as compared to the other two interview modes that is found in our study is a direct consequence of the mode of interview. It is caused by something else. One other aspect, besides the mode of interview, that sets the group interviewed face-to-face apart from the other respondents is that the interviews with the former took place a few months later than those with the latter.

It is not inconceivable that part of the misreporting of the voter turnout has been caused by memory effects. This also explains why the response bias for the telephone and mail groups goes mainly in one direction (i.e., overreporting of voting), while for the face-to-face group the misreporting by both the voters and nonvoters increases, but the relative increase of misreporting among the voters (i.e., many more voters claim to have not voted) is much larger than that among the nonvoters. This is also in line with results reported by Belli et al. (1999), who found that an experimental question they used to reduce overreporting by memory failure did work much better for people interviewed later in the

data collection period. The fact that many of the people interviewed face-to-face are not interested in politics in the first place, as they belong to the group that was most difficult to persuade to participate in the survey, will reinforce these memory effects, as it is known that memory effects will be stronger for respondents that have no interest in the topic of the survey (Schwarz and Sudman 1994). When the follow-up of the nonrespondents with a different interview mode is done within a short interval, memory effects can be prevented. Still, it is important to keep in mind that there should be some time between the refusal and the following up of these refusers, not to run the risk that these people will get irritated when they are contacted again.

Our results show that mixed mode surveys and adding the possibility for nonrespondents to answer so-called central questions will lead to larger response rates and to less nonresponse bias. Although using different modes of interview brought some response bias with it, the total bias was still lower than if only one mode of interview had been used. All in all, we conclude that a mixed mode design is an efficient way of fighting bias in survey research.

5. References

Abramson, P.R. and Claggett, W. (1989). Race-Related Differences in Self-Reported and Validated Turnout in 1986. Journal of Politics, 51, 397–408.

Andersson, H.E. and Granberg, D. (1997). On the Validity and Reliability of Self-Reported Vote: Validity without Reliability? Quality and Quantity, 31, 127–140.

Belli, R.F., Traugott, M.W., Young, M., and McGonagle, K.A. (1999). Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring. Public Opinion Quarterly, 63, 90–108.

Belli, R.F., Traugott, M.W., and Beckmann, M.N. (2001). What Leads to Voting Overreports? Contrast of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies. Journal of Official Statistics, 17, 479–498.

Bernstein, R., Chadha, A., and Montjoy, R. (2001). Overreporting Voting. Why It Happens and Why It Matters. Public Opinion Quarterly, 65, 22–44.

Biemer, P.P. (2001). Nonresponse Bias and Measurement Bias in a Comparison of Face to Face and Telephone Interviewing. Journal of Official Statistics, 17, 295–320.

Biemer, P.P. and Lyberg, L. (2003). Introduction to Survey Quality. New York: John Wiley.

Collins, M. and Sykes, W. (1987). The Problems of Non-coverage and Unlisted Numbers in Telephone Surveys in Britain. Journal of the Royal Statistical Society, Series A, 150, 241–253.

Day, N.A., Dunt, D.R., and Day, S. (1995). Maximizing Response to Surveys in Health-Program Evaluation at Minimum-Cost Using Multiple Methods. Evaluation Review, 19, 436–450.

De Leeuw, E.D. (1992). Data Quality in Mail. Telephone and Face-to-Face Surveys. Amsterdam: Vrije Universiteit.

De Leeuw, E.D., Mellenbergh, G.J., and Hox, J.J. (1996). The Influence of Data Collection Method and Structural Models: A Comparison of a Mail, a Telephone and a Face-to-Face Survey. Sociological Methods and Research, 24, 443–472.

De Leeuw, E.D. and van der Zouwen, J. (1988). Data Quality in Telephone and Face to Face Surveys: A Comparative Meta-Analysis. In Telephone Survey Methodology, R.M. Groves, P.P. Biemer, L. Lyberg, J. Massey, W.L. Nicholls, II and J. Waksberg (eds). New York: John Wiley, 283–299.

Dillman, D. (1978). Mail and Telephone Surveys. The Total Design Method. New York: Wiley.

Epstein, J.F., Barker, P.R., and Kroutil, L.A. (2001). Mode Effects in Self-Reported Mental Health Data. Public Opinion Quarterly, 65, 529–549.

Fowler, F.J., Roman, A.M., and Di, Z.X. (1998). Mode Effects in a Survey of Medicare Prostate Surgery Patients. Public Opinion Quarterly, 62, 29–46.

Fox, R.J., Crask, M.R., and Kim, J. (1988). Mail Survey Response Rate; A Meta-Analysis of Selected Techniques for Inducing Response. Public Opinion Quarterly, 52, 467–491.

Galobardes, B., Sunyer, J., Anto, J.M., Castellsague, J., Soriano, J.B., and Tobias, A. (1998). Effect of the Method of Administration, Mail or Telephone, on the Validity and Reliability of a Respiratory Health Questionnaire. The Spanish Centers of the European Asthma Study. Journal of Clinical Epidemiology, 51, 875–881.

Goyder, J.C. (1985). Face-to-Face Interviews and Mail Questionnaires: The Net Difference in Response Rates. Public Opinion Quarterly, 49, 234–252.

Granberg, D. and Homberg, S. (1991). Self-Reported Turnout and Voter Validation. American Journal of Political Science, 35, 448–459.

Greenfield, T.K., Midanik, L.T., and Rogers, J.D. (2000). Effects of Telephone versus Face-to-Face Interview Modes on Reports of Alcohol Consumption. Addiction, 95, 277–284.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: John Wiley.

Groves, R.M. and Kahn, R.L. (1979). Surveys by Telephone: A National Comparison with Personal Interviews. New York: Academic Press.

Hagenaars, J.A. (1993). Loglinear Models with Latent Variables. Newbury Park: Sage.

Harbaugh, W.T. (1996). If People Vote Because They Like to, Then Why Do so Many of Them Lie? Public Choice, 89, 63–76.

Harvey, L. (1987). Factors Effecting Response Rates to Mailed Questionnaires: A Comprehensive Literature Review. Journal of the Market Research Society, 29, 341–353.

Heberlein, T.A. and Baumgartner, R.M. (1978). Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature. American Sociological Review, 43, 447–462.

Hochstim, J.R. (1967). A Critical Comparison of Three Strategies of Collecting Data from Households. Journal of the American Statistical Association, 62, 976–989.

Hochstim, J.R. and Athanasopoulos, D.A (1970). Personal Follow-up in a Mail Survey: Its Contribution and Its Cost. Public Opinion Quarterly, 34, 69–81.

Hox, J.J. and De Leeuw, E. (1994). A Comparison of Nonresponse in Mail, Telephone, and Face-to-Face Surveys – Applying Multilevel Modeling to Metaanalysis. Quality and Quantity, 28, 329–344.

Kalfs, N. (1993). Hour By Hour. Effects of the Data Collection Mode in Time Use Research. Amsterdam: NIMMO.

Presser, S. (1990). Can Context Change Reduce Vote Overreporting ? Public Opinion Quarterly, 54, 586–593.

Presser, S. and Stinson, L. (1998). Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance. American Sociological Review, 63, 137–145.

Rasinski, K.A., Mingay, D., and Bradburn, N.M. (1994). Do Respondents Really "Mark All That Apply" on Self-Administred Questions? Public Opinion Quarterly, 58, 400–408.

Reubrand, K.H. and Blassius, J. (1996). Face-to-Face, Telefonische und Postalische Befragungen. Ausschöpfungsquoten und Antwortmuster in einer Großstadt-Studie. Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie, 48, 296–318.

Saris, W.E. (1997). Adjustment for Differences between Face to Face and Telephone Interviews. In Eurobarometer – Measurement Instruments For Opinions in Europe W.E.Saris and M.Kaase (eds). Mannheim: ZUMA, 141–156.

Saris, W.E. and Hagenaars, J.A. (1997). Mode Effects in the Standard Eurobarometer Questions. In Eurobarometer – Measurement Instruments For Opinions in Europe W.E. Saris and M. Kaase (eds). Mannheim: ZUMA, 87–100.

Schwarz, N., Strack, F., Hippler, H.J., and Bishop, G. (1991). The Impact of Administrative Mode on Response Effects in Survey Measurement. Applied Cognitive Psychology, 5, 193–212.

Schwarz, N. and Sudman, S. (1994). Autobiographical Memory and the Validity of Retrospective Reports. New York: Springer Verlag.

Shettle, C. and Mooney, G. (1999). Monetary Incentives in Government Surveys. Journal of Official Statistics, 15, 231–250.

Sudman, S. and Bradburn, N. (1974). Response Effects in Surveys: A Review and Synthesis. Chicago: Aldine.

Sudman, S. and Bradburn, N.M. (1982). Asking Questions: A Practical Guide To Questionnaire Design. San Franciso: Jossey-Bass.

Sykes, W. and Collins, M. (1988). Effects of Mode of Interview: Experiments in the UK. In Telephone Survey Methodology, R.M. Groves, P.P. Biemer, L. Lyberg, J. Massey, W.L. Nicholls, II and J. Waksberg (eds). New York: John Wiley, 301–320.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). Psychology of Survey Response. Cambridge: Cambridge University Press.

Vermunt, J.K. (1997). LEM: A General Program for the Analysis of Categorical Data. Tilburg: Tilburg University.

Received March 2003
Revised October 2004