Open Access Repository

www.ssoar.info

The effects of question, respondent and interviewer characteristics on two types of item nonresponse

Silber, Henning; Roßmann, Joss; Gummer, Tobias; Zins, Stefan; Weyandt, Kai Willem

Mitglied der

Leibniz-Gemeinschaft

gesis
Leibniz-Institut
für Sozialwissenschaften

ORIGINAL ARTICLE

The effects of question, respondent and interviewer characteristics on two types of item nonresponse

Henning Silber[1] | **Joss Roßmann[1]** | **Tobias Gummer[1]** | **Stefan Zins[2]** | **Kai Willem Weyandt[1]**

[1]GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

[2]IAB Institute for Employment Research, Nuremberg, Germany

Correspondence
Henning Silber, GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany.
Email: henning.silber@gesis.org

Abstract

In this article, we examine two types of item nonresponse in a face-to-face population survey: 'don't know' (DK) and 'item refusal' (REF). Based on the cognitive model of survey response, the theory of survey satisficing and previous research, we derive explanatory variables on three levels: question, respondent and interviewer characteristics. The results of our cross-classified model show that while the two levels question and respondents' characteristics affected both types of item nonresponse, interviewer characteristics affected only DK answers. Our results also confirm that DK and REF are substantially different item nonresponse types resulting from distinguishable disruptions of the cognitive response process. Since most results are in line with prior theoretical predictions, they suggest that survey practitioners are well-advised by continuing to follow the large body of practical guidance derived from the theories tested here.

KEYWORDS
cognitive response process, do not know answers, item refusals, measurement error, question design, response behaviour

1 | INTRODUCTION

Nonresponse error is one of the most important challenges to data quality in survey research (Groves, 2004). When a person initially agrees to respond to a survey, there remains the likelihood that the person does not provide 'valid' answers to individual survey questions. Item nonresponse is considered to be the result of at least two cognitive decisions (see Beatty & Herrmann, 2002): First, respondents have to evaluate whether they can provide a 'valid' answer, and second, respondents must evaluate whether they are willing to give a 'valid' answer. If one of these decisions is negative for a question, a respondent will not respond to it. Item nonresponse can take the form of '*item refusals*' (REF) or '*don't know*' (DK) responses. However, only a few studies (e.g. Juster & Smith, 1997; Shoemaker et al., 2002) have systematically differentiated between the two item nonresponse types and proposed explanatory concepts that lead to one or the other.

The previous studies that have investigated at least one type of item nonresponse used explanatory characteristics on three levels: characteristics of the questions such as question wording (e.g. Coombs & Coombs, 1976; Tourangeau et al., 2000), characteristics of the respondents such as education (e.g. Bishop et al., 1986; Schuman & Presser, 1996) and interviewer effects (e.g. Singer & Kohnke-Aguirre, 1979). Yet, only a few studies have employed a multilevel model that includes all these three levels at once, while coding almost all questions of a specific survey (Dahlhamer et al., 2020; Dykema, Schaeffer, & Garbarski, 2016; Dykema et al., 2020; Holbrook et al., 2006; Holbrook et al., 2016; Olson et al., 2018). Of those studies, only Olson et al. (2018) and Dahlhamer et al. (2020) used multilevel modelling to investigate item nonresponse (both with telephone surveys). However, the focus of Olson et al. (2018) was on the cognitive response process, and Dahlhamer et al. (2020) did not distinguish between different types of item nonresponse. Thus, the current knowledge on the causes of item nonresponse and their interplay remains fragmentary.

The present study contributes to this evidence by systematically examining sources for both REF and DK on the level of the question, the respondents and the interviewer in the 2013 Pre-Election Survey of the German Longitudinal Election Study (Rattinger et al., 2014). Specifically, our analysis includes the majority of the survey questions, considers the type of (non)response to them as the dependent response variable in a multinomial logit model and uses characteristics of the questions, respondents and interviewers as predictors of these response types. By doing that, our research design allows us to examine parallels and differences in the explanation of REF and DK on each of these three levels.

2 | THEORY OF ITEM NONRESPONSE

The cognitive model of the response process to a survey question includes four cognitive steps: comprehension of the question, retrieval of relevant information from memory, judgement based on the retrieved information and mapping the answer on the response scale (Bradburn et al., 2004; Tourangeau et al., 2000).

If the comprehension step is not successful, a nonresponse is a likely outcome. Especially, a DK response is likely because the person has difficulties to understand the meaning of the question (see Olson et al., 2018). With respect to the retrieval step, Beatty and Herrmann (2002) proposed four cognitive states. The first state describes the information, which is available and can be retrieved with minimal effort. The second state describes the information, which is accessible and can be retrieved with effort or prompts. The third state describes the information that is generatable, which means that it can be estimated, and the fourth state describes the information, which is not known and there is

no basis for estimation. When considering the cognitive processes of judgement and response, and given that respondents decide against editing their answer to conform with social norms such as not admitting that they did not vote in a previous election (see, e.g. DeMaio, 1984), the first three states of Beatty and Herrmann's model lead to a substantive response and the fourth to item nonresponse. However, respondents can decide to edit their responses, which results in misreporting and measurement error. Specifically, misreporting of item nonresponse occurs when a respondent could give a substantive answer, but does not (error of omission), and misreporting of substantive response occurs when a respondent cannot give a substantive answer but does give one anyway (error of commission). When considering the difference between REF and DK, DK is more likely if the information is not accessible and not estimable, while a REF answer is more likely if a response is edited due to social desirability (e.g. Olson et al. 2018).

Besides comprehension problems, inestimability and errors due to omission, respondents can also decide to shortcut the response process by providing an answer which can be given without much thinking (see Krosnick, 1991). Such 'satisficing response strategies' include giving substantive responses, such as selecting the first reasonable response option, but also providing DK responses, in particular.

Past research has shown important differences between REF and DK responses. For instance, Shoemaker et al. (2002) provided evidence that DK was associated with cognitive effort, whereas REF was associated with both cognitive effort and question sensitivity. In addition, Juster and Smith (1997) showed important differences between REF and DK responses to income questions.

While any survey interview is an interactive process between the researcher and respondent (Toepoel & Dillman, 2011), the communicative nature of an interview is particularly salient in interviewer-administered surveys. With respect to the question–answer sequence, Dykema et al. (2020) proposed a model that displays a response as a function of the interaction between characteristics of the question, the respondent and the interviewer. In the heart of that model is the 'interactional processing', which, together with the behaviour of the respondent and interviewer, ultimately leads to item (non) response. Moreover, Japec (2008) reasoned that interviewer burden might lead to interviewer satisficing. Consequently, *interviewers' ability* and *motivation* may also influence the likelihood of REF and DK answers.

In the following, we will outline our expectations regarding the probability of item nonresponse when considering characteristics on the question, respondent and interviewer level.

2.1 | Question characteristics

In line with Tourangeau et al. (2000), we propose that cognitive processing varies across different *question types*. Specifically, information retrieval is substantially different for questions about attitudes and behaviours, on the one hand, and for questions about facts and knowledge, on the other hand (see Olson et al., 2018). While questions about attitudes or behaviours are likely to require that respondents generate information that is not in their memories, questions of the two remaining types are more likely to be available or accessible (see Beatty & Herrmann, 2002). Thus, based on the cognitive state model, we expect that questions about facts or knowledge result in lower amounts of DK responses than questions about attitudes or behaviours. In contrast, we assume questions about facts or knowledge to result in higher amounts of REF because respondents are expected to have a higher likelihood of editing their answers to present a socially desirable image of themselves or to avoid admitting a lack of knowledge (e.g. when asked to report their income). With respect to attitudinal questions, Bradburn et al. (2004, p.137) emphasized that, in contrast to behavioural questions, there is 'no

[retrievable] "true" answer'. Thus, we expect that the higher cognitive effort will result in more DKs for attitudes than behavioural questions, whereas the number of REF responses should not be affected.

Other characteristics of a question that are likely to influence the number of REF and DK answers are its difficulty, sensitivity and context (see Holbrook et al., 2006; Olson et al., 2018). For instance, a long and complex wording increases *question difficulty* and makes it more burdensome for respondents to comprehend a question, which is likely to increase both types of item nonresponses (Krosnick, 1991; Olson et al., 2018). With respect to retrieval, question difficulty is likely to increase the number of DK responses because the relevant information might be challenging to access or generate (Beatty & Herrmann, 2002; Olson et al., 2018). This might be specifically the case for hypothetical questions, questions about intentions and recall questions. In contrast, the difficulty of the retrieval process should not affect the likelihood of REF responses.

In addition, previous research has also provided evidence that the *question format*, which is closely linked to the cognitive processes of judgement and mapping, influences the likelihood of item nonresponse (e.g. Biemer & Lyberg, 2003; Olson et al., 2018). While respondents who answer open-ended questions can provide any kind of response in their own words, closed questions require that respondents map their answers onto a given response scale (see Olson et al., 2018; Singer & Couper, 2017). With respect to item nonresponse, we expect that closed questions are less burdensome to answer and lead to less REF and DK answers than open-ended questions on attitudes and behaviours because respondents can infer the universe of appropriate answers from the response scale (e.g. Schuman & Presser, 1996; Singer & Couper, 2017). Similarly, we expect that (closed) multiple-choice questions, in which respondents can select more than one response category, are likely to result in more DK and REF responses because they are cognitively more demanding than standard closed single-choice questions.

Question sensitivity refers to whether a question asks about attitudes or behaviours that respondents consider being confidential or socially undesirable (Shoemaker et al., 2002). Research about the connections between question sensitivity and item nonresponse provided mixed evidence on that relationship (see Olson et al., 2018; Tourangeau & Yan, 2007). While high question sensitivity, on one the hand, encourages respondents to edit their answer and possibly give item nonresponse (e.g. Shoemaker et al., 2002), it, on the other hand, may encourage them to report an edited substantive answer (Olson et al., 2018) because providing an item nonresponse could be seen as admitting to having engaged in, for instance, a stigmatized behaviour such as smoking or drug consumption. Behind this inconclusive evidence, we expect that question sensitivity will not be associated with either type of item nonresponse.

The *question context* also influences the amount of item nonresponse (Bradburn et al., 2004; Dickinson & Kirzner, 1985). With respect to the position of the question within a questionnaire, two contradicting hypotheses exist (see Olson et al., 2018). First, due to respondents' fatigue (e.g. Gummer & Roßmann, 2015; Olson et al., 2018), questions with a later position in the questionnaire may have a higher likelihood of REF. Second, due to learning effects (Holbrook et al., 2016; Olson et al., 2018; Warren & Halpern-Manners, 2012), respondents may find the cognitive response process easier during later stages of the questionnaire, which may decrease the amount of DK answers. Behind this evidence, we expect to find more REF (due to respondents' fatigue) but fewer DK responses (due to learning effects) at later stages in the questionnaire.

2.2 | Respondent characteristics

While both the survey instruments and the interviewers impact the response process, it is the respondent who finally must evaluate whether he or she can provide a 'valid' answer and whether he or she is willing to do so. Consequently, respondent characteristics should play a major role in explaining item

nonresponses. Following the assumptions by the theory of survey satisficing (Krosnick, 1991), which states that respondents low in cognitive ability and motivation are more likely to shortcut cognitive processing and to provide item nonresponse, we included these factors on the respondents' level.

Respondent ability consists of the respondents' general cognitive ability as well as task-related abilities (Krosnick, 1991). General cognitive abilities have been referred to as cognitive sophistication (Krosnick, 1991) or working memory capacity (e.g. Knäuper, 1999). They refer to the general abilities to retrieve information from memory and to integrate the information into summary judgements. In contrast, task-related abilities refer, for instance, to prior topical knowledge (Roßmann, 2017; Roßmann et al., 2018). In line with the theoretical expectations, previous research has demonstrated that low respondent ability increases the chances of DK responses (e.g. Krosnick et al., 2002; Pickery and Loosveldt, 1998; Pickery et al., 2001), REF (e.g. Craig & McCann, 1978) or item nonresponse of both types (e.g. Lenzner, 2012; Yan & Curtin, 2010). Thus, we expect respondents' abilities to be negatively associated with the number of REF and DK responses meaning that high cognitive ability is likely to decrease item nonresponses.

Respondent motivation is likely to be influenced by a respondents' general tendency to enjoy effortful mental exercises, the saliency of a question's topic to the respondent, the perceived importance of a survey, as well as distractions and increasing fatigue in the course of an interview (Krosnick, 1991). Prior research has provided evidence in favour of the assumption that higher respondent motivation is linked to lower frequencies of DK responses (e.g. Krosnick et al., 2002; Pickery et al., 2001) or item nonresponse of both types (e.g. Lenzner, 2012; Roßmann et al., 2018). Thus, we expect that higher respondent motivation relates to a decreased probability of both DK and REF responses.

In addition, previous studies have reported findings on *interviewer observations* of respondents' behaviour during the survey interview, which assess respondents' ability and motivation and give further insights into the cognitive response process (see Holbrook et al. 2014). Specifically, interviewer evaluations of respondents' disengagement (an indicator for (low) motivation), respondents' comprehension problems (an indicator for (low) ability) and distraction during the interview are expected to be positively related to the number of item nonresponses. In contrast, the expectation of the association between respondents' reluctance and item nonresponse is less conclusive (see Olson, 2013). While some previous studies concluded that reluctant respondents do not provide worse data quality with respect to measurement than nonreluctant respondents (e.g. Hox et al., 2012; Kaminska et al., 2010), other studies presented evidence for a relation between unit and item nonresponse (e.g. Korkeila et al., 2001; Schmidt et al., 2005; Yan & Curtin, 2010). In their review of that literature, Olson (2013) saw evidence for a negative relationship between respondents' willingness to cooperate and the number of item nonresponses, so that we also assume that respondents, with whom it is difficult to gain cooperation, have a higher probability of REF. This, however, should not affect the likelihood of DK responses.

