methods, data, analyses | Vol. 14(1), 2020, pp. 3-34 DOI: 10.12758/mda.2019.05

Effects of Respondent and Survey Characteristics on the Response Quality of an Open-Ended Attitude Question in Web Surveys

Katharina Schmidt, Tobias Gummer & Joss Roßmann
GESIS – Leibniz Institute for the Social Sciences

Abstract

Open-ended questions have a great potential for analyses, but answering them often imposes a great burden on respondents. Relying on satisficing theory as an overarching theoretical framework, we derived several hypotheses about how respondent and survey level characteristics, and their interactions, might affect the quality of the responses to an open-ended attitude question in self-administered surveys. By applying multilevel analyses to data from 29 web surveys, we examined the effects of respondent and survey level characteristics on three indicators of response quality: response length, response latency, and the interpretability of the answers. With respect to all three indicators, we found that more educated and more motivated respondents provided answers of significantly better quality compared to other respondents. However, the present study provides evidence that analyzing response quality exclusively with process-generated measures of quality may produce a misleading picture. Therefore, the addition of content-related indicators, such as the interpretability of responses, provides a more informative result. We found that the further the open-ended question was located towards the end of the questionnaire, the fewer interpretable answers were given. Our results also indicated that if the survey was carried out in close proximity to a federal election, responses were more likely to be interpretable. Overall, our study suggests that the characteristics at the respondent and survey levels influence the response quality of open-ended attitude questions and that these characteristics interact to a small degree.

Keywords: Open-ended questions, response quality, web surveys, multilevel modeling, satisficing

Researchers make use of open-ended questions in surveys because they allow respondents to report facts, behaviors, or attitudes without being restricted to a fixed set of answer choices. Open-ended questions can produce a much more diverse set of answers compared to closed-ended questions, which influence respondents' answers by providing cues to what kind of information is being sought via their response format (Dillman, Smyth, & Christian, 2014; Fuchs, 2009; Reja, Manfreda, Hlebec, & Vehovar, 2003; Schuman & Presser, 1996; Tourangeau, Rips, & Rasinski, 2000). Although it is well established that open-ended questions are advantageous because researchers can collect rich and detailed information from respondents on a topic of interest, these questions often suffer from comparably lower response quality as, for instance, is indicated by higher levels of item nonresponse (Reja et al., 2003; Schuman & Presser, 1996).

However, the implications for response quality also depend on the type of open-ended questions. Factual or behavioral open-ended questions – for instance, questions on behavioral frequencies or personal characteristics (cf. Fuchs, 2009; Holbrook et al., 2014) – usually limit the universe of adequate responses because the requested form of answers is rather obvious. Yet, particularly open-ended questions that ask about frequencies often suffer from the problem that respondents provide rounded answers (cf. Holbrook et al., 2014; Tourangeau et al., 2000; Turner, Sturgis, & Martin, 2015). Answering open-ended attitude questions is usually more demanding for respondents because they ask for a detailed response that might include several themes and elaboration on these themes (Holland & Christian, 2009; Smyth, Dillman, Christian, & McBride, 2009). Thus, responding to open-ended attitude questions often requires substantial cognitive effort from respondents, which is more burdensome and can lead to respondent fatigue (Dillman et al., 2014; Gummer & Roßmann, 2015; Holland & Christian, 2009). Consequently, respondents may use *satisficing* response strategies to reduce the burden of answering cognitively demanding open-ended attitude questions, which results in answers of lower quality (Holland & Christian, 2009; Krosnick, 1991, 1999).

The susceptibility of open-ended attitude questions to satisficing response behavior is particularly relevant for self-administered surveys, which lack a human interviewer who can motivate respondents and guide them through the response process (Holland & Christian, 2009; Rada & Dominguez-Alvarez, 2013; Reja et al., 2003). A substantial body of methodological research has examined the effects of questionnaire and question design on response quality for web surveys (e.g., Couper, Tourangeau, Conrad, & Zhang, 2013; Smyth et al., 2009; Tourangeau, Couper, & Conrad, 2004). With regard to these considerations, the present study

Direct correspondence to
Katharina Schmidt, GESIS – Leibniz Institute for the Social Sciences
katharina.schmidt@gesis.org

aims to answer the following research question: What characteristics affect the response quality of open-ended attitude questions in web surveys? Identifying relevant characteristics at the respondent and survey levels should effectively support researchers in designing web surveys that generate high-quality responses to open-ended questions.

Previous studies have compared the response quality of open-ended and closed questions (e.g., Reja et al., 2003), or have examined the mode differences in the response quality of web and paper questionnaires with respect to open-ended questions (e.g., Kwak & Radler, 2002; Rada & Dominguez-Alvarez, 2013). In addition, existing research has mostly examined the effects of a limited number of characteristics on response quality, such as the interest of the respondent in the topic (e.g., Galesic & Bosnjak, 2009; Holland & Christian, 2009; Olson & Peytchev, 2007), mobile device usage (Revilla & Ochoa, 2015a; Toepoel & Lugtig, 2014), and gender, age, or education (Couper & Kreuter, 2013; Denscombe, 2007; Yan & Tourangeau, 2008). The present study complements these studies in at least two ways. First, by applying multilevel modeling to data from 29 web surveys, we examined the characteristics of response quality at the respondent and survey levels, and the interaction of the variables at both levels. Second, with the notable exception of Holland and Christian (2009) and Smyth et al. (2009), prior studies have mostly used response length (e.g., Galesic, 2006; Galesic & Bosnjak, 2009; Grauenhorst, Blohm, & Koch, 2016; Kwak & Radler, 2002; Mavletova, 2013; Rada & Dominguez-Alvarez, 2013) or response time (e.g., Callegaro, Yang, Bhola, & Dillman, 2004; Galesic & Bosnjak, 2009) as indicators of response quality. The present study extends this research by using the interpretability of the responses to open-ended questions as an additional indicator of quality. As we argue later in the study, the interpretability of responses is potentially an even more appropriate and informative indicator of response quality than response length or latency.

The remainder of this study is organized as follows. The next section introduces *satisficing* as the theoretical framework for our study. Therefore, we present our expectations on how respondents cope with the cognitive demands of open-ended attitude questions, and review the indicators of response quality that previous research has used. Then, by using satisficing theory, we derive a set of hypotheses on the effects of several survey and respondent characteristics on the response quality of open-ended attitude questions. The following sections describe the data, the operationalization of the independent and dependent variables, and the methods used in the empirical analysis. The last sections present and discuss the results and close with recommendations for further research.

Theoretical Background

In the present study, we use satisficing theory (Krosnick, 1991, 1999) to measure and explain the response quality of open-ended attitude questions. Satisficing theory provides theoretical mechanisms that link the characteristics of questions and respondents with the use of response strategies that negatively affect response quality.

Satisficing theory assumes that answering survey questions usually requires respondents to pass through four stages of cognitive processing (Tourangeau & Rasinski, 1988; Tourangeau et al., 2000) – comprehension, information retrieval, judgment, and response selection. The response strategy that involves the complete and effortful execution of these cognitive processes is termed *optimizing*. However, if the difficulty of a question is high and a respondent is low in ability and/or motivation, the respondent might decide to use a *satisficing* response strategy (Krosnick, 1991, 1999). While weak forms of satisficing imply less effortful cognitive processing, strong satisficing involves skipping altogether the cognitive processes of question comprehension, information retrieval, and judgment. Hence, satisficing enables respondents to reduce the burden of responding to cognitively demanding survey questions (Krosnick, 1991, 1999). Consequentially, it follows from the propositions of the satisficing framework that the quality of responses should be poorer when respondents adopt weak or strong satisficing than when they optimize.

