Journal of Official Statistics, Vol. 17, No. 3, 2001, pp. 337–350

An Exploration of Question Characteristics that Mediate Interviewer Effects on Item Nonresponse

Jan Pickery[1] *and Geert Loosveldt*[1]

In this article we analyze the item nonresponse to several questions in a face-to-face survey and assess the effect of respondent and interviewer characteristics on this nonresponse. We distinguish different kinds of questions (threatening or not, easy or complex) and different kinds of nonresponse (missing value versus ''don't know'' or ''no opinion'' answer, with or without a ''don't know'' or ''no opinion'' filter). Using multilevel logistic regression we try to discern respondent, interviewer and question related sources of item nonresponse. The main purpose of the article is to examine which item nonresponse is subject to interviewer effects. Our results suggest that question difficulty and the scope of the interviewer task might explain the size of the interviewer effects on item nonresponse.

Key words: Survey; missing value; no opinion answer; interviewer item nonresponse; multilevel logistic regression.

1. Introduction

When analyzing survey data, a large amount of item nonresponse can cause serious problems. There is a significant amount of literature about dealing with item nonresponse, e.g., different imputation methods (for an overview see Little and Rubin (1989/1990)). Though less omnipresent in the literature, the causes of item nonresponse have also gained some attention in social research (see Groves 1989, pp. 156–157). This literature sometimes does not distinguish between missing values and ''don't know'' and ''no opinion'' answers. Actually the latter are more widely discussed, and theories are chiefly developed to explain those answers.

Krosnick (1991), for instance, argues that answering ''don't know'' or ''no opinion'' is a form of satisficing. Satisficing occurs when a respondent is not motivated to expend the mental effort necessary to generate optimal answers. A ''no opinion'' answer is an acceptable answer, but it is the result of a 'weak' cognitive process. Satisficing is a function of task difficulty and the respondent's ability and motivation. This theoretical reasoning is consistent with the finding that using a ''don't know'' or ''no opinion'' filter in the question increases the proportion of respondents who give this answer, particularly among respondents with little formal education and respondents who consider an issue to be personally less important (Schuman and Presser 1981, p. 143). Following this

[1] Department of Sociology, University of Leuven, E. Van Evenstraat 2B, 3000 Leuven, Belgium. Email: jan.pickery@soc.kuleuven.ac.be and geert.loosveldt@soc.kuleuven.ac.be

Acknowledgments: An earlier draft of this article was presented at the International Conference on Survey Nonresponse, Portland, Oregon, October 28–31, 1999. This research was funded by the Flemish Foundation of Scientific Research.

argument, answering ''don't know'' is mainly explained by respondent characteristics that can be related to the cognitive aspect of answering questions and by characteristics of the questions.

Furthermore, the role of the interviewer is also stressed. Answering questions is not only a cognitive process of the respondent, it is also a communicative process (Schwarz and Sudman 1995). Within this process the interviewer plays an important role. There is substantial literature about the interviewer as a source of measurement error (for an overview see Groves 1989, pp. 357–406). The main idea is that interviewers are not 'neutral' collectors of data, but that they can influence the answers obtained. Item nonresponse is also subject to interviewer effects, as was shown long ago by Hanson and Marks (1958) and Bailar et al. (1977). The interviewers' effect on the use of ''no opinion'' answers has been demonstrated, too (for a recent example see Pickery and Loosveldt 1998). Finally, there is some evidence of an interaction of the question characteristics and the interviewer effects. The existing research is not consistent, however, regarding the influence of the question form on interviewer effects (Groves 1989, pp. 374–376).

This article attempts to contribute to this literature with a nonexperimental study. We analyze the item nonresponse to several questions in a face-to-face survey and assess the effect of respondent and interviewer characteristics on it. We distinguish different kinds of questions (threatening or not, easy or complex) and different kinds of nonresponse (missing value versus ''don't know'' or ''no opinion'' answer, with or without the anticipated possibility of that answer in the question). The main purpose of the article is to examine which item nonresponse is subject to interviewer effects.

This research question is closely related to that of Catania et al. (1997), who examine the effect of interviewer characteristics and item wording on responses to questions concerning sexual behavior. Their experimental study deals with several specific question wordings, and one of the conclusions is that many of the effects of item wording on item nonresponse mediate or are mediated by the interviewer (Catania et al. 1997, p. 369). We tackle a similar research question, focusing on general item nonresponse.