2.3 | Interviewer characteristics

In their systematic review of interviewer effects, West and Blom (2017) reported that the *experience of an interviewer* had a negative association with unit nonresponse. In contrast, they noted that the *age and gender of an interviewer* did not affect the amount of unit nonresponse substantially. With respect to item nonresponse, the study by Berk and Bernstein (1988) found evidence of a weak effect of interviewer age on DK responses and REF: Younger interviewers were associated with slightly higher rates of item nonresponse than older interviewers. However, Berk and Bernstein (1988) did not find evidence that other characteristics of the interviewers, as for instance, education or prior experience,

affected item nonresponse rates. These results received further support by Pickery et al. (2001), who found that neither interviewer age, sex, education and experience nor the number of interviews had significant effects on item nonresponse. Finally, Daikeler and Bosnjak (2020) showed in their meta-analysis that only specific aspects of interviewer trainings were successful in reducing the amount of item nonresponse. In summary, previous research suggests that the direct effects of interviewer characteristics on item nonresponse are rather limited. We expect that these findings generalize to both types of item nonresponse.

However, following sociological findings on homophily (McPherson et al., 2001), we assume that respondents may be more motivated to comply and interact with interviewers that they perceive as physically and intellectually similar to themselves. Thus, the similarity in age, gender and education between respondents and interviewers may motivate respondents to complete the cognitive response process carefully and, by that, diminish the likelihood of providing DK and REF answers.

3 | METHOD

3.1 | Data

The present study used data of the 2013 German Pre-Election Survey of the German Longitudinal Election Study (Rattinger et al., 2014), which was fielded between 29 July 2013 and 21 September 2013. The AAPOR (2016) Response Rate 5 was 32.1%. In total, 2,003 respondents answered the questionnaire in computer-assisted personal interviews (CAPI). Interviewers were instructed to stick to the wording of the questionnaire, refer to each question's interviewer instructions and not to mention 'don't know' categories explicitly. Yet, they were instructed to accept DK and REF as an answer, if given by a respondent. The questionnaire covered 214 questions on political attitudes and behaviours as well as the sociodemographic characteristics of the respondents. Overall, the respondents answered between 229 and 314 of the 397 question items ($M = 247.7$), depending on their individual skip patterns. The average interview duration was 65 min and 53 s. With respect to its questionnaire content and length, survey mode and survey design, we consider this survey to be representative of many large-scale surveys carried out in the field of social sciences.

For our analyses, the dataset was reshaped into the long format, resulting in a dataset on the item instead of the respondent level (i.e. wide format). Hence, each observation (row) is the answer to one question by one respondent. Finally, we merged the item-level dataset with additional information on each of the questions resulting in a final dataset of 545,178 observations.

3.2 | Measures

The dependent variable of our study was whether respondents provided item nonresponse (REF or DK) to the questions in the survey. These response outcomes were then linked to the characteristics of the questions, respondents and interviewers.

Item nonresponse. For each question, we coded whether respondents provided a substantive response or a nonresponse by saying that they refuse to answer that question or by saying that they cannot give an answer because they do not know (substantive response, DK, REF). On average, respondents did not respond substantively to 6.5% of the questions (REF = 1.7% and DK = 4.8%).

Question characteristics. We developed a coding scheme for the characteristics of the 214 questions in the questionnaire, which is based on categories derived from theory and previous studies

(see Table A1 in the Online Appendix for a detailed description of the categories). Each question was independently coded by three coders—who were research assistants with a background in social sciences—regarding its type, format, difficulty, sensitivity and context (see Rattinger et al., 2014, for the wording of all questions in the questionnaire). The final coding scheme and instructions for coders are provided in Table A2 in the Online Appendix. If all three coders agreed in their assessment, the coding was assigned as the final code. If the three coders did not agree in their assessments, a decision on the final code was reached in an expert assessment among the authors. On average, in the initial round of coding, across all question characteristics, an agreement of 84% between all three coders was reached (see Table A3 in the Appendix for more details on inter-coder agreement). For those remaining cases, the authors considered question specifics, interpretations by each coder, partial agreement between two coders and refined the coding scheme to cover more specific categories to even fit questions that resulted in disagreement (see Bais et al., 2019). Based on the refined coding scheme, final codes for all instances in which disagreement between the three coders existed were decided upon by the authors.

Building on the fact-attitude terminology for survey *question types* (Tourangeau et al. 2000), we distinguished between questions about attitudes, facts, knowledge, behaviours and other questions (i.e. all remaining questions that did not fit into the classification, for instance, questions on personality and emotions). We differentiated between four common *question formats*: closed, semi-closed (i.e. closed questions with an open-ended 'other' response option), open-ended and multiple-choice questions. To closed, semi-closed and open-ended, only a single response can be provided, whereas for multiple-choice questions, several of the response options can be chosen. *Question difficulty* was measured by the number of words in a question, and by whether the question was hypothetical, a behavioural intention or a recall. In addition, it was determined whether the question was part of an item battery, and in that case, the total number of items in that battery was counted. *Question sensitivity* was measured in two categories (0 = 'nonsensitive', 1 = 'sensitive'). Examples of sensitive questions include questions asking about a respondent's vote choice or income. *Question context* was measured by its position in the questionnaire. A higher number refers to a later placement. Individual skip patterns were not considered.

Respondent characteristics. Our analyses included three measures of respondents' ability, two measures of respondents' motivation and four measures of interviewer observations (see Online Appendix Table A4 for the question wordings). General *cognitive ability* was measured by three binary indicators for low, medium or high levels of education (see e.g. Krosnick et al., 2002), and age in decades (see e.g. Knäuper, 1999; Olson et al., 2018). In line with Roßmann et al., (2018), we used respondents' political knowledge (proportion of correct answers to five knowledge questions) as an indicator for *task-related ability* (see also Lenzner et al., 2010; Roßmann, 2017). Referring to Roßmann et al., (2018), we used political interest (asked using a five-point rating scale from 0 = 'not at all' to 1 = 'very strongly') as a measure for *respondents' motivation* (see also Pickery et al., 2001). As we further expected that involvement in the survey's topic increases motivation (Roßmann, 2017), we included a dichotomous indicator for identification with a political party (0 = 'low', 1 = 'high'). In our analyses, we used four measures of *interviewer observation* on the respondents' answer behaviour (see also Holbrook et al., 2014; Kaminska et al., 2010). Based on these, we computed evaluations of the difficulty to gain cooperation with a respondent, the level of a respondent's disengagement, frequencies of comprehension difficulties and the frequency of distractions during an interview (all asked using a four- or five-point rating scale from 0 = 'low' to 1 = 'high'). We included gender (0 = 'male', 1 = 'female') as an additional variable on the respondent level.

Interviewer characteristics. Data on the interviewers included sociodemographic characteristics (i.e. age, gender and education) as well as information on what kind of survey-specific training an

interviewer received. This information was provided by the field institute, which conducted the survey. On average, each of the 183 interviewers conducted 10.95 interviews.

An interviewer's *cognitive ability* was measured by three binary indicators for low, medium or high levels of education, and by age in decades. *Task-related ability* was measured by the training of an interviewer (0 = 'by phone', 1 = 'in person'). The study included gender (0 = 'male', 1 = 'female') as an additional variable on the interviewer level.

Cross-level interactions. We included three measures that crossed the respondent and the interviewer level. First, a binary measure indicates whether a respondent and an interviewer had the *same gender* or not (0 = 'different gender', 1 = 'same gender'). Second, a measure was calculated as the squared *age difference* (in decades) between the respondent and the interviewer for each survey interview. To obtain additional information on the age difference, we also included a dummy variable that indicated whether the respondent was older than the interviewer. Third, to measure *education differences*, we included all possible education combinations between respondents and interviewers using low education in both cases as the reference category.

Measurement endogeneity. While all question and interviewer characteristics are exogenous measures, meaning that they were not influenced by the survey response process, some measures of respondents' characteristics were derived during the response process. This is especially the case for all interviewer observations, which were based on the evaluations of the respondents' behaviour during the interview. Specially, interviewer observation of respondents' disengagement, comprehension difficulty and distraction might be directly related to both types of nonresponse (see Kirchner et al., 2017), so that those observations might be—at least partially—a result of respondents' item nonresponse behaviour during the interview. While we acknowledge that those measures are endogenous and are, therefore, problematic, we see value in including them in the analysis because they provide additional insights and alternative measures were not available. To ensure that including those measures does not alter the final model with respect to the remaining variables, we provide the results without those observations in Table A5 in the Online Appendix as a robustness check. In fact, the conclusions for the other variables would not have changed if the interviewer observations would have been omitted.

Missingness. All explanatory variables had below 1% missing responses, so that we used complete case analysis.

3.3 | Model structure and estimation

Our variable of interest y has three possible outcomes $c = 1, 2, 3$ for substantive response, DK or REF respectively. We use y_{kqi} to denote the value of y for the kth respondent, answering the qth question, asked by the ith interviewer.

We formulate our model for y_{kqi} as a multinomial logistic regression model, that is, we have:

$$\phi_{kqi,c} = Pr y_{kqi} = c | x_{kqi}, z_{kqi} \ c = 1, 2, 3, \tag{1}$$

where x_{kqi} is the vector of the values for all measures mentioned in Section 3.2 and z_{kqi} is the vector of indicators, for the kth respondent, the qth question and the ith interviewer.

Our assumed model is:

$$log\left(\frac{\phi_{kqi,c}}{\phi_{kqi,1}}\right) = x_{kqi}\beta_c + u_{q,c} + v_{i,c} + w_{ki,c} \ c = 2, 3, \tag{2}$$

where we have fixed effects given by β_c, which is the vector of regression coefficients for the included measures for outcome c. For outcome c, we also include random effects in the model which we denoted by $u_{q,c}$ for the random effect related to question q, $v_{i,c}$ for the random effect related to interviewer i, $w_{ki,c}$ for the random effect related to respondent k interviewed by the ith interviewer. We assumed that the random effects in Equation (2) are all mutually independent of each other, that is, we assumed for all respondents, questions and interviewers: $u_{qc} \sim N(0, \sigma_{u,c}^2)$, $v_{ic} \sim N(0, \sigma_{v,c}^2)$, $w_{kic} \sim N(0, \sigma_{w,c}^2)$, that is, we assume normal distributions with zero mean and variances $\sigma_{u,c}^2$, $\sigma_{v,c}^2$ and $\sigma_{w,c}^2$ for random effects related to question, interviewer and respondent–interviewer interaction respectively. Based on the assumed random effect variances, we find the following covariance structure for terms on the right-hand side of Equation (2):

$$
\begin{cases}
\sigma_{u,c}^2 & \text{for } k \neq l, \ i \neq j, \ q = p \\
\sigma_{u,c}^2 + \sigma_{v,c}^2 & \text{for } k \neq l, \ i = j, \ q = p \\
\sigma_{w,c}^2 + \sigma_{v,c}^2 & \text{for } k = l, \ i = j, \ q \neq p \quad , \ c = 2, 3. \\
\sigma_{v,c}^2 & \text{for } k \neq l, \ i = j, \ q \neq p \\
0 & \text{else}
\end{cases}
$$

Our hypothesis is that observations from the same respondents are related to each other, as well as observations for the same questions. As each respondent answered multiple questions, and as multiple respondents participated in the survey, we decided to model the effects of respondent and question characteristics with cross-classified random effects (Goldstein, 1994; Rasbash & Goldstein, 1994). Further, we assumed that measures taken by the same interviewer are also related to one other. Since an interviewer can interview multiple respondents, but one respondent is only interviewed by a single interviewer, there is a hierarchical structure between respondents and interviewers. Hence, the random effect that is included for the respondents is nested within the interviewers.

It is possible to represent our logistic model in Equation (2) also as a so-called latent variable model (Hedeker, 2003), with a residual error that follows a logistic distribution with mean zero and scale one and has thus a constant variance of $\pi^2/3$ (Train, 2009). In combination with the covariance structure of our model, we define the following intraclass correlation coefficients (ICCs) for respondents ICC_{Rc}, questions ICC_{Qc} and interviewers ICC_{Ic} as:

$$ICC_{Rc} = \rho_{wc} + \rho_{vc}, \ ICC_{Qc} = \rho_{uc}, \ ICC_{Ic} = \rho_{vc}, \tag{3}$$

$c = 2,3$

where, $\sigma_c^2 = \sigma_{u,c}^2 + \sigma_{v,c}^2 + \sigma_{w,c}^2 + \frac{\pi^2}{3}$, $\rho_{u,c} = \frac{\sigma_{u,c}^2}{\sigma_c^2}$, $\rho_{v,c} = \frac{\sigma_{v,c}^2}{\sigma_c^2}$, $\rho_{w,c} = \frac{\sigma_{w,c}^2}{\sigma_c^2}$, for c = 2, 3. That is, with the ICCs in Equation (3) we are interested in measuring the similarity between two observations of the same outcome, for the same respondent, the same question or the same interviewer respectively (Hox, 2010, p. 34).

Because of technical limitations, we could not estimate our model in Equation (2) in its multinomial form (see Online Appendix, Section B). However, Begg and Gray (1984) show that a multinomial logit model can also be formulated as a set of binary logistic regressions. For this, we condition $\phi_{kqi,c}$ in Equation (1), also on $I\left(y_{kqi} = c\right) + I\left(y_{kqi} = 1\right) = 1$ for $c = 2, 3$, that is, we consider only observations for which we have outcome c or the reference outcome 1, 'substantive response'. Then, we can estimate our model in Equation (2), by fitting two separate logistic regression for outcome DK and REF. Although Begg and Gray (1984) do not cover random effect models, their method is equally applicable for our model, which is clear from the latent variable representation of our model (Hedeker, 2003). Hence, we can assume the covariance structure as described above, as we would for a model

with a continuous response, apart from the residual variance (for further details on estimation, see Online Appendix, Section B).

4 | RESULTS

We first discuss the ICCs (see Equation (3)) obtained from estimating the full cross-classified random effect model (see Equation (2)), including all fixed effects. Then, the results are presented separately for REF and DK answers. Afterwards, the results for the two item nonresponse types are compared with another.

For DK answers, the ICC on the question level was higher than for REF ($ICC_{Q\ DK} = 0.45$; $ICC_{Q\ REF} = 0.24$). In contrast, the ICC on the interviewer and respondent level was higher for REF than for DK answers (interviewer: $ICC_{I\ DK} = 0.11$; $ICC_{I\ REF} = 0.21$; respondent: $ICC_{R\ DK} = 0.18$; $ICC_{R\ REF} = 0.43$). Two conclusions follow from these findings: First, the decision for DK was more strongly influenced by a question itself, rather than by the different respondents or interviewers. Second, we can infer that the decision for REF is more strongly influenced by the interviewer or respondent than by the individual question. These results support the assumption that DK and REF are the results of distinct cognitive processes. It can be noted that the ICCs do not change in order or substantially in magnitude if they are estimated for the so-called null model, that is, estimation of our model without the fixed effects.

To give some perspective on the absolute size of the ICCs, we compared them to the ICCs for interviewers estimated for the European Social Survey (ESS) Rounds 1 to 6 (Beullens & Loosveldt, 2016). In the ESS, most countries had an ICC of around 0.1 or below, and values above 0.2 were considered as very high. Thus, the ICCs that we measure seem to be within the midrange and on the higher end for DK or REF, respectively, compared to the interviewer ICCs of the ESS. Furthermore, we consider the ICCs for DK on the question level (0.45) and for REF on the respondent level (0.43) as very high.

4.1 | Results of the analyses for REF

On the level of question characteristics, the results for REF were mixed (see Table 1). With respect to question type, we found a significantly lower probability of REF for questions about facts than for attitudinal questions, while there were no significant differences for questions about knowledge and behaviour. Of those, only the nonsignificant effect of behavioural questions was in line with our expectations. Question difficulty showed a significant effect in the expected direction for the length of the wording with a higher number of words in the question stem resulting in a higher probability of REF. Also in line with our expectations, we did not observe a significantly higher probability of REF for hypothetical questions, behavioural intentions and recalls. As expected with respect to the question format, we found that semi-closed and multiple-choice questions elicited a significantly higher probability of REF than closed questions. Also as expected, higher question sensitivity did not result in a higher probability of REF, and later placement in the questionnaire led to a higher likelihood of REF.

Our results were also mixed on the level of respondent characteristics. With respect to respondents' general ability, we did not observe that the probability of REF was higher among respondents with lower levels of education or older respondents. In contrast, we found evidence in favour of the assumed effect of task-related ability on REF: The probability of REF was lower among respondents with higher political knowledge. As expected with respect to respondents' motivation, we found that

respondents who were involved in the survey's topic—as measured by identification with a political party—had a lower probability of REF. However, we did not observe a corresponding effect of political interest on the likelihood of REF. In conformance with our expectations, we found that reluctant and disengaged respondents tend to have a higher probability of REF. For the remaining two interviewer observations on question comprehension difficulty and distractions, we did not observe the expected effects on REF.

On the interviewer level, all effects conformed to our expectations. With respect to interviewer ability, our measures of education, age and training did not show significant effects, thereby supporting the findings by Pickery et al. (2001). In contrast to our theoretical expectations, the three cross-level interactions for similarity in gender, age or education did not elicit significant effects on the probability of REF.

4.2 | Results of the analyses for DK

On the level of question characteristics, the results for DK responses were mixed (see Table 1). With respect to question type, we found the expected significantly lower probability of DK answers for questions about facts and behaviours than for attitudinal questions. However, we did not find a significant effect of knowledge questions on the likelihood of DK. Four of the five measures of question difficulty showed the expected positive effects on the likelihood of DK responses. Specifically, a higher number of items in a battery, hypothetical questions, behavioural intentions and recalls all resulted in a higher probability of DK answers, whereas a higher number of words in the question stem did not show a significant effect. With respect to the question format, we did not find the expected effects on DK answers. The nonsignificant effect for question sensitivity was in line with our hypothesis, whereas we did not find the expected negative relationship between question position and DK.

Moving to the level of respondent characteristics, we found that the indicators for respondents' task-specific—but not general—abilities exerted the expected effect on the likelihood of DK responses: The chance of a DK answer was not significantly higher for respondents with a lower level of education or older respondents, whereas respondents with a higher level of political knowledge were less likely to give DK answers. With respect to respondents' motivation and involvement, higher political interest and identification with a political party elicited the expected lower probability of DK answers. Of the four interviewer observations, disengagement and question comprehension difficulty exerted the expected increase on the probability of DK answers. The remaining two interviewer observations on cooperation difficulty and distractions during the interview had no significant effects on this outcome. While we did not expect that the difficulty to gain cooperation with a respondent would affect the likelihood of DK responses, the lacking effect of respondents' distractions on DK answers stood in contrast to our assumption.

With respect to the interviewer level, we found that older interviewers had an increased likelihood of prompting DK answers and that a larger age difference between a respondent and an interviewer also increased the likelihood of DKs. Thus, our study provides some evidence that interviewer ability can affect the likelihood of DK responses. An interviewer and a respondent having the same gender did not show a significant effect on the probability of DK answers, nor did the education difference between respondent and interviewer show a significant effect. These results may suggest that respondents feel more comfortable admitting that they do not know an answer if the interviewer is older, especially if the age difference is large, or that older interviewers were more likely to prompt DK answers.

TABLE 1 Results of the cross-classified random effect model for item nonresponse

Characteristics on three levels	Item refusals (REF)			Don't know (DK)		
Fixed effects	OR	AME	Expectation	OR	AME	Expectation
Question characteristics						
Question type (reference: attitudinal)						
Factual	0.399**	−0.0118	+	0.048***	−0.1009	−
Knowledge	0.533	−0.0084	+	0.751	−0.0095	−
Behavioral	0.919	−0.0011	0	0.050***	−0.0999	−
Other	0.691	−0.0047	none	0.089***	−0.0806	none
Question difficulty						
Number of words	1.020**	0.0003	+	0.986	−0.0004	+
Number of items (battery)	1.027	0.0003	+	1.101*	0.0032	+
Hypothetical	0.940	−0.0008	0	27.956***	0.1110	+
Intention	0.414	−0.0114	0	34.337***	0.1179	+
Recall	0.978	−0.0003	0	6.734***	0.0636	+
Question format (reference: closed)						
Semi-closed	4.688**	0.0199	+	2.256	0.0271	+
Open-ended	1.246	00.0028	+	1.077	0.0025	+
Multiple-choice	66.985***	0.0541	+	1.746	0.0186	+
Question sensitivity	2.157	0.0099	0	0.688	−0.0125	0
Question position	1.006**	0.0001	+	0.999	−0.0001	−
Respondent characteristics						
Respondent ability						
Education (reference: low)						
Medium	1.246	0.0029	−	0.865	−0.0014	−
High	1.376	0.0022	−	0.794	−0.0060	−
Age	1.055	0.0007	+	1.028	0.0005	+
Political knowledge	0.260***	−0.0173	−	0.167***	−0.0596	−
Respondent motivation						
Political interest	1.329	0.0037	−	0.368***	−0.0333	−
Party identification	0.492***	−0.0091	−	0.695***	−0.0121	−
Interviewer observations						
Cooperation difficulty	1.516*	0.0053	+	0.959	−0.0014	0
Disengagement	1.948**	0.0086	+	1.293*	0.0086	+
Comprehension difficulty	1.538	0.0055	+	1.994***	0.0230	+
Distraction	1.430	0.0046	+	1.155	0.0048	+
Additional measure						
Female	0.975	−0.0003	0	1.250***	0.0074	0

(Continues)

TABLE 1 (Continued)

Characteristics on three levels	Item refusals (REF)			Don't know (DK)		
Fixed effects	**OR**	**AME**	**Expectation**	**OR**	**AME**	**Expectation**
Interviewer characteristics						
Interviewer ability						
Education (reference: low)						
Medium	1.139	0.0010	0	0.822	−0.0039	0
High	0.805	−0.0031	0	1.012	0.0019	0
Age	1.033	0.0004	0	1.180*	0.0059	0
Training (reference: in person)						
Telephone	0.870	−0.0018	0	1.296	0.0086	0
Additional measure						
Female	1.320	0.0036	0	1.090	0.0029	0
Cross-level interactions (respondent-interviewer)						
Gender						
Same gender	1.129	0.0016	–	0.9271	−0.0025	–
Age						
$Age_{resp} > Age_{int}$	1.017	0.0002	none	1.064	0.0021	none
$(Age_{resp} - Age_{int})^2$	0.996		+	1.011*		+
Education (Reference: low)						
$Medium_{resp}*Medium_{int}$	1.069		0	1.322		0
$High_{resp}*Medium_{int}$	0.716		+	0.894		+
$Medium_{resp}*High_{int}$	0.963		+	1.003		+
$High_{resp}*High_{int}$	0.952		0	1.198		0
Variances of random effects						
Question level						
σ^2_{uREF} 2.400						
σ^2_{uDK} 3.910						
Respondent level (nested in interviewers)						
σ^2_{wREF} 2.142						
σ^2_{wDK} 0.613						
Interviewer level						
σ^2_{vREF} 2.081						
σ^2_{vDK} 0.941						

Notes: OR = odds ratios; AME = average marginal effects; + = positive association expected; − = negative association expected; 0 = no association expected; none = no expectation. We omitted σ^2_e and R^2 statistics because of unclear definitions of the quantities in that model.

The ORs show high values due to the small number of occurrences for semi-closed and multiple-choice questions with respect to REF, and hypothetical, intention and recall questions with respect to DK responses. We recommend considering the AMEs for a substantial interpretation of the effects.

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

4.3 | Comparison between the results for REF and DK

When comparing the results for REF and DK answers, we found many similarities but also differences in their connection to explanatory characteristics on the question, respondent and interviewer level. On the level of question characteristics, only questions about facts affected both types of item nonresponse similarly. Number of words in a question, semi-closed questions, multiple-choice questions and the question position affected the probability of REF responses but not the likelihood of DKs, whereas behavioural questions, number of items in a battery, hypothetical questions, behavioural intentions and recalls had an impact on DK responses but did not influence the likelihood of REF.

On the level of respondents' characteristics, we found significant effects in the same direction for political knowledge and party identification. Notably, respondent's education and age did not show a significant effect on the likelihood of REF nor DK answers. Of the four interviewer observations, respondents' disengagement was connected to both types of item nonresponses, while the difficulty to gain cooperation was connected to REF and comprehension difficulty to DK answers.

On the level of interviewer characteristics, all our measures did not exert effects on the probability of REF. For DK answers, the age of an interviewer and the age difference between interviewers and respondents had a significant effect.

5 | DISCUSSION

Building on earlier research (e.g. Beatty & Herrmann, 2002; Juster & Smith, 1997; Shoemaker et al., 2002), the present study set out to examine and distinguish between two types of item nonresponse—REF and DK—by proposing explanatory variables on three levels: the question, the respondent and the interviewer level. Overall, the study showed that the characteristics of the questions and the respondents affected both types of item nonresponses, whereas interviewer characteristics only affected the likelihood of DK answers. The relevance of acknowledging the different explanatory levels is also reflected in the predominantly high ICCs of the cross-classified model. Of the explanatory variables, some variables showed similar effects for REF and DK answers (e.g. factual questions and political knowledge), whereas other variables exerted different effects on the two types of item nonresponse (e.g. hypothetical questions and question position).

On the level of question characteristics, only factual questions showed a lower probability of REF as well as DK answers. However, a variety of other measures (e.g. number of words in the question stem) showed different effects on DK and REF. Comparing our results with findings of previous studies, our study reproduced the positive association between question difficulty and both types of item nonresponse (e.g. Krosnick, 1991; Shoemaker et al., 2002), but not the positive association of question sensitivity and REF (Shoemaker et al., 2002). The inclusion of five indicators of question difficulty allowed us to study their different effects on REF and DK responses. Comprehension difficulty measured by the number of words and the length of an item battery seems to increase the probability of REF and DK, whereas DK responses are seemingly more likely in the case of retrieval difficulties, especially when a question is hypothetical, asks about a behavioural intention or requests the respondent to recall information from memory.

On the level of respondent characteristics, we found that political knowledge and party identification both reduced the probability of item nonresponse. In addition, a higher interest in politics led to a significantly lower likelihood of DK answers. The finding that both types of item nonresponse were associated to respondents' ability and motivation is in line with Krosnick (1991), who suggested that these two aspects are important in understanding the occurrence of nonoptimal

responses. With respect to interviewer observations, our study found that the difficulty of gaining cooperation with respondents was positively related to the likelihood of REF. This result contradicts findings from previous studies by Hox et al. (2012) and Kaminska et al. (2010) on the response quality of reluctant respondents, which suggested that there would be no relationship. However, the result is in line with findings of other studies (e.g. Korkeila et al., 2001; Schmidt et al., 2005; Yan & Curtin 2010), which provided evidence in favour of a positive relationship between unit and item nonresponse propensities.

On the level of interviewer characteristics, most of the indicators of interviewers' ability did not have an impact on either type of item nonresponse. However, the measures age and age difference influenced the number of DK answers, but they did not influence the number of REF. The absence of an effect of interviewers' ability for REF is opposite to the reasoning by Japec (2008), who suggested that interviewers' ability influences the responses during a survey interview. However, our results are in line with the conclusions by Berk and Bernstein (1988) and Pickery et al. (2001), who suggested that interviewer performance can hardly be predicted on the basis of basic sociodemographic characteristics, as, for instance, interviewer sex, age or education.

The results of the present research complement and extend those of Olson et al. (2018), who investigated the impact of question, respondent and interviewer characteristics on response behaviour in a telephone survey with an average duration of about 15 min. Similar to the study by Olson et al. (2018), we found that those three levels contribute to explaining why respondents choose to provide nonsubstantive responses in a face-to-face population survey. While Olson and her colleagues have set out to better understand the response process by coding each response into one of six different response behaviours ('adequate response', 'qualified response', 'uncodable response', 'don't know', 'refusal' and 'clarification request'), our study differentiated between three response behaviours ('substantive response', 'item refusal' and 'don't know'). The focus on item nonresponse allowed us to more closely investigate similarities and differences between those two types of (non)response behaviours.

5.1 | Limitations

Our study has certain limitations. First, the data for the present study were collected by interviewers as CAPI and the results are consequently limited to this mode of administration. REF and DK answers may be affected differently in other modes of data collection. Therefore, future studies could investigate to which degree our findings can be generalized to self-administered surveys. Second, our results are based on a single dataset. However, because our analytical strategy made use of all responses on the question level, which increased the number of available cases compared to analyses on the respondent level, we expect that many of our results are likely to be reproducible in other interviewer-administered surveys. Third, the present survey was part of the German Longitudinal Election Study and, thus, had a specific topic. Future research could replicate our analyses by using surveys on different topics and in other national or international contexts. Fourth, our dataset did not allow us to examine whether there was a clarification request or whether an initial REF/DK was converted into a 'valid' answer. Looking closer into the interaction between respondents and interviewers during the survey interview will likely help to shed more light on the origins of item nonresponse. Lastly, the data collection included only limited paradata. Thus, future research could explore the impact of response latencies, change patterns and other paradata on the probability of item nonresponse.

5.2 | Conclusion

Altogether, we tested hypotheses regarding characteristics of the question, the respondent and the interviewer based on theory and previous findings and showed that item nonresponse is related to each of the three levels. Most measures on the question level (question type, format, difficulty and context) related to item nonresponse. The key variables on the respondent level were respondents' ability and motivation, and the key variable on the interviewer level was an interviewer's age. By that the present study confirms that our knowledge about the cognitive response process and its implications regarding information processing when answering survey question is quite sensible. This leads us to remind survey designers on classic recommendations regarding task difficulty and cognitive effort (e.g. Krosnick, 1991; Saris & Gallhofer, 2007; Bradburn et al., 2004), such as that the number and the length of batteries in a questionnaire should be minimized, multiple-choice questions should be avoided if possible and the interview length should be kept at a minimum to reduce response burden, and, as a consequence, the probability of item nonresponse. With respect to REF, researchers should be especially attentive when it is difficult to gain respondents to cooperate. With respect to DK responses, researchers should be particularly attentive when designing questions about hypothetical content, intentions or questions that require respondents to recall information from their memories. Regarding this, interviewer observations on respondents having a hard time to comprehend a question may provide valuable information on problematic questions. Also, interviewer-administered surveys offer the opportunity to specifically train interviewers to implement strategies that aim at mitigating item nonresponse such as repetition of the question if needed.

ORCID

Henning Silber https://orcid.org/0000-0002-3568-3257
Joss Roßmann https://orcid.org/0000-0003-2701-0690
Tobias Gummer https://orcid.org/0000-0001-6469-7802

REFERENCES

AAPOR. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (9th ed.). The American Association for Public Opinion Research. https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf. [Accessed 5 May 2021].

Bais, F., Schouten, B., Lugtig, P., Toepoel, V., Arends-Tòth, J., Douhou, S., Kieruj, N., Morren, M., & Vis, C. (2019). Can survey item characteristics relevant to measurement error be coded reliably? A case study on 11 Dutch General Population Surveys. *Sociological Methods and Research*, 48(2), 263–295.

Beatty, P., & Herrmann, D. (2002) To answer or not to answer: Decision processes related to survey item nonresponse. In: Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, J. R. (Eds.) *Survey nonresponse*. NY: John Wiley & Sons, pp. 71–85.

Begg, C. B., & Gray, R. (1984) Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika*, 71(1), 11–18.

Berk, M. L., & Bernstein, A. B. (1988) Interviewer characteristics and performance on a complex health survey. *Social Science Research*, 17, 239–251.

Beullens, K., & Loosveldt, G. (2016) Interviewer effects in the European Social Survey. *Survey Research Methods*, 10(2), 103–118.

Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality* (Vol. 335). NY: John Wiley & Sons.

Bishop, G. F., Tuchfarber, A. J., & Oldendick, R. W. (1986) Opinions on fictitious issues: The pressure to answer survey questions. *Public Opinion Quarterly*, 50(2), 240–250.

Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design - For market research, political polls, and social and health questionnaires*. San Francisco: John Wiley & Sons.

Coombs, C. H., & Coombs, L. C. (1976) "Don't Know" item ambiguity or respondent uncertainty? *Public Opinion Quarterly*, 40(4), 497–514.

Craig, C. S., & McCann, J. M. (1978) Item nonresponse in mail surveys: extent and correlates. *Journal of Marketing Research*, 15(2), 285–289.

Daikeler, J., & Bosnjak, M. (2020). How to conduct effective interviewer training: A meta-analysis and systematic review. In: Olson, K., Smyth, J. D., Dykema, J., Holbrook, A. L., Kreuter, F. & West, B. T.(Eds) *Interviewer Effects from a Total Survey Error Perspective*. Boca Raton: CRC Press, pp. 47–62.

Dahlhamer, J. M., Maitland, A., Ridolfo, H., Allen, A. & Brooks, D. (2020) Exploring the associations between question characteristics, respondent characteristics, interviewer performance measures, and survey data quality. In: Beatty, P. C., Collins, D., Kaye, L., Padilla, J. L., Willis, G. B. & Wilmot, A. (Eds.) *Advances in Questionnaire Design*. Development, Evaluation and Testing, Hoboken, NJ: Wiley, pp. 153–192.

DeMaio, T. J. (1984) Social desirability and survey measurement: A review. *Surveying Subjective Phenomena*, 2, 257.

Dickinson, J. R., & Kirzner, E. (1985) Questionnaire item omission as a function of within-group question position. *Journal of Business Research*, 13(1), 71–75.

Dykema, J., Schaeffer, N. C., & Garbarski, D. (2016). Introduction to special issue of survey practice on interviewer-respondent interaction. *Survey Practice*, 9(2), 1–3.

Dykema, J., Schaeffer, N. C., Garbarski, D. & Hout, M. (2020) The role of question characteristics in designing and evaluating survey questions. In: Beatty, P. C., Collins, D., Kaye, L., Padilla, J. L., Willis, G. B. & Wilmot, A. (Eds.) *Advances in questionnaire design, development, evaluation and testing*. NJ: Wiley, pp. 117–152.

Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods and Research*, 22(3), 364–375.

Groves, R. M. (2004). *Survey errors and survey costs* (Vol. 536). NY: John Wiley & Sons.

Gummer, T., & Roßmann, J. (2015). Explaining interview duration in web surveys: A multilevel approach. *Social Science Computer Review*, 33(2), 217–234.

Hedeker, D. (2003). A mixed-effect multinomial logistic regression model. *Statistics in Medicine*, 22(9), 1433–1446.

Holbrook, A. L., Anand, S., Johnson, T. P., Cho, Y. I., Shavitt, S., Chávez, N., & Weiner, S. (2014). Response heaping in interviewer-administered surveys: is it really a form of satisficing? *Public Opinion Quarterly*, 78(3), 591–633.

Holbrook, A., Cho, Y. I., & Johnson, T. (2006). The impact of question and respondent characteristics on comprehension and mapping difficulties. *International Journal of Public Opinion Quarterly*, 70(4), 565–595.

Holbrook, A. L., Weiner, S., Chavez, N., Shavitt, S., Cho, Y. I., & Johnson, T. P. (2016). Do interviewer errors help explain the impact of question characteristics on respondent difficulties? *Survey Practice*, 9(2), 1–11.

Hox, J. J. (2010). *Multilevel analysis: Techniques and Applications* (2nd ed.). New York: Routledge.

Hox, J. J., De Leeuw, E. D., & Chang, H. T. (2012) Nonresponse versus measurement error: Are reluctant respondents worth pursuing? *Bulletin de Méthodologie Sociologique*, 113(1), 5–19.

Japec, L. (2008). Interviewer error and interviewer burden. In Lepkowski, J. M., Tucker, C., Brick, J. M., de Leeuw, E. D., Japec, L., Lavrakas, P. J., Link, M. W. & Sangster, R. L. (Eds.) *Advances in telephone survey methodology*. NJ: Wiley, pp. 187–211.

Juster, F. T., & Smith, J. P. (1997). Improving the quality of economic data: Lessons from the HRS and AHEAD. *Journal of the American Statistical Association*, 92(440), 1268–1278.

Kaminska, O., McCutcheon, A. L., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, 74(5), 956–984.

Kirchner, A., Olson, K., & Smyth, J. D. (2017). Do Interviewer postsurvey evaluations of respondents' engagement measure who respondents are or what they do? A behavior coding study. *Public Opinion Quarterly*, 81(4), 817–846.

Knäuper, B. (1999). The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly*, 3(1), 347–370.

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H., & Koskenvuo, M. (2001). Nonresponse and related factors in a nation-wide health survey. *European Journal of Epidemiology*, 17(11), 991–999.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., et al. (2002) The impact of "no opinion" response options on data quality. *Public Opinion Quarterly*, 66(3), 371–403.

Lenzner, T. (2012). Effects of survey question comprehensibility on response quality. *Field Methods*, 24(4), 409–428.

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, 24(7), 1003–1020.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415–444.

Olson, K. (2013). Do non-response follow-ups improve or reduce data quality? A review of the existing literature. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1), 129–145.

Olson, K., Smyth, J. D., & Ganshert, A. (2018). The effects of respondent and question characteristics on respondent answering behaviors in telephone interviews. *Journal of Survey Statistics and Methodology*, 7, 275–308.

Pickery, J., & Loosveldt, G. (1998). The impact of respondent and interviewer characteristics on the number of "No Opinion" answers. *Quality & Quantity*, 32, 31–45.

Pickery, J., Loosveldt, G., & Carton, A. (2001). The effects of interviewer and respondent characteristics on response behavior in panel surveys: A multilevel approach. *Sociological Methods & Research*, 29(4), 509–523.

Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, 19(2), 337–350.

Rattinger, H., Roßteutscher, S., Schmitt-Beck, R., Weßels, B., & Wolf, C.. (2014). Pre-election Cross-section (GLES 2013). GESIS Data Archive, Cologne. ZA5700 Data file Version 2.0.0, doi: https://doi.org/10.4232/1.12000.

Roßmann, J. (2017). *Satisficing in Befragungen: Theorie, Messung und Erklärung*. Berlin: Springer-Verlag.

Roßmann, J., Gummer, T., & Silber, H. (2018). Mitigating satisficing in cognitively demanding grid questions: evidence from two web-based experiments. *Journal of Survey Statistics and Methodology*, 6(3), 376–400.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. NY: John Wiley & Sons.

Schmidt, J. B., Calantone, R. J., Griffin, A., & Montoya-Weiss, M. M. (2005). Do certified mail third-wave follow-ups really boost response rates and quality? *Marketing Letters*, 16(2), 129–141.

Schuman, H., & Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. NY: Academic Press.

Shoemaker, P. J., Eichholz, M., & Skewes, E. A. (2002). Item nonresponse: Distinguishing between don't know and refuse. *International Journal of Public Opinion Research*, 14(2), 193–201.

Singer, E., & Couper, M. P. (2017). Some methodological uses of responses to open questions and other verbatim comments in quantitative surveys. *Methods, Data, Analyses*, 11(2), 115–134.

Singer, E., & Kohnke-Aguirre, L. (1979). Interviewer expectation effects: A replication and extension. *Public Opinion Quarterly*, 43(2), 245–260.

Toepoel, V., & Dillman, D. A. (2011). How visual design affects the interpretability of survey questions. In Das, M., Ester, P. & Kaczmirek, L. (Eds.) *Social and behavioral research and the internet. Advances in applied methods and research strategies*. NY: Routledge, pp. 165–190.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, England: Cambridge University Press.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883.

Train, K. (2009). *Discrete choice methods with simulations* (2nd ed.). Cambridge, England: Cambridge University Press.

Warren, J. R., & Halpern-Manners, A. (2012). Panel conditioning in longitudinal social science surveys. *Sociological Methods & Research*, 41(4), 491–534.

West, B. T., & Blom, A. G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, 5(2), 175–211.

Yan, T., & Curtin, R. (2010). The relation between unit nonresponse and item nonresponse: a response continuum perspective. *International Journal of Public Opinion Research*, 22(4), 535–551.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.