Coping with the Cognitive Demands of Open-Ended Attitude Questions

Satisficing theory states that under the condition of weak satisficing, respondents superficially or incompletely execute the processes of question comprehension, information retrieval, integration of the information into a summarizing judgment, and response reporting (Krosnick, 1991, 1999). Consequently, we assumed that respondents provide shorter and less detailed answers to an open-ended attitude question if they retrieve incomplete information or if they do not put sufficient effort into generating a well-formulated response. In line with previous research, we also suggest that short response latencies might indicate shortcuts and simplifications in the response process (see e.g., Greszki, Meyer, & Schoen, 2015; Roßmann, 2017; Roßmann, Gummer, & Silber, 2018; Smyth et al., 2009).

If respondents pursue strong satisficing as a response strategy, they completely skip the cognitive processes of question comprehension, information retrieval, and judgment (Krosnick, 1991, 1999). Therefore, shortcutting the processes of comprehension can result in answers that do not correspond to the question. Furthermore, a failure to retrieve information and integrate it into a judgment may tempt respondents to provide a response that lacks interpretability because it contains non-substantive information, such as “don’t know,” or nonsense entries, such as, for example, “:-)” (cf. Baker et al., 2010; Revilla & Ochoa, 2015b).

Measuring the Response Quality of Open-Ended Attitude Questions

Prior research on open-ended questions in self-administered surveys has used several indicators to study response quality. However, in most instances, these indicators were not derived from a unifying theoretical framework that links respondents’ response strategies with the quality of their responses.

First, previous studies often have related the accuracy of a response to its extensiveness, reasoning that the longer an open-ended response is, the more detailed and informative (e.g., Galesic, 2006; Galesic & Bosnjak, 2009; Grauenhorst, Blohm & Koch, 2016; Kwak & Radler, 2002; Mavletova, 2013; Rada & Dominguez-Alvarez, 2013). However, we need to be aware that longer responses are not necessarily more accurate than shorter ones. What we consider to be a high quality response also depends on the type of open-ended question. This study investigated open-ended attitude questions, specifically those that asked about the most important problem facing a country. For this particular type of open-ended question, shorter answers may be sufficient for accurately expressing an attitude, compared to open-ended questions that ask for more narrative responses. Depending on the content of a question, an inherent trade-off may exist between the extensiveness and accuracy of a response: Up to a certain point, the accuracy of a response increases with its length. However, at some point, a further increase in length may indicate that respondents put insufficient effort into integrating their retrieved information into a summarizing judgment. In these cases, it is often difficult to identify the information that the question asked for. Thus, response length alone may not be an ideal indicator of response quality.

Second, a growing body of research has suggested that longer response latencies indicate more effortful and thorough cognitive processing and, as a consequence, a higher quality response, compared to shorter response times (see, e.g., Greszki et al., 2015; Roßmann, 2017; Roßmann et al., 2018; Smyth et al., 2009). Accordingly, existing research has contended that respondents who do not put much effort into answering an open-ended question will tend to write less and have shorter response times (e.g., Revilla & Ochoa, 2015b). However, we have to acknowledge that longer latencies also may signal response problems or flawed questions (Bassili & Scott, 1996). For instance, some respondents might have difficulties understanding and answering a question because its wording is not concise or because it addresses several different topics at once. In this case, longer response latencies do not necessarily indicate higher quality. Short response latencies may also be the result of highly accessible attitudes and, thus, indicate responses of high quality (cf. Fazio, 1990; Mayerl, Sellke, & Urban, 2005). Thus, we suggest that response latencies should not be used as the sole indicator of response quality.

Third, only a few studies have examined the richness of detail and interpretability of responses to open-ended questions. Two of these studies looked at nonsubstantive and nonsense answers (Mavletova, 2013; Revilla & Ochoa, 2015a), and Smyth et al. (2009) coded the content of open-ended answers with regard to the number of themes that respondents addressed and the additional elaboration they provided. For this purpose, Smyth et al. (2009) defined a *theme* as "a concept or subject that answered the question and was independent of all other concepts within the response" (p. 327). In line with their reasoning, we suggest that answers that cannot be interpreted (i.e., answers that do not constitute a theme) indicate low quality, since they lack informative content or do not correspond to the question at all.

Although response length and, particularly, response latency are essentially process-generated measures, the interpretability of answers is a content-related indicator of response quality. In our view, this distinction is important because the different indicators of response quality may convey different information, and thus, their use in analyses might lead to different or even contradictory conclusions. As we assume throughout the present study, the interpretability of answers is likely a more appropriate indicator of response quality, compared to response length or latency, because it is less sensitive to conflicting assumptions about its association with quality. However, with regard to the majority of previous research and the naive expectations derived from satisficing theory, we base our analyses on the assumption that longer answers and longer response times reflect higher response quality.

Effects of Survey and Respondent Level Characteristics on the Quality of the Answers to an Open-Ended Attitude Question

In this section, we draw on *satisficing* as a theoretical framework to derive a comprehensive set of hypotheses to address the effects of explanatory factors on both survey and respondent levels (for an overview of our hypotheses, see Table 1).

Survey Level Characteristics

With respect to the survey level, an important factor is the location of the open-ended question in the questionnaire. According to satisficing theory, the response burden accumulates over the course of a questionnaire, which in turn may lead to respondent fatigue (Krosnick, 1991, 1999). Consequentially, the later an open-ended question is placed in a questionnaire, the higher are the chances that respon-

Table 1 Overview of the hypotheses

	Hypothesis
Survey level	
Hypothesis 1	The later open-ended questions are asked in a survey, the lower will be the response quality.
Hypothesis 2	The closer a survey is conducted to an event that is related to a question topic, the higher will be the response quality.
Hypothesis 3	The respondents of a probability-based online panel provide better quality answers, compared to the respondents of an opt-in online panel.
Respondent level	
Hypothesis 4	Higher educated respondents give better quality answers to open-ended questions, compared to less educated respondents.
Hypothesis 5	Older respondents give lower quality answers to open-ended questions, compared to younger respondents.
Hypothesis 6	Highly motivated respondents give better quality answers, compared to less motivated respondents.
Hypothesis 7	Respondents using a mobile device to answer open-ended survey questions give lower quality answers, compared to respondents using a PC.
Cross-level interactions	
Hypothesis 8	The later open-ended questions are asked in a survey, the larger is the effect of the respondents' motivation on response quality.
Hypothesis 9	The closer a survey is conducted to an event that is related to a question topic, the smaller is the effect of the respondents' abilities on response quality.

dents already will be fatigued and that they will perceive answering the question as taxing. Thus, we expect that the later open-ended questions are asked in a survey, the lower will be the response quality (Hypothesis 1).

Another factor at the survey level is the context of an interview. Surveys are conducted within broader societal environments that are characterized by events of which at least some will receive significant attention by the population under study. According to satisficing theory, it can be expected that if relevant information or pre-formulated attitudes are easily accessible, respondents should be motivated to optimize their responses – specifically, their cognitive processes of information retrieval should require much less effort (Krosnick, 1991, 1999). If a topic-related event occurs in close proximity to a survey, respondents should have more easily accessible information. Thus, we hypothesize that the closer a survey is conducted to an event that is related to a question topic, the higher will be the response quality (Hypothesis 2).

Further, we assume that opt-in online panelists and probability-based panelists differ in the quality of their responses. For example, a study by Silber, Lischewski, and Leibold (2013) compared the response behavior of the professional respondents of two online access panels with the less professional respondents of two web surveys. Their results showed that the respondents of the online access panels had lower break-off rates and were more likely to answer an open-ended attitude question. However, their answers were shorter and less often meaningful compared to the responses of the less professional respondents (Silber et al., 2013). In addition, due to the self-selection in the recruitment process, members of the opt-in online panels were more likely to hold multiple memberships in different online panels (Hillygus, Jackson, & Young, 2014). Therefore, we assume that the respondents from the opt-in online panels are used to answering large quantities of surveys. Moreover, since opt-in panelists presumably do more web surveys, compared to probability-based panelists, they may be less motivated to work through all four steps of cognitive processing, and satisfice more often (Baker et al., 2010). Thus, we expect that the respondents of a probability-based online panel to provide answers of better quality to open-ended questions, compared to the respondents of an opt-in online panel (Hypothesis 3).