For the analyses we make use of multilevel models. It has been argued before that multilevel models offer the best possibility of analyzing the effect of both interviewer and respondent characteristics on survey data (see e.g., Hox 1994; Hox et al. 1991). Other statistical techniques require mutual independence of interviewer and respondent characteristics, which is often not the case because of the hierarchical structure of the data. Respondent and interviewer characteristics can become confounded since respondents from a specific area are most likely to be interviewed by interviewers from the same area. In a multilevel model both the regression coefficients and the variance components are conditional on the explanatory variables in the model, which is a useful property if such incomplete orthogonalization of interviewer and respondent variables is present (Hox 1994, p. 307). If the relevant respondent variables are put in the model, interviewers are equalized by statistical means. We use multilevel logistic models for dichotomous responses (item nonresponse: yes/no).

In the next section we present the data and the variables used in the analysis. Section 3 briefly describes the multilevel logistic model. Section 4 presents the results of the analyses. Section 5 tries to find general tendencies in these results, and Section 6 provides conclusions.

2. Data and Variables Used in the Analyses

The data used in the analyses come from the Belgian Election Studies. A second wave of a panel survey was set up after the 1995 General Election in Belgium. The survey made use of face-to-face interviews. In this article, we use the data from the Flemish region which cover 2,099 respondents interviewed by 158 interviewers.[2]

Question characteristics that seemed relevant in choosing the questions for the analysis were the sensitive and threatening nature and task difficulty. Furthermore we wanted to distinguish missing values from ''no opinion'' and ''don't know'' answers. Of course these question traits are not that clear-cut, and even the differences in item nonresponse might not always be that obvious. Furthermore, another important variable makes the situation more complex, especially – but not only – for the ''no opinion'' answer: the presence of a filter or a quasi-filter (the availability of that answer in the questionnaire). Because our study is a nonexperimental one that relies on re-analyses of data that we did not collect ourselves, we were not able to manipulate question form characteristics. Eventually we selected five questions that faced considerable item nonresponse and showed some variation in the recited characteristics. We evaluate these questions from the perspective of these characteristics.

Income questions are commonly considered to be sensitive or threatening, because they invade privacy (Tourangeau, Lance, and Rasinski 2000). These questions usually also are subject to considerable item nonresponse. Moore, Stinson, and Welniak (1999) report that more than a quarter of the wage and salary data in the Current Population Survey is missing or incomplete because of the respondents' lack of knowledge or because respondents are reluctant to report. In the survey we analyzed, the respondent's income is asked for with both an open and a closed question. If the respondent did not answer the open question, the interviewer had to ask the closed question. Apart from being sensitive, it is also a rather difficult question since the subject is the net family income. The open and closed questions are combined into one variable (whether or not a meaningful answer is recorded for the respondent). The interviewer had a 'no answer' alternative for both questions, but it was never mentioned to the respondent. Item nonresponse amounts to (10.5 percent; 221 respondents). This nonresponse probably comprises various kinds of response behavior like refusal or ignorance.

Questions regarding political preference are easier than the income question, because they only refer to the respondent's own judgment. They are not, however, neutral. Overreporting of voting was repeatedly observed (Presser 1990), which indicates that questions about political behavior are also sensitive. One might expect that this applies to questions about political preference as well. In this survey there were three questions regarding party preference (most preferred, second most preferred, and least preferred). These were open questions without any anticipated availability of a ''no answer'' alternative. We combined the nonresponse to these questions into one variable (no item nonresponse versus at least

[2] A detailed description of the sampling scheme plan can be found in Beerten et al. (1997, pp. 11–17). We thank the ISPO-PIOP Centre for Electoral Research for providing us with the data. The Flemish data were originally collected by Jacques Billiet, Marc Swyngedouw, Ann Carton and Roeland Beerten. The ISPO-PIOP is supported by the Federal Services for Technical, Cultural and Scientific Affairs. Neither the Centre nor the original collectors bear any responsibility for the analyses or interpretations presented here.

one question without a meaningful answer). For more than one third of the respondents, we counted at least one item nonresponse (35.6 percent; 640 respondents). The amount of nonresponse to these three questions is not the same. The question regarding the second most preferred party faced the largest nonresponse (30 percent, about twice as much as the other two questions), which clarifies that the absence of a preference probably makes up an important part of the item nonresponse to these questions, apart from refusals.

In the other three analyses, we restrict ourselves to ''don't know'' and ''no opinion'' answers. One of the more complex tasks of the questionnaire with an anticipated possibility of a ''don't know'' answer was *rating six parties on eight different 11-point scales* (Catholicism, economic liberalism, immigrant rights . . .). This is a difficult task. It involves a thorough knowledge of the six major Flemish parties, and not every party has a clear position on each of the scales. Answering these questions requires serious cognitive effort on the part of the respondent (Loosveldt 1999). The task is not, however, very threatening, since the respondent's position it not at stake and it is not presented as a knowledge test. Furthermore, an explicit ''don't know'' filter was included in the question, though it was not mentioned on the show card (quasi-filter). These numerous difficult questions resulted in a lot of nonresponse: 1,159 respondents made use of the ''don't know'' alternative at least once (55.2 percent).

The respondents were also asked the extent of *approval of 19 statements* that were being discussed during the General Election. Since the question concerns the respondent's own opinion and the statements refer to actual electoral topics, we consider it to be an easier and not so threatening question. Also in this case a ''no opinion'' filter was included in the question but was not on the show card. Again the number of ''no opinion''s is dichotomized: at least one/none (item nonresponse: 27.1 percent; 568 respondents).

Afterwards the respondents had to select three items out of the 19 statements that were important to them when voting for a political party and *evaluate the last government's policy regarding these items*. For these three items the ''no opinion'' alternative was supplied in the questionnaire, but it was not mentioned in the question or on the card for the respondent. For this question the availability of that response alternative lies between the income question and the rating of parties and political statements questions. Nonresponse is considerably lower for these questions. It only sums up to 8.3 percent (175 respondents used the ''no opinion'' answer at least once).

We dichotomized all dependent respondent variables. The occurrence of item nonresponse in each of the presented questions is the dependent variable of our analyses. This implies some loss of information, but the uniformity in the dependent variables facilitates comparison of the results.

Of course these dependent variables vary considerably, e.g., in question length, complexity and sensitivity. These variations (partly) account for differences in the amount of nonresponse. It is not surprising that the highest nonresponse is found for the longest question (rating of parties question, which actually comprises 48 questions). The main purpose of our analyses is not, however, to explain the amount of item nonresponse. After all, we only selected questions which faced substantial nonresponse. We want to examine which item nonresponse is subject to interviewer effects. Accordingly, we

compare various questions with a certain amount of ''unusable'' data. The existence of interviewer effects would imply that item nonresponse it not only a matter of respondent behavior, but also of interviewer behavior. Our analyses might clarify whether the same question traits that explain nonresponse also explain the interviewer effects on nonresponse. That is, question traits could determine interviewer behavior. We separate interviewer effects from respondent effects by including both with the respective variables explicitly in multilevel models.

The selection of the **independent respondent variables** is largely based on previous research about the use of the ''no opinion'' and ''don't know'' answer. Relevant variables are age (see e.g., Groves 1989, pp. 441–443 and Bell 1984, p. 212), education (e.g., Sudman and Bradburn 1974, pp. 98–99), sex (e.g., Hox et al. 1991, p. 450) and a measure of involvement or interest in the subject (e.g., Groves 1989, p. 419). We use age (centered around the grand mean), sex (0: man, 1: woman), two dummies for education (low education (0: no, 1: yes) and high education (0: no, 1: yes), so the base category is mean education) and also two dummies for political interest (low political interest and high political interest, based on three items and identically coded).

Previous research results with regard to the effect of **interviewer characteristics** do not offer similar specific guidelines concerning which variables to select. According to some authors the effects of the socio-demographics of the interviewer seem to be negligible (Groves 1989, p. 408; Hox et al. 1991, p. 459; De Leeuw and Collins 1997, p. 213). Berk and Bernstein (1988: p. 245), on the other hand, find that younger interviewers have higher nonresponse rates. The results for role dependent characteristics, variables related to the interviewing task, are not uniform. Research dealing with interviewer experience is confusing. Fowler and Mangione (1990, p. 134) conclude that some experience is better than no experience, but that very experienced interviewers take too many liberties. Singer et al. (1983, pp. 78–79) found no significant effect of the interviewers' experience on the amount of nonresponse.

We use the same socio-demographic variables at the interviewer level as at the respondent level, although we decided in favor of only one dummy for education (0: low, 1: high) and political interest (identically coded). The reason is that there is less variation between interviewers than between respondents for these variables. We also include the number of interviews and degree of experience as role dependent interviewer variables. The latter variable is a dummy that records whether or not the interviewer had already collaborated in an ISPO survey. Considering the former variable it could be argued that interviewers who conducted more interviews might have worked faster and more sloppily, resulting in more item nonresponse. An opposite hypothesis could be that those interviewers become more proficient in their job and more familiar with the questionnaire and hence better at obtaining meaningful answers from the respondents. As a last interviewer characteristic, we bring in the interviewer item nonresponse as a measure of interest. The interviewers also had to complete the questionnaires themselves, and their item nonresponse can be included in the models. The hypothesis is that if they do not fill in the question, they might also consider it too threatening, too difficult or too irrelevant to insist on meaningful answers from respondents.

3. The Multilevel Logistic Regression Model

A logistic regression is an appropriate technique to analyze a dichotomous outcome. Goldstein (1995, pp. 77–111) shows how to include random coefficients in models for discrete response data. We present the elaboration for the logistic model.

In a multilevel model for respondent and interviewer effects the respondents constitute the first level and the interviewers the second. In a logistic model the dependent variable becomes a probability. The probability of a ''success'' (e.g., item nonresponse) for respondent i, interviewed by interviewer $j(\pi_{ij})$ is modeled as follows:

$$\pi_{ij} = \frac{\exp(\beta_{0j} + \beta_{1j} x_{1ij})}{1 + \exp(\beta_{0j} + \beta_{1j} x_{1ij})} \tag{1}$$

or

$$\log \text{it}(\pi_{ij}) = \beta_{0j} + \beta_{1j} x_{1ij} \tag{2}$$

Subscript i refers to the respondent and subscript j to the interviewer. In this model there is one independent respondent variable (x_{1ij}), but of course more independent variables are possible. The regression coefficients (β's) are interviewer specific; some or all have a subscript j, thus implying an interviewer residual. The elaboration of the Level 2 equations is similar to the general multilevel model. The residuals at the interviewer level can be denoted by u:

$$\beta_{0j} = \beta_0 + u_{0j}, \quad \beta_{1j} = \beta_1 + u_{1j} \tag{3}$$

If only the parameter associated with the constant varies across interviewers, the model is often called a random intercepts model. If the regression coefficients are also allowed to vary, the model is called a random slopes model. The β's can also be made dependent on higher level variables (interviewer characteristics), allowing for generalization across interviewers.

$$\beta_{0j} = \beta_0 + \gamma_{01} z_{1j} + u_{0j}, \quad \beta_{1j} = \beta_1 + \gamma_{11} z_{1j} + u_{1j} \tag{4}$$

We define one higher level variable (z_{1j}), but of course more interviewer variables can be included in the model. Mostly the same second level variables are included in all β-equations, but this is not necessary.

The observed response (0 or 1) for respondent i interviewed by j equals the probability plus an error term: $y_{ij} = \pi_{ij} + e_{ij}$. The model is completed by specifying a distribution for the observed response $y_{ij} \mid \pi_{ij}$. Usually a binomial distribution is assumed. This distributional assumption can be implemented by the introduction of an extra explanatory variable at the first level, w_{ij}. Given the binomial distribution w_{ij} is defined as follows:[3]

$$w_{ij} = \sqrt{\pi_{ij}(1 - \pi_{ij})} \tag{5}$$

The model then becomes:

$$y_{ij} = \pi_{ij} + e_{ij} w_{ij} \tag{6}$$

[3] When analyzing a table instead of individual level data with a logistic model – which is formally equivalent – an extra term (cell total) is needed in this equation.

and the distributional assumption is imposed by constraining the Level 1 variance (σ_e^2) to be one. By relaxing this constraint on σ_e^2 the model allows for extra binomial variation.

This model can be fitted in MLwiN with a Quasilikelihood estimation procedure (see Rasbash et al. 2000, pp. 99–116). All models in the following section are fitted using second order PQL.[4] A disadvantage of Quasilikelihood estimation is that the likelihood value is only approximate and likelihood ratio tests are always unreliable. Consequently we will not give overall test statistics for the models. Significance tests are based on the standard errors (WALD-test).[5]

4. Results of the Analyses

For every analysis we present the model with all independent variables that were significant in at least one of the five different analyses, even if these variables did not make a significant contribution in the analysis involved. We do not present any test of the assumption of the binomial variation.[6] Including the same variables and constraining the variance for all models makes the results more comparable.

Table 1 reports the results. Note that there are two values for the interviewer variance (σ_{u0}^2) in the table. The first one corresponds to the final model with all variables and is the most important one. The second one (marked with a $^{\#}$) corresponds to a model without interviewer variables, but with the respondent variables. In those models we just included the interviewer as a random factor, but we do not try to explain the interviewer variance with interviewer variables. Those $\sigma_{u0}^{2\,\#}$-values show that the inclusion of interviewer variables in some models that (partly) account for the interviewer variance does not affect the relative sizes of those variances. This is important when comparing the magnitude of the interviewer effects in Section 5.

We firstly clarify the results of the table by looking at all analyses separately. Nonresponse to the *income question* is partially explained by the respondent's age and sex. Older respondents are less reluctant to report their income. Women, on the other hand, are more likely not to give a meaningful answer to this question. The parameters associated with these variables can be further clarified by calculating odds ratios. The odds of item nonresponse (yes/no) to the income question for women are 1.634 ($= e^{0.491}$) times the odds for men. For the numerical variable age the odds ratio corresponds with a one-unit change in age. The interviewer variance is large and clearly significant. In other words item nonresponse to the income question is subject to interviewer effects. However, there is only one interviewer variable that accounts partly for the interviewer effect: the interviewer's political interest. The item nonresponse to this question is smaller if the interviewer has a greater political interest. This is the only analysis with a significant

[4] Two different Quasilikelihood procedures are possible in MLwiN: Marginal Quasilikelihood (MQL) and Predictive Quasilikelihood (PQL). MQL tends to underestimate the values of both the fixed and the random parameters. PQL is more accurate but computationally less stable. Furthermore a second order approximation also enhances accuracy (see Goldstein 1995, pp. 83–86 and 99–101 for details). If possible, second order PQL should be used.

[5] Bootstrapping is an alternative for the Wald-statistics for significance testing if the likelihood ratio test is unreliable. MLwiN provides some bootstrapping facilities.

[6] Two analyses showed very small underdispersion (with σ_e^2 estimates of approximately 0.95), but in those models the estimates of all other coefficients remained almost completely the same and all significance tests showed entirely similar results.

Table 1. The results of the multilevel analyses

	Family income		Party preference		Rating of parties		Political statements		Government policy	
	Parameter	s.e.	Parameter	s.e.	Parameter	s.e.	Parameter	s.e.	Parameter	s.e.
Fixed										
Respondent level										
constant	−2.434	0.378**	−1.046	0.154**	−0.538	0.214**	−1.949	0.190**	−3.214	0.281**
age	−0.020	0.006**	−0.001	0.003	−0.006	0.004	0.009	0.004*	−0.010	0.006
sex	0.491	0.168**	0.106	0.100	0.949	0.105**	0.911	0.115**	0.938	0.192**
low education	0.126	0.221	0.352	0.131**	0.107	0.137	0.463	0.148**	0.392	0.232
high education	0.176	0.213	−0.015	0.132 **	−0.210	0.134 **	−0.177	0.155 **	−0.185	0.249 *
low political interest	0.215	0.194	0.683	0.119**	0.319	0.132*	0.748	0.131**	0.648	0.194**
high political interest	−0.219	0.205	−0.384	0.120** **	−0.367	0.118** **	−0.336	0.143* **	−0.600	0.262* **
Interviewer level										
political interest	−0.530	0.217*	0.041	0.135	0.186	0.190	−0.065	0.162	−0.200	0.211
same item nonresponse	−0.153	0.319	0.373	0.125**	0.553	0.190**	0.478	0.194*	0.081	0.725
Random										
Interviewer level										
σ^2_{u0}	0.623	0.191**	0.217	0.066**	0.937	0.155**	0.478	0.111**	0.427	0.180**
Respondent level										
σ^2_e	1.000	0.000†	1.000	0.000†	1.000	0.000†	1.000	0.000†	1.000	0.000†
$\sigma^{2\ \#}_{u0}$	0.713	0.204**	0.235	0.068**	1.011	0.163**	0.523	0.116**	0.435	0.181**

*$p < 0.05$, **$p < 0.01$.

†constrained.

#model without Level 2 (interviewer) independent variables.

effect of interviewer political interest, which is a rather strange result since the income question – as opposed to the other questions – has no direct reference to the subject of the survey: political attitudes and voting behavior.

Political interest is an important factor in explaining nonresponse to the question regarding *political party preference*. Both dummies (low political interest/high political interest) have a significant contribution, with a higher chance of nonresponse for respondents with low political interest and a smaller chance for respondents with high political interest. Consequently the joint test of significance[7] shows that both dummies together are highly significant. Education also has an effect on the nonresponse to these questions. Lower educated respondents have a higher chance of nonresponse. The last dummy is far from significant, but the joint test for both dummies proves significance. Again there is a general interviewer effect (significant interviewer variance) and only one interviewer characteristic that accounts for this effect: item nonresponse to the same question by the interviewer. Item nonresponse for the respondent is more likely if the interviewer did not answer this question him or herself. The interviewer variance is significant, but much smaller for this variable than for the income question.

The third analysis deals with the very complex task of *rating the six major parties* on eight different scales. These questions resulted in a lot of nonresponse (''don't know'' answers). The respondent's sex and his/her political interest explain the use of the ''don't know'' answer to these questions. Women make use of the ''don't know'' answer more often than men do and political interest has an effect on the response, following the expectations and the results of the previous analysis. The separate and joint tests of significance yield the same results. As far as education is concerned the single dummies for low and for high education are not significant, but the joint test shows that there are indeed differences in the use of the ''don't know'' answer according to the respondent's educational level (more for respondents with a lower level of education, less for ones with a higher level of education). The interviewer effect for the rating of parties question is the strongest of all five analyses. Again there is only one significant interviewer variable: item nonresponse to the same question, with a positive correlation between the respondent's and the interviewer's nonresponse. This is a complex and very extensive task for the respondents. The results of this analysis indicate that the interviewers certainly have an effect on their response behavior and that this effect is determined by how they carried out the task themselves.

The following analysis considers the use of the ''no opinion'' answer to 19 questions regarding approval of *political statements*. This was an easier question than the previous one and respondents made markedly less use of the ''no opinion'' alternative. All respondent variables included in the analysis contribute to the explanation of the use of the ''no opinion'' answer. Older respondents use it more frequently than younger respondents, and women more than men. Education and political interest again follow the expectations: a higher chance for respondents with a lower level of education and less interest in politics and a smaller chance for their counterparts. The high education dummy is not significant at the $\alpha = 0.05$ level, but the joint test for both education dummies clearly passes the significance test. Also for this analysis the interviewer

[7] H_0: $\beta_{\text{low political interest}} = \beta_{\text{high political interest}} = 0$.

variance is significant. So the interviewer again has an effect on the use of the ''no opinion'' answer. Entirely similar to the results of the two previous analyses, the only interviewer variable that partly accounts for the interviewer variance is his or her nonresponse to the same question. If the interviewer used the ''no opinion'' answer at least once, the chance of the respondent's using that answer increases.

The last analysis deals with a follow-up question to the previous one. The respondent had to select the three items out of the 19 that they considered important and assess the *government's policy* towards these topics. Respondent variables explaining the use of the ''no opinion'' answer to this question are sex, education and political interest. All effects follow the expectations and the results of the previous analyses. But only the joint test of both education dummies shows significance. Also for this variable there is an interviewer effect, but there is no interviewer variable that accounts for the interviewer variance.

5. General Tendencies in These Results?

When it comes to the effects of **respondent characteristics** these analyses were largely confirmations of previous research findings. The most important factor in explaining item nonresponse in general and the ''no opinion'' and ''don't know'' answer more particularly, was the respondent's political interest. The only question for which there was no evidence that political interest affected item nonresponse was the income question, a question that is not directly related to the subject of the survey. This effect of the interest in the subject is not really surprising. Sex was also an important explanatory factor. This can also be considered as a confirmation of the recurring finding that women usually have a lower interest in and knowledge of politics (see e.g., Carton 1993). Education was significant in four out of the five analyses, though sometimes only when a joint test for both dummies was considered. Every time it followed the expectations. Age had two opposite effects. For the income question, older respondents had a smaller chance of nonresponse; for the 19 political statements, older respondents had a higher chance. The latter finding is more in line with previous findings than the former, but again the income question is a rather special question that is not directly related to the subject of the survey. All these effects are more or less in line with expectations.

When it comes to the **interviewer effects and interviewer variables** of interest, the first remarkable result is that all analyzed item nonresponse was subject to interviewer effects. The answer to the main research question of this article ''Which item nonresponse is subject to interviewer effects?'' is indeed simple: all the item nonresponse we analyzed!

Secondly, we found a significant interviewer variance for all intercepts, but no varying regression coefficients. None of the analyses proved evidence of random slopes. So the chance of item nonresponse differs from interviewer to interviewer, but the effect of e.g., political interest on nonresponse did not vary significantly between interviewers.

Thirdly, interviewer item nonresponse to the same question turned out to be the interviewer variable that accounted most for the interviewer variance. In three out of the five analyses it made a significant contribution. This finding confirms the hypothesis that interviewers who do not fill in a particular question because they find it too threatening or too difficult will accept nonresponse from the respondents more easily. We suspect that they

Table 2. The interviewer variance and the level of item nonresponse

Question	σ^2_{u0}	Respondent item nonresponse	Interviewer item nonresponse
1) rating of parties	1.011	0.552	0.503
2) family income	0.713	0.105	0.121
3) political statements	0.523	0.271	0.206
4) government policy	0.435	0.083	0.024
5) party preference	0.235	0.356	0.301

Table 3. The interviewer variance and the question characteristics

Question	σ^2_{u0}	Nature of nonresponse	Task difficulty	Threatening nature	Availability in questionnaire
1) rating of parties	1.011	don't know	+	−	+
2) family income	0.713	all	0	+	0
3) political statements	0.523	no opinion	−	−	+
4) government policy	0.435	no opinion	−	−	0
5) party preference	0.235	all	−	+	−

may not even ask the question (properly). In general, though, there are very few (of the available) interviewer variables that account for the interviewer effects. Interviewer age, sex, education, experience and number of interviews were not significant in any of the models. This is a confirmation of the results of Berk and Bernstein (1988, p. 250). They conclude that interviewer performance cannot be predicted on the basis of easily measured interviewer characteristics like education and experience.

Finally, we found significant interviewer variances in all analyses but we can still try to explain the magnitude of those interviewer effects. Possible explanatory factors are the level of nonresponse for the respondents and/or the interviewers and the question characteristics as described in Section 2. Table 2 and Table 3 rank the five questions according to the size of the interviewer effect on their item nonresponse.[8] Furthermore they report the level of nonresponse (Table 2) and a score on the question characteristics (Table 3).

As Table 2 shows, neither the level of the item nonresponse for the respondents nor the same level for the interviewers seems to be a good predictor of the size of the interviewer effect. The largest effect was found for the question with the largest item nonresponse, but the question with the second largest nonresponse had the smallest interviewer effect. Apart from that there is no pattern at all. This finding is similar for respondent and interviewer item nonresponse. Furthermore those seem to show a considerable correlation.

[8] For this comparison we use the interviewer variance in the models without interviewer variables ($\sigma^{2\#}_{u0}$). These σ^2_{u0} values refer to models that include all the respondent variables, but account only for the interviewer as a random factor. So the interviewer identification is part of the model, but not the interviewer variables. $\sigma^{2\#}_{u0}$ could be called the ''explainable interviewer variance.'' In some analyses part of that variance is explained or modeled by the inclusion of the interviewer variables. Since we are interested in the size of the general interviewer effect, comparing the total amount of explainable variance is more correct. (Actually it does not make any difference since only a small portion of the variance is explained.)

Interviewer effects are often compared by expressing them as intra-class correlations (ICC). In a model with binomial response the ICC can be calculated by setting the Level 1 variance at $\pi^2/3$ (Hedeker and Mermelstein 1996, p. S216). π is the number denoted by the Greek letter and should not be confused with the probability (π_{ij}) in the formulas (1)–(6). Since we have only models with binomial responses and we constrained all Level 1 variances, we can as well compare the Level 2 variances. The result of that comparison is entirely identical to that of a comparison of ICC's.

In Table 3 we tried to score the questions on the relevant question characteristics described in the second section. In a way these scores are arbitrary (especially for task difficulty and threatening nature). They should certainly be interpreted as being relative to the other four questions. A + means more difficult, more threatening or more easily available, a − less difficult, less threatening or less easily available, and a 0 represents a middle position on these dimensions. As far as the nature of the item nonresponse (''no opinion'' or ''don't know'' versus other/all) is concerned, no pattern can be found, so this trait does not seem to explain the size of the interviewer effect at all. Nor is there any connection with the threatening nature of the questions. For the availability there is no straight pattern either, but it is striking that the smallest effect was found where there was no anticipated availability of a ''no answer'' alternative at all. For the party preference question it was not mentioned to the respondent and it was not provided in the questionnaire for the interviewer. With task difficulty there also seems to be a connection. The interviewer effect is the largest for the most difficult question. We also considered the income question rather difficult because the subject was the net family income. For hourly-paid workers and members of families with several incomes this is not really obvious.

The combination of these last two question traits (difficulty and availability of the filter) also extends the opportunity for the interviewer to intervene. The rating of parties question is very long and interviewers can interrupt after each scale or when encountering a particular party. The ''don't know'' answer is a tactful chance of escape for them as well. The income question consisted of two questions (an open and a closed one) and the interviewer had to move over to the second one if the respondent did not answer the first one. Of course the analysis of five questions is not enough to draw general conclusions, but we hypothesize that the combination of task difficulty for the respondent and a more complex job description for the interviewer (of which partial availability of item nonresponse is an important aspect) are significant factors in explaining the size of the interviewer effect.

A possible drawback of this comparison is that the effect of the independent respondent variables varies across question type as well. Nonresponse to the political statements question is better explained in terms of respondent variables than nonresponse to the income question. But with the inclusion of the independent variables the respondents of each interviewer are basically comparable. It is unlikely that other respondent variables that explain item nonresponse are correlated with the interviewer classification. Consequently we assume that more or other respondent variables will not alter the results of the comparison of the relative sizes of the interviewer effects.

6. Conclusion

Our analysis confirms previous research about item nonresponse. It is related to the interest in the subject of the survey, to the respondent's sex and, to a lesser extent, to his/her age and education. Furthermore it is subject to interviewer effects. Since our research was nonexperimental and we examined only a selection of questions, we cannot give a full explanation of the size of the interviewer effects. But the results of the analysis prompt us to conclude that the interviewer effects on item nonresponse may be a function of the task difficulty of the question and the corresponding complexity of the interviewing

task. The presence of a quasi-filter is an important aspect of the latter. Further research that also considers questions with complete filters, apart from quasi-filters is needed to check this hypothesis. The questionnaire we examined did not have such questions. Anyway the omnipresence of the interviewer effects and the effect of the interviewer item nonresponse in our analyses prove that item nonresponse is not only a matter of answering questions or not, but also of asking them and how that is done.

These conclusions are based on a small-scale nonexperimental study. Future research is needed to elucidate the findings. Firstly, more questions from different surveys need to be analyzed. Secondly, experiments would further facilitate the analysis of the effect of the question characteristics. There is plenty of experimental research focusing on the effects of question characteristics on item nonresponse (see Groves 1989, Chapter 10). Similar experiments should include the interviewer as a subject of the analysis. Finally, the recurring finding of the effect of the interviewer item nonresponse calls for further research. If the information about the interviewer item nonresponse is available at the beginning of the interviewer training, it can be used to trace questions that will probably face a large nonresponse rate and interviewers that will probably obtain more item nonresponse. It is worth examining whether an extra training for these interviewers or extra focus on these particular questions can reduce item nonresponse.

7. References

Bailar, B., Bailey, L., and Stevens, J. (1977). Measures of Interviewer Bias and Variance. Journal of Marketing Research, 14, 337–343.

Beerten, R., Billiet, J., Carton, A., and Swyngedouw, M. (1997). 1995 General Election Study. Codebook and Questionnaire. Leuven: ISPO.

Bell, R. (1984). Item Nonresponse in Telephone Surveys: An Analysis of Who Fails to Report Income. Social Science Quarterly, 65, 207–215.

Berk, M. and Bernstein, A. (1988). Interviewer Characteristics and Performance on a Complex Health Survey. Social Science Research, 17, 239–251.

Carton, A. (1993). Blijft Politiek een Mannenzaak? Een Balans op Basis van de Resultaten van het Onderzoek naar de Politieke Houdingen van de Vlaamse Vrouwen en Mannen. In Kiezen is Verliezen. Onderzoek naar de Politieke Opvatingen van Vlamingen, eds. M. Swyngedouw, J. Billiet, A. Carton, and R. Beerten. Leuven: Acco.

Catania, J.A., Binson, D., Canchola, J., Pollack, L.M., Hauck, W., and Coates, T.J. (1996). Effects of Interviewer Gender, Interviewer Choice, and Item Wording on Responses to Questions Concerning Sexual Behavior. Public Opinion Quarterly, 60, 345–375.

De Leeuw, E.D. and Collins, M. (1997). Data Collection Methods and Survey Quality. In Survey Measurement and Process Quality, eds. L.E. Lyberg, P.P. Biemer, M. Collins, E.D. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin. New York: Wiley.

Fowler, F.J. and Mangione, T.W. (1990). Standardized Survey Interviewing: Minimizing Interviewer-related Error. Newbury Park: Sage.

Goldstein, H. (1995). Multilevel Statistical Models. London: Edward Arnold.

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

Hanson, R.H. and Marks, E.S. (1958). Influence of the Interviewer on the Accuracy of Survey Results. Journal of the American Statistical Association, 53, 635–655.

Hedeker, D. and Mermelstein, R.J. (1996). Application of Random-Effects Regression Models in Relapse Research. Addiction (Supplement), 91, S211–S229.

Hox, J.J. (1994). Hierarchical Regression Models for Interviewer and Respondent Effects. Sociological Methods and Research, 22, 300–318.

Hox, J.J., De Leeuw, E.D., and Kreft, I.G. (1991). The Effect of Interviewer and Respondent Characteristics on the Quality of Survey Data: A Multilevel Model. In Measurement Errors in Surveys, eds. P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman. New York: Wiley.

Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. Applied Cognitive Psychology, 5, 213–236.

Little, R.J.A. and Rubin, D.B. (1989/1990). The Analysis of Social Science Data with Missing Values. Sociological Methods and Research, 18, 292–326.

Loosveldt, G. (1999). The Interviewer as Informant about the Interview Process. In ASC '99 – Leading Survey & Statistical Computing into the New Millennium. Proceedings of the ASC International Conference, September 1999, eds. R. Banks, C. Christie, J.J. Francis, P. Harris, B. Lee, J. Martin, C. Payne, and A. Westlake. Edinburgh: The Association for Survey Computing – University of Edinburgh.

Moore, J., Stinson, L., and Welniak, E. (1999). Income Reporting in Surveys: Cognitive Issues and Measurement Error. In Cognition and Survey Research, eds. M. Sirken, D. Herrmann, S. Schechter, N. Schwartz, J. Tanur, and R. Tourangeau. New York: Wiley.

Pickery, J. and Loosveldt, G. (1998). The Impact of Respondent and Interviewer Characteristics on the Number of ''No Opinion'' Answers. A Multilevel Model for Count Data. Quality & Quantity, 32, 31–45.

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

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., and Lewis, T. (2000). A User's Guide to MLwiN. London: Multilevel Models Project, Institute of Education, University of London.

Schuman, H. and Presser, S. (1981). Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context. New York: Academic Press.

Schwarz, N. and Sudman, S. (1995). Answering Questions. San Francisco: Jossey-Bass.

Singer, E., Frankel, M., and Glassman, M. (1983). The Effect of Interviewer Characteristics and Expectations on Response. Public Opinion Quarterly, 47, 68–83.

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

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

Received December 1999
Revised May 2001