Respondent Level Characteristics

With respect to the respondent level, we expect a set of individual characteristics to affect the efforts of respondents to form and report an interpretable response. According to our theoretical framework, respondents with greater ability are used to performing complex mental processes; they are practiced at thinking about the topic of a question and in formulating judgments (Krosnick, 1991, 1999). Previous research has shown that older respondents and those with lower levels of education often provide answers of worse quality (Couper & Kreuter, 2013; Denscombe, 2007; Knäuper, 1999; Loosveldt & Beullens, 2013; Olson & Peytchev, 2007; Roßmann et al., 2018; Yan & Tourangeau, 2008). Thus, we assume that higher educated and younger respondents give answers of better quality to open-ended questions, compared to less educated and older respondents, respectively (Hypothesis 4 and Hypothesis 5).

At the respondent level, another important factor is a respondent's motivation to answer questions accurately. Motivated respondents are more likely to perform all steps of the response process thoroughly, and thus, take their time to read and answer open-ended questions. In line with this assumption, previous studies have suggested that less motivated respondents give faster and shorter responses (e.g., Galesic & Bosnjak, 2009; Holland & Christian, 2009; Olson & Peytchev, 2007).

Thus, we hypothesize that highly motivated respondents give answers of higher quality, compared to less motivated respondents (Hypothesis 6).

In the past decade, the usage of Internet-capable mobile devices, like smartphones and tablets, has increased substantially (Gummer, Quoß, & Roßmann, 2019). Previous research has demonstrated that the use of these mobile devices affects response quality (De Bruijne & Wijnant, 2013; Mavletova, 2013; Peytchev & Hill, 2010; Stapleton, 2013). In particular, with regard to screen size, smartphones differ considerably from personal computers (PCs) and tablets. Since a smaller screen size limits the amount of visible information, respondents sometimes need to scroll or zoom to see the whole question. In addition, selecting a response on a touch screen may take longer due to the smaller screen size (Couper & Peterson, 2017). Thus, answering survey questions on a smartphone may require more effort from respondents (Couper & Peterson, 2017; De Bruijne & Wijnant, 2013; Mavletova, 2013; Peytchev & Hill, 2010; Stapleton, 2013) and therefore increase the burden of providing open-ended responses. Apart from that, respondents may use their smartphones and tablets more often to respond to surveys when they are outside of their home (Mavletova, 2013), and they may be more likely to multitask while completing web surveys (Couper & Peterson, 2017). Therefore, distractions or interruptions may be more common among users of mobile devices, which in turn can negatively affect response quality. This scenario is particularly important with respect to open-ended questions because users of smartphones or tablets usually need to enter their answer on a virtual keyboard, which often is more difficult, and thus, slower than using a regular keyboard with a desktop or notebook computer. In line with these assumptions, studies by Mavletova (2013) and Lugtig and Toepoel (2016) found that the use of smartphones to answer web surveys was associated with shorter responses to open-ended questions. Thus, we expect respondents using a mobile device to give answers of lower quality to open-ended attitude questions, compared to respondents using a PC (Hypothesis 7).

Cross-Level Interactions

Although the factors discussed above are conceptually located at different levels, we assume that they interact. According to satisficing theory, respondents differ in their response strategy depending on the position of the open-ended questions in the survey, and their motivation (Krosnick, 1991, 1999). Whereas higher motivated respondents probably invest more effort in answering open-ended questions, regardless of their position in the survey, less motivated respondents are likely to experience respondent fatigue earlier and switch their response strategy to satisficing (cf. Hypothesis 6). The closer an open-ended question is located near the end of the questionnaire, the larger are the differences between the respondents who are low in motivation and those who are highly motivated. Thus, we assume that the later the open-ended questions are asked in a survey, the larger is the effect of the respondents' motivation on response quality (Hypothesis 8).

Similarly, we expect an interaction between the proximity of a survey to a topically relevant event and respondents' ability to answer thoroughly an open-ended question on that topic. We also assume that the increased availability of topic-related attitudes and information diminishes the differences between highly able and less able respondents. In this regard, we hypothesize that the closer a survey is conducted to an event that is related to a question topic, the smaller is the effect of the respondents' abilities on response quality (Hypothesis 9).

Data

The present study draws on pooled data from 29 cross-sectional web surveys that were conducted between 2009 and 2015 as part of the German Longitudinal Election Study (Rattinger et al., 2009-2015). Building on the foundations of a repeated cross-section design, key questions were asked repeatedly in each survey, which covered topics such as political attitudes and behaviors, and socio-demographics. Surveys 1-16 used samples from a large German opt-in online panel with about 65,000 to 100,000 active panelists who were recruited to answer surveys on specific issues via online advertisements or via blogs and social media channels. In contrast, surveys 17-29 were sampled from a German probability-based online panel that was comprised of about 40,000 active panelists who were recruited at the end of regular computer-assisted telephone surveys (CATI) that drew on random digit dialing sampling. Comparable quotas on age, sex, and education were used to select each of the 29 samples for the web-based cross-sectional surveys. Accordingly, we calculated each survey's completion rate (AAPOR, 2016) following the recommendations of Callegaro and DiSogra (2008). On average, the completion rate was 82% (for details, see Appendix Table A.1). The pooled data set had 32,494 respondents (~1,120 per survey).

For our analyses, we selected a question measuring public opinion that is regularly asked in open-ended form in surveys (cf. Schuman & Presser, 1996): a question about the most important problem facing the country. The wording of the question was the same for all 29 surveys: "In your opinion, what is the most important political problem facing Germany at the moment?" The original German wording was: "Was ist ihrer Meinung nach gegenwärtig das wichtigste politische Problem in Deutschland?" While the wording of the question was constant across the surveys, the design of the question was slightly changed in some surveys. From survey 18 onwards, the maximum length of respondents' answers was technically limited to 100 characters, which forced respondents to shorten their response. Also, in surveys 21-24, the question was supplemented with additional features that made respondents aware of the response length limit. In our analyses, we included these changes in design as controls (see below). Answering the open-ended question was voluntary in each of the 29 surveys, so respondents could decide whether they would give an answer or leave the text box empty.

We created three indicators of the quality of the responses to the open-ended attitude question that served as dependent variables in our analyses: response length, response latency, and the interpretability of the answers. We operationalized response length by counting the number of characters. Since the character-based measure of length was skewed to the right (*Skewness*=2.64), we used the natural logarithm of the length for our further analyses. This transformation reduced the skewness to 0.44.

We measured response latency to the open-ended question in seconds.[1] As before, we used the natural logarithm to account for the skewness of the response latency measure. This reduced the skewness from 1.40 to -0.65.

Furthermore, we used the interpretability of the responses as a content-related indicator of response quality. During data processing, we coded respondents' answers to the open-ended questions into categories using a predefined coding scheme developed and extensively tested by the project team of the German Longitudinal Election Study.[2] We used the categories of this coding scheme to create a dummy variable that indicated whether the answers were interpretable or not (0 = not interpretable / 1 = interpretable). Answers that could not be interpreted (e.g., "asdf", "---"), did not mention a problem ("don't know"), or represented a refusal were coded as not interpretable. Answers that corresponded to the question and mentioned specific themes (e.g., "unemployment") were coded as substantive responses.

To explain response quality, we drew on a set of independent variables at the survey and respondent level. In addition, we included two cross-level interactions. Table 2 presents the descriptive statistics for all the variables we used in our analy-

1 An issue with response latencies is that their distributions are almost inevitably skewed (Fazio, 1990). Particularly in the absence of an interviewer in web-based surveys, we observed a characteristic long tail of slow latencies in the distribution. Since we do not know whether extremely slow latencies are caused by situational factors (e.g., distractions) or by lower abilities of respondents, we used a common outlier detection method and, in each survey, set response latencies that were longer than the mean plus two times the standard deviation of the distribution to missing (see e.g., Bassili & Fletcher, 1991). Therefore, we first omitted extreme outliers (5 minutes or more to answer the question) that would have skewed the distribution and affected the mean-based outlier criterion. Applying this approach, we classified 7.9% of the data points as response time outliers.

2 The development of the coding scheme for the open-ended question on the most important problem facing Germany was complemented by extensive tests of inter-coder reliability. Then, the coders received a comprehensive coding scheme, which included further information and detailed coding instructions to ensure high coding quality.

ses. A more detailed discussion on the operationalization of respondent level variables is provided in Appendix B.

At the survey level, we included three variables to test Hypotheses 1, 2, and 3. First, the web surveys used in this study applied a paging design (Couper, 2008). Thus, the screen on which the open-ended question appeared was a very good estimate of its position in the questionnaire (Hypothesis 1). For instance, in survey 29, the open-ended question appeared on the 10th screen. For an easier interpretation of the effects in our models (see Section 5), we rescaled the variable to a range of 0 to 1. Second, we included a dummy variable that indicated whether a survey was conducted within 6 months before or after the German federal elections in the years 2009 and 2013 (Hypothesis 2). Elections are among the most important political events in democratic societies. Since both the open-ended attitude question and web surveys were strongly related to political issues and elections in particular, it is likely that respondents have more readily available attitudes in times when a multitude of these issues are central to the public debate. Political information should be highly available for respondents due to election campaigns, which they can follow on advertising posters, television, or the Internet. In addition, during election campaigns, the appearance of specific political issues in the media and their handling by the candidates is higher (Huber, Rattinger, & Wagner, 2009; Schumann & Schoen, 2009). Third, we created a dummy variable that indicated whether the survey used respondents from an opt-in (surveys 1–16) or a probability-based (surveys 17–29) online panel (Hypothesis 3). For controls, we included two variables that indicated whether the response length was technically limited (0 = no / 1 = yes) and whether the question was supplemented with additional features to make respondents aware of the 100 characters response length limit (0 = no / 1 = yes).

With respect to the respondent level, we used education (0 = low / 1 = intermediate / 2 = high) as an indicator of the respondent's ability (Hypotheses 4). Since we also assumed that ability is associated with age, we included it (0 = 18–29 / 1 = 30–39 / 2 = 40–49 / 3 = 50–59 / 4 = 60+) as a second indicator (Hypothesis 5). Hypothesis 6 suggests that a respondent's motivation may influence their response behavior. Accordingly, we included three related variables: interest in the survey topic (0 = low interest in politics / 1 = intermediate interest in politics / 2 = high interest in politics), strength of the respondent's identification with a political party (0 = none / 1 = moderate / 2 = strong), and (intended) turnout to vote in a federal election (0 = no / 1 = yes). To examine the effects of different devices on response quality (Hypothesis 7), we identified whether respondents used a PC (desktop or notebook), tablet, or smartphone to complete the survey. The information on the device was extracted from the user agent string using the Stata command parseuas (Roßmann & Gummer, 2016). For control variables, we included the respondent's sex (0 = male / 1 = female) and region of residence (0 = East Germany / 1 = West Germany).

Table 2 Variables used to explain the response quality of open-ended attitude questions

Variable	M	Min	Max	N
Survey level				
Position of open-ended question	0.15	0	1	29
Proximity to election	0.41	0	1	29
Probability-based online panel	0.45	0	1	29
Respondent level				
Age				
18–29	0.23	0	1	32,494
30–39	0.20	0	1	32,494
40–49	0.24	0	1	32,494
50–59	0.16	0	1	32,494
60+	0.16	0	1	32,494
Education				
low	0.31	0	1	32,209
intermediate	0.39	0	1	32,209
high	0.30	0	1	32,209
Interest in politics				
low	0.21	0	1	32,458
intermediate	0.40	0	1	32,458
high	0.39	0	1	32,458
Intention to vote	0.85	0	1	32,449
Strength of party identification				
none	0.28	0	1	32,426
moderate	0.28	0	1	32,426
strong	0.44	0	1	32,426
Device				
personal computer	0.93	0	1	32,491
smartphone	0.04	0	1	32,491
tablet	0.03	0	1	32,491
Control Variables				
Technical limit of answer to 100 characters	0.41	0	1	29
Information on 100 characters limit	0.14	0	1	29
Sex: Female	0.50	0	1	32,494
Region: West Germany	0.80	0	1	32,487

Note. M = mean. Statistics at the respondent level variables are calculated with N = number of respondents. Statistics at the survey level are calculated with N = number of surveys.

As argued previously in the present study, interactions between respondent and survey characteristics can be assumed to partially explain response behavior. Thus, we created a cross-level interaction between the location of the open-ended attitude question in the questionnaire and the respondents' interest in the survey topic (Hypothesis 8). Further, to test whether topic-related events enhance the availability and accessibility of relevant attitudes and information, we created a second cross-level interaction between the survey's proximity to a federal election and the respondents' ability as indicated by their level of education (Hypothesis 9).

Methods

To statistically account for the multilevel structure of our data – individuals clustered in surveys – and to test the hypotheses and interactions of two conceptual levels (respondent and survey level), we applied multilevel modeling (Hox, 2010; Luke, 2004; Rabe-Hesketh & Skrondal, 2008; Snijders & Bosker, 1999) using Stata 14.1. This approach explicitly modeled that the characteristics of the lower level (i.e., respondents) depend on the higher level (i.e., surveys). Our mathematical expressions mainly refer to the work of Snijders and Bosker (1999) and Luke (2004).

We fitted a random intercept model with fixed slopes and cross-level interactions. Since we assumed that the location of a question in a survey and the proximity of a survey to a topic-related event explain the variation in the coefficients of respondents' ability and motivation (Hypotheses 8 & 9), the slopes were fixed. In the following, Y_{ij} denotes an individual i's response behavior in survey j. X_{pij} is a vector of p characteristics at the respondent level, whereas Z_{qj} is a vector of q characteristics at the survey level. $X_{pij}Z_{qj}$ is a vector of cross-level interactions. Thus, γ_{0p}, γ_{q0}, and γ_{qp} are the respective regression coefficients. γ_{00} is the grand mean, u_{0j} is the survey level residuals, and r_{ij} is the respondent level residuals. Consequently, our final (linear) model used to explain response length and latency is denoted in single-equation form as follows:

$$Y_{ij} = \gamma_{00} + \sum \gamma_{0p} X_{pij} + \sum \gamma_{q0} Z_{qj} + \sum \gamma_{qp} X_{pij} Z_{qj} + u_{0j} + r_{ij}$$

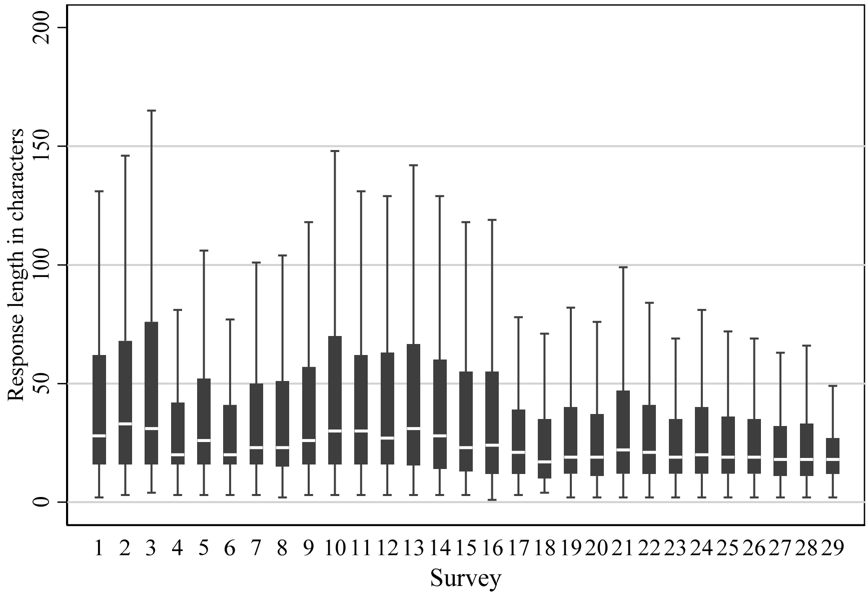
in which $Y_{ij} = \ln\left(Y_{ij}\right)$ is the transformed response length and latency that we used as dependent variables. Due to their operationalization, we assumed that both response length and latency indicators are approximately normally distributed.

Since our theoretical reasoning remained the same for our binary dependent variable, the respective logistic multilevel model also was specified as a random intercept fixed slope model with cross-level interactions. Accordingly, we modelled the probability of respondent i giving an interpretable answer P_{ij} in survey j as follows:

$$logit(P_{ij}) = \gamma_{00} + \sum \gamma_{0p} X_{pij} + \sum \gamma_{q0} Z_{qj} + \sum \gamma_{qp} X_{pij} Z_{qj} + u_{0j}$$

Results

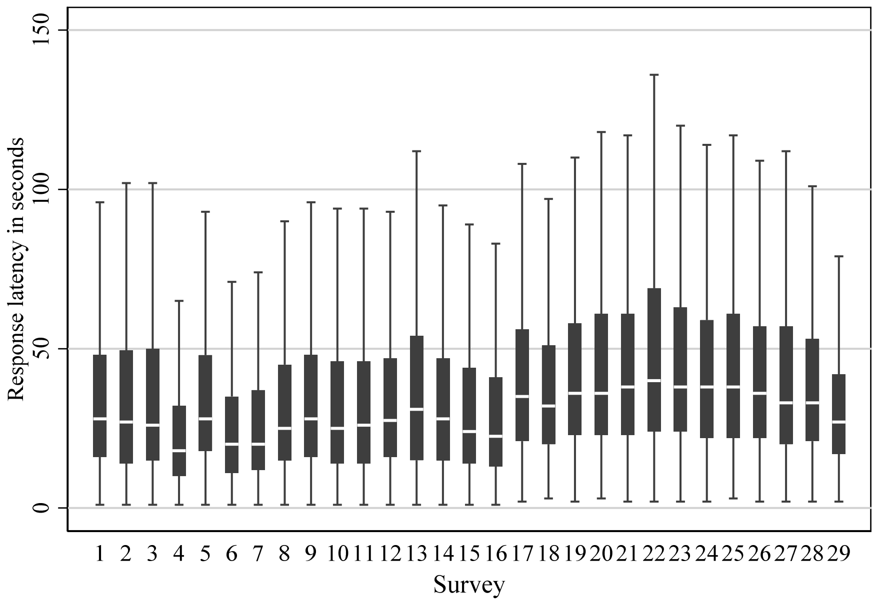
Before reporting the results of our multilevel models, we will briefly discuss the variation in the length of responses, response latencies, and interpretable answers across surveys. Figure 1 shows the length of the responses and visualizes the variations between surveys, which is particularly evident for the surveys 18-29 in which the response length was limited to 100 characters.



Note. Outliers were excluded from the analysis.

Figure 1 Boxplots of the response length to the question on the most important political problem facing Germany for 29 web surveys

The boxplots of Figure 2 illustrate the variations in response latencies across surveys. These plots suggest that a strong variation exists in latencies within each survey, and between surveys. Apparently, the average response latencies increased after the sampling switched from an opt-in panel to a probability-based panel in survey 17.



Note. Outliers were excluded from the analysis.

Figure 2 Boxplots of response latency to the question on the most important political problem facing Germany for 29 web surveys

Figure 3 depicts the variation in the share of interpretable answers to the open-ended question. On average, 88.05% of the answers were interpretable with a strong variation across surveys. Notably, more between-survey variation of interpretable answers seems to occur when an opt-in panel was used; and a more homogeneous (but larger) share of interpretable answers seems to occur between surveys when a probability-based panel was used. Table 3 shows the results of the multilevel models for the three indicators of response quality: response length (Model 1), response latency (Model 2), and interpretability of the answers (Model 3).

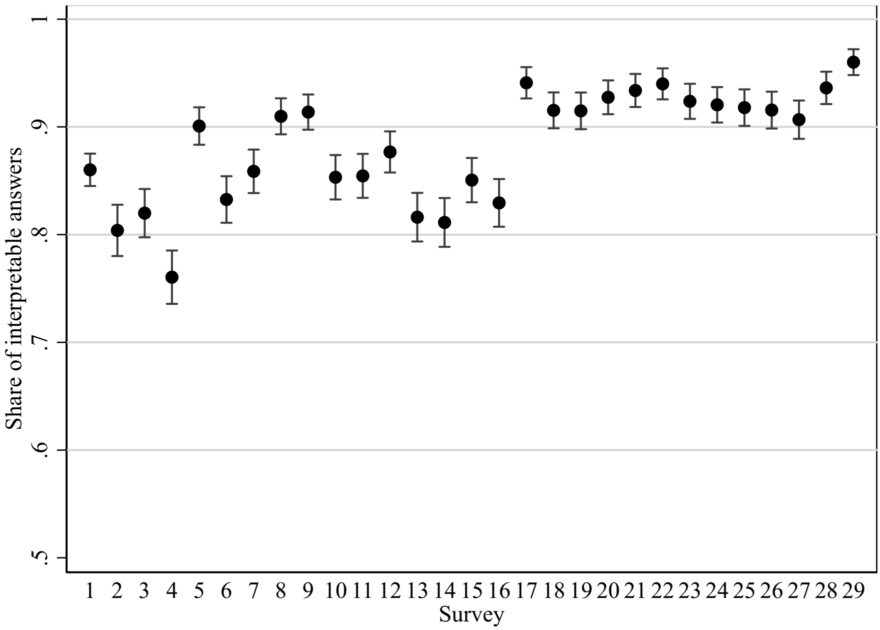


Figure 3 Share of interpretable answers to the question on the most important political problem facing Germany for 29 web surveys

Note that all variables reported in Table 3 range from 0 to 1. Thus, the coefficients of Model 1 and 2 provide the marginal effects of each of the dependent variables. This enables an easy interpretation of the coefficients, since a 1-unit change in any of the independent variables is equivalent to comparing a respondent at the minimum value of the respective variable to a respondent at the maximum value. Similarly, we report the average marginal effects (AMEs) for logistic Model 3. AMEs enable an intuitive interpretation as the average effect on the probability over all cases in the sample (Best & Wolf, 2015).[3]

3 We tested multiple hypotheses in our study what could possibly result in capitalizing on chance (Type I error, i.e., rejecting too many null hypotheses). However, popular adjustment methods such as the Bonferroni correction come at the price of lowering statistical power and, in turn, increase the chance of Type II errors (Gelman, Hill, & Yajima, 2012; Rothman, 1990). Against this caveat, we remain skeptical whether to correct for this potential issue. In addition, we derived all hypotheses from satisficing theory and prior research, what lays ground for a careful assessment of the plausibility of effects that we found to be statistically significant. In our view, this approach should limit the negative consequences that potential Type I errors might have. Accordingly, we argue that the problem of multiple comparisons is not likely to be a major issue in the present study.

To assess the explanatory power of our models, we calculated the intraclass correlation coefficient (ICC) for all three dependent variables based on empty models. For the response length (ICC = .05), the response latency (ICC = .07), and the interpretability of the answers (ICC = .07), part of the variance can be attributed to the survey level. The residual variances of our three final models (Table 3) further indicate that including our covariates reduced the proportion of unexplained variation between surveys.

Table 3 Multilevel models of the response quality to open-ended attitude questions

	Model 1 Response Length	Model 2 Response Latency	Model 3 Response Interpretability	
	b (SE)	b (SE)	b (SE)	AME (SE)
Survey level effects				
Position of open-ended question	-0.041	-0.452***	-0.886***	-0.083***
	(0.104)	(0.107)	(0.241)	(0.023)
Proximity to election	0.034	0.123*	0.466***	0.044***
	(0.048)	(0.053)	(0.125)	(0.012)
Probability-based online panel	-0.296***	0.343***	0.918***	0.086***
	(0.089)	(0.100)	(0.244)	(0.023)
Respondent level effects				
Education: low	Ref.	Ref.	Ref.	Ref.
intermediate	0.041*	-0.006	0.259***	0.024***
	(0.018)	(0.016)	(0.059)	(0.006)
high	0.111***	0.049**	0.463***	0.044***
	(0.019)	(0.017)	(0.069)	(0.007)
Age: 18–29	Ref.	Ref.	Ref.	Ref.
30–39	-0.026	0.079***	0.262***	0.025***
	(0.017)	(0.015)	(0.050)	(0.005)
40–49	-0.043**	0.187***	0.662***	0.062***
	(0.016)	(0.014)	(0.052)	(0.005)
50–59	-0.050**	0.247***	0.817***	0.077***
	(0.018)	(0.016)	(0.064)	(0.006)
60+	0.056**	0.405***	0.896***	0.084***
	(0.018)	(0.017)	(0.070)	(0.007)
Interest in politics: low	Ref.	Ref.	Ref.	Ref.
intermediate	0.102***	0.177***	0.681***	0.064***
	(0.018)	(0.016)	(0.052)	(0.005)
high	0.210***	0.228***	1.354***	0.127***
	(0.019)	(0.017)	(0.069)	(0.007)

	Model 1 Response Length	Model 2 Response Latency	Model 3 Response Interpretability	
	b (SE)	b (SE)	b (SE)	AME (SE)
Intention to vote	-0.065*** (0.017)	0.039** (0.015)	0.233*** (0.046)	0.022*** (0.004)
Strength of party identification: none	Ref.	Ref.	Ref.	Ref.
moderate	-0.048** (0.015)	0.083*** (0.013)	0.387*** (0.046)	0.036*** (0.004)
strong	-0.082*** (0.014)	0.082*** (0.013)	0.657*** (0.047)	0.062*** (0.005)
Device: personal computer	Ref.	Ref.	Ref.	Ref.
smartphone	-0.163*** (0.028)	0.129*** (0.025)	0.073 (0.102)	0.007 (0.010)
tablet	-0.152*** (0.031)	0.010 (0.029)	0.110 (0.135)	0.010 (0.013)
Cross-level interaction effects				
Interest in politics: intermediate × Position of open-ended question	0.022 (0.067)	0.056 (0.054)	-0.361* (0.150)	-0.034* (0.014)
Interest in politics: high × Position of open-ended question	-0.009 (0.066)	0.122* (0.054)	-0.510** (0.170)	-0.048** (0.016)
Education: intermediate × Proximity to election	-0.056* (0.026)	-0.023 (0.023)	-0.076 (0.084)	-0.007 (0.008)
Education: high × Proximity to election	-0.053 (0.028)	-0.076** (0.025)	-0.247* (0.098)	-0.023* (0.009)
Control variables				
Technical limit of answer to 100 characters	-0.073 (0.091)	-0.039 (0.102)	-0.206 (0.248)	-0.019 (0.023)
Information on 100 characters limit	0.091 (0.059)	0.059 (0.066)	-0.167 (0.155)	-0.016 (0.015)
Sex: female	0.026* (0.011)	-0.046*** (0.010)	-0.071 (0.038)	-0.007 (0.004)
Region: West Germany	-0.007 (0.013)	-0.046*** (0.012)	-0.157** (0.048)	-0.015** (0.005)
Intercept	3.346*** (0.039)	2.820*** (0.039)	0.132 (0.102)	
$\sigma^2_{u_0}$	0.006	0.008	0.036	
σ^2_r	0.781	0.683	3.290	
N	28,264	29,520	32,062	

Note. *p*-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Survey level effects

With respect to the location of the open-ended question in the questionnaire (Hypothesis 1), we did not find any evidence that later placement affected the length of answers. However, respondents took less time to answer the question the later it was placed in a questionnaire. Further, respondents gave interpretable answers to a significantly lesser extent the later the open-ended question was asked in the survey. The latter two findings are in line with the expectation that fatigue increases over the course of a questionnaire, which increases the likelihood of respondents adopting a satisficing response strategy.

Also, we hypothesized that if the survey was conducted in proximity to a topic-related event (i.e., the German federal elections 2009 and 2013), respondents would provide answers of higher quality due to the increased availability and accessibility of pre-formulated attitudes and relevant information (Hypothesis 2). The results of Models 2 and 3 showed that respondents took more time to answer and that responses were more likely to be interpretable if the survey was carried out 6 months before or after an election, which we hypothesized to be an effect of more accessible information. We did not observe significant effects on response length.

Next, we expected that respondents of a probability-based online panel would provide higher quality responses to an open-ended attitude question, compared to opt-in panelists (Hypothesis 3). Again, our findings are mixed. We found that membership in the probability-based online panel had a negative effect on response length. However, this negative effect was not particularly surprising, since the introduction of the 100 character limit in survey 18 almost perfectly coincided with the change of the panel provider in survey 17. Accordingly, we refrain from overinterpreting this finding. In contrast, we found a positive effect of membership in a probability-based online panel on response latency. Our results also revealed that respondents of the probability-based online panel gave interpretable answers at a significantly higher rate than the opt-in panelists. The latter findings supported our theoretical expectation that the sample of the probability-based online panel was composed of less over-surveyed, and thus, more motivated respondents who engaged in providing interpretable responses, compared to the sample of the opt-in online panel.

Respondent level effects

As we had expected in Hypothesis 4, our results confirmed that highly educated respondents gave answers of higher quality. On average, their answers were longer, and they took more time to respond to the open-ended question, compared to less educated respondents. Higher educated respondents also gave interpretable answers at a higher rate. These findings are in line with the assumption that respondents high in ability are more likely to carefully execute all steps of cognitive processing. In contradiction to Hypothesis 5, we found that the group of the oldest respondents needed more time to answer the open-ended question and gave interpretable responses to a greater extent, compared to younger age groups. On the basis of these findings, we rejected Hypothesis 5. On the one hand, the lower response quality of younger respondents was surprising, since previous studies (e.g., Knäuper, 1999) and satisficing theory (Krosnick, 1991) have suggested that younger respondents tend to provide better responses due to a higher working memory capacity (i.e., ability). On the other hand, we interpret our results as an indication that age might not be a well-suited measure for determining respondents' abilities to thoroughly answer open-ended questions (cf. Holbrook, Krosnick, Moore, & Tourangeau, 2007).

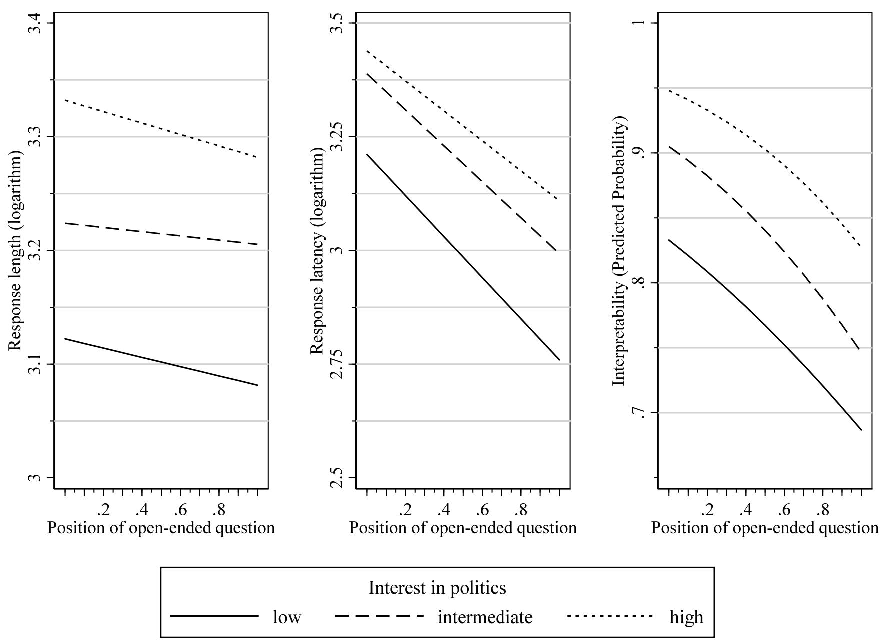
With regard to motivation (Hypothesis 6), we found that respondents with a higher interest in a survey topic provided a significantly better response quality than less interested respondents. They gave longer answers, took more time to respond, and provided interpretable answers at a higher rate. These results indicate that interest in a survey topic plays an important role in shaping the quality of the responses to open-ended questions (see, e.g., Holland & Christian, 2009). In contradiction to our expectations, respondents who turned out, or intended to vote gave significantly shorter responses, compared to those who did not intend to vote. A similar pattern emerged with respect to respondents' identification with a political party: respondents who reported a strong or at least a moderate psychological attachment gave significantly shorter answers than those who did not identify themselves with a party at all. However, in line with Hypothesis 6, a moderate or strong party identification had significant positive effects on the response latency and interpretability of the response. Thus, our findings regarding the effects of motivation on response quality largely confirm Hypothesis 6.

In line with previous studies (e.g., Lugtig & Toepoel, 2016; Mavletova, 2013), our results showed that tablet or smartphone usage negatively affected the number of characters entered, compared to the use of a PC (desktop or notebook). Smartphone users also took more time to answer the open-ended attitude question than respondents using a PC. As discussed previously in the present study, respondents may take longer answering survey questions with a smartphone due to the smaller screen size and the use of virtual keyboards (e.g., Couper & Peterson, 2017; Mavletova, 2013). However, we found no significant effects of mobile device usage on the rate of interpretable answers. Thus, Hypothesis 7 was only partly confirmed with respect to the length of answers and response latency for smartphone users. We suggest that these findings indicate that the use of mobile devices - particularly smartphones - has notable effects on the process of entering open-ended responses, but not necessarily on the quality of the content.

Cross-level interaction effects

In the last step of our analyses, we examined whether the respondent and survey level factors interacted across conceptual levels to affect the quality of the responses to open-ended questions.

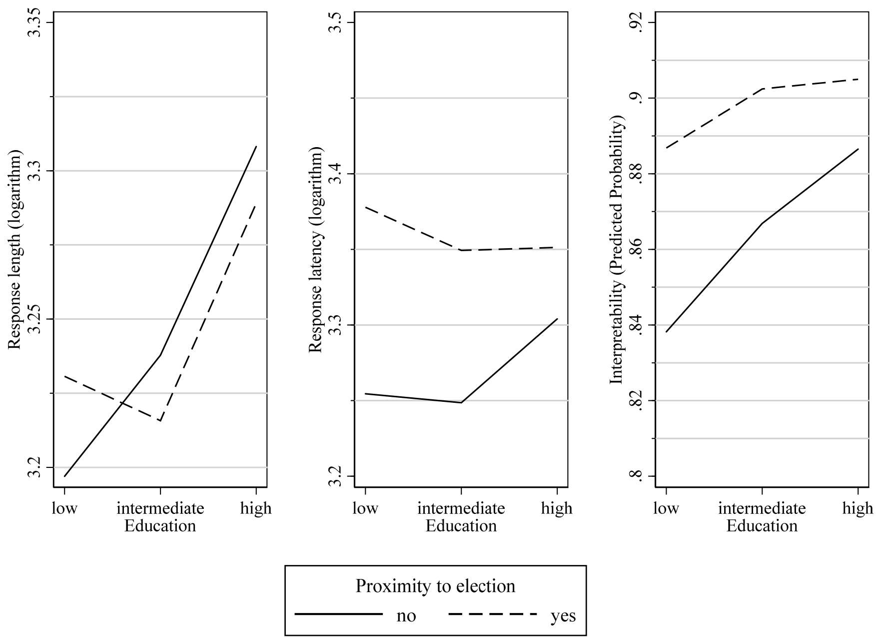
In particular, we assumed that the later the open-ended attitude questions are asked in a survey, the larger is the effect of the respondents' motivation on their response quality (Hypothesis 8). For the purpose of illustration, Figure 4 presents interaction plots for each of the three indicators.



Note. Predicted values based on models presented in Table 3: Model 1 = response length, Model 2 = response latency, and Model 3 = interpretability of response.

Figure 4 Cross-level interactions between the position of an open-ended question in a survey and the effects of respondents' motivation on response quality

In contrast to our expectation, we found no significant effects on response length due to the cross-level interaction of respondents' interest in a survey topic and the location of the open-ended attitude question in the questionnaire (Figure 4, first plot). Although Models 2 and 3 showed significant interaction effects on the response latency (for high political interest) and interpretability of responses, the visual presentation of these effects (Figure 4, second and third plot) suggests that the differences between respondents who are low in motivation and those who are highly motivated is only slightly different if the open-ended attitude question is located towards the end of the questionnaire. Consequently, even though we found that effects occurred, their impact seems to be limited. With regard to Hypothesis 9, we found a significant negative effect on the length of answers due to the interaction between the intermediate levels of education and the proximity of a survey to a federal election (Figure 5, first plot).



Note. Predicted values based on models presented in Table 3: Model 1 = response length, Model 2 = response latency, and Model 3 = interpretability of response.

Figure 5 Cross-level interactions between the proximity of a survey to a federal election and respondents' education level

In addition, significant impacts occurred on the interpretability of responses and response latencies due to the interaction of high education and the proximity of a survey to a federal election (Figure 5, second and third plot). With respect to the respondent level, the study found that highly educated respondents provided answers of higher quality, compared to lower educated respondents (cf. Hypothesis 3), and all respondents gave more interpretable answers during a time period of 6 months before or after a federal election (cf. Hypothesis 2). However, the cross-level interaction effects imply a more complex association. In line with our theoretical expectations, it seemed that in times where political issues were central in the public debate, related attitudes and information also were more readily accessible for less educated respondents. The results indicated that the differences in response quality between low and highly educated respondents were reduced by the occurrence of a topic-related event (i.e., a federal election) (Hypothesis 9), albeit only to a small extent as Figure 5 illustrates.

Conclusion

The present study investigated the effects of respondent and survey characteristics on the response quality of open-ended attitude questions in web surveys, which complements previous research in several ways. First, we analyzed a pooled data set of 29 web surveys on the political attitudes and behaviors of German Internet users. These data not only enabled us to study respondent level effects, but also to gain new insights into the effects of survey design, and the interaction of this design and respondent characteristics on the response quality of an open-ended attitude question.

Second, we used three different indicators of response quality. Nevertheless, the results of our analyses with these three indicators did not provide unambiguous evidence for every hypothesis. Thus, the question arises as to whether short responses or latencies consistently imply bad response quality or not. In other words, we need to ask whether the relationship between these indicators and response quality is more complex than the majority of previous studies have suggested (see Section 2). Our results indicate that analyzing the response quality to an open-ended attitude question exclusively with single indicators, for instance, with response length or latency, may create a misleading picture. Including content-related indicators such as the interpretability of responses provided us with more differentiated insights, compared to the exclusive use of process-generated measures of quality (i.e., response length or latency). Moreover, for the majority of the survey and respondent level variables, their effects on the content-related measure of response quality were in line with the theoretical expectations. We believe this result is an indication that the content-related measure captured what is most generally understood as the response quality of open-ended questions. Thus, in future studies on the quality of responses to open-ended questions in surveys, we recommend using content-related indicators, such as the number of themes that were addressed (Smyth et al., 2009) or the interpretability of answers. For future research, studying a variety of response quality indicators and exploring the empirical and theoretical relationships between them certainly seems worthwhile.

Third, we used satisficing theory to analyze the response quality of open-ended questions. The analyses we carried out lend support to several hypotheses on the effects of respondent and survey level characteristics on response quality, which we derived from the satisficing framework. In particular, our empirical results support the assumption that motivated respondents and those high in ability provided higher quality responses. These results are in line with previous studies that have found that respondents who are more interested in a survey topic or who are more highly educated are more likely to provide an open-ended response of good quality, compared to less motivated or less able respondents (Denscombe, 2007; Holland & Christian, 2009; Knäuper, 1999; Smyth et al., 2009).

Fourth, by including cross-level interactions in our models, we found that factors on different conceptual levels were not completely independent in affecting response quality. This finding emphasizes the need for further studies on the effects on answer quality caused by the cross-level interactions between respondent and survey level characteristics. Moreover, the finding that significant, albeit small, differences exist with respect to interpretable responses - due to the interaction of the location of the open-ended question in the survey and respondents' low and high in interest in a survey topic - highlights the importance of considering a respondent's motivation when designing web surveys. This finding supports the results from experimental studies that have demonstrated that altering the visual design of a survey can stimulate less motivated respondents to provide responses of better quality (cf. Holland & Christian, 2009; Smyth et al., 2009). For example, Smyth et al. (2009) found that using an introduction that emphasizes the importance of answers to the researchers increased the respondents' elaboration of themes. Also, the results of our study indicate that survey designers should take into account the societal context during the data collection period, since the response quality of an open-ended attitude question can be influenced by topic-related events that occur in proximity to the survey (e.g., a federal election). The present study has shown that the occurrence of such an event can diminish the differences in response quality that normally are caused by the differences in respondents' abilities. This finding is particularly important when analyzing (pooled) longitudinal data sets, which are comprised of interviews that were conducted in close proximity to topic-related events and others that were not. As our findings suggest, measurement errors are not homogenous across surveys; instead, they differ systematically. During analyses, these errors may be mistaken for a substantive change over time or surveys.

The following limitations of the present study should, however, be considered. First, with regard to the political topic of a survey and the particular type of open-ended attitude question (the most important problem), we suggest that follow-up studies should further examine how findings can vary across different survey topics or hold for other types of open-ended questions (e.g., open-ended questions that require more narrative responses). Second, in the present study, we limited the number of survey level characteristics because we decided to pool similar surveys. In light of this limitation, future studies could compile a more diversely designed set of surveys to test more interactions of more factors at the respondent and survey levels. Compiling a larger collection of surveys should help future studies to arrive at findings that are more robust. Finally, a further interesting opportunity for upcoming research would be to develop additional content-related indicators of data quality to measure how strongly responses correspond to the actual open-ended question, and whether these responses are interpretable.

References

AAPOR. (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*: The American Association for Public Opinion Research.

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., . . . Zahs, D. (2010). Research Synthesis: AAPOR Report on Online Panels. *Public Opinion Quarterly, 74*(4), 711-781. doi:10.1093/poq/nfq048

Bassili, J. N., & Fletcher, J. F. (1991). Response-Time Measurement in Survey Research. A Method for CATI and a New Look at Nonattitudes. *Public Opinion Quarterly, 55*(3), 331-346. doi:10.1086/269265

Bassili, J. N., & Scott, B. S. (1996). Response Latency as a Signal to Question Problems in Survey Research. *Public Opinion Quarterly, 60*(3), 390-399. doi:10.1086/297760

Best, H., & Wolf, C. (2015). Logistic Regression. In H. Best & C. Wolf (Eds.), *Regression Analysis and Causal Inference* (pp. 153-171). Los Angeles: Sage.

Callegaro, M., & DiSogra, C. (2008). Computing Response Metrics for Online Panels. *Public Opinion Quarterly, 72*(5), 1008-1032. doi:10.1093/poq/nfn065.

Callegaro, M., Yang, Y., Bhola, D., & Dillman, D. A. (2004). *Response Latency as an Indicator of Optimizing. A Study Comparing Job Applicants and Job Incumbents' Response Time on a Web Survey.* Paper presented at the Proceedings of the RC 33 Sixth International Conference on Social Science Methodology, Amsterdam.

Couper, M. P. (2008). *Designing Effective Web Surveys.* New York: Cambridge University Press.

Couper, M. P., & Kreuter, F. (2013). Using Paradata to Explore Item Level Response Times in Surveys. *Journal of the Royal Statistical Society, 176*(1), 271-286. doi:10.1111/j.1467-985X.2012.01041.x

Couper, M. P., & Peterson, G. J. (2017). Why Do Web Surveys Take Longer on Smartphones? *Social Science Computer Review*, 35(3), 357-377. doi:10.1177/0894439316629932

Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The Design of Grids in Web Surveys. *Social Science Computer Review, 31*(3), 322-345. doi:10.1177/0894439312469865

De Bruijne, M., & Wijnant, A. (2013). Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment with a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey. *Social Science Computer Review, 31*(4), 482-504. doi:10.1177/0894439313483976

Denscombe, M. (2007). The Length of Responses to Open-Ended Questions: A Comparison of Online and Paper Questionnaires in Terms of a Mode Effect. *Social Science Computer Review, 26*(3), 359-368. doi:10.1177/0894439307309671

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*: John Wiley & Sons.

Fazio, R. H. (1990). A Practical Guide to the Use of Response Latency in Social Psychological Research. In C. Hendrick & M. S. Clark (Eds.), *Research Methods in Personality and Social Psychology* (pp. 74-97). Thousand Oaks, CA: Sage.

Fuchs, M. (2009). Differences in the Visual Design Language of Paper-and-Pencil Surveys Versus Web Surveys A Field Experimental Study on the Length of Response Fields in Open-Ended Frequency Questions. *Social Science Computer Review, 27*(2), 213-227.

Galesic, M. (2006). Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey. *Journal of Official Statistics, 22*(2), 313-328.

Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly, 73*(2), 349-360. doi:10.1093/poq/nfp031

Gelman, A., Hill, J., & Yajima, M. (2012). Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness, 5*(2), 189-211.

Grauenhorst, T., Blohm, M., & Koch, A. (2016). Respondent Incentives in a National Face-to-Face Survey Do They Affect Response Quality? *Field Methods*, 28(3), 266-283. doi:10.1177/1525822X15612710

Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the Effects of Removing "Too Fast" Responses and Respondents from Web Surveys. *Public Opinion Quarterly, 79*(2), 471-503. doi:10.1093/poq/nfu058

Gummer, T., Quoß, F., & Roßmann, J. (2019). Does Increasing Mobile Device Coverage Reduce Heterogeneity in Completing Web Surveys on Smartphones? *Social Science Computer Review*, 37(3), 371-384. doi:10.1177/0894439318766836

Gummer, T., & Roßmann, J. (2015). Explaining Interview Duration in Web Surveys: A Multilevel Approach. *Social Science Computer Review, 33*(2), 217-234. doi:10.1177/0894439314533479

Hillygus, S. D., Jackson, N., & Young, M. (2014). Professional Respondents in Nonprobability Online Panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 219-237): Chichester: Wiley.

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. doi:10.1093/poq/nfu017

Holbrook, A. L., Krosnick, J. A., Moore, D., & Tourangeau, R. (2007). Response Order Effects in Dichotomous Categorical Questions Presented Orally. The Impact of Question and Respondent Attributes. *Public Opinion Quarterly, 71*(3), 325-348. doi:10.1093/poq/nfm024

Holland, J. L., & Christian, L. M. (2009). The Influence of Topic Interest and Interactive Probing on Responses to Open-Ended Questions in Web Surveys. *Social Science Computer Review, 27*(2), 196-212.

Hox, J. (2010). *Multilevel Analysis. Techniques and Applications*. New York: Routledge.

Huber, S., Rattinger, H., & Wagner, C. (2009). No Matter When? Eine Analyse von Feldzeiteffekten. In S. Huber, H. Rattinger, & C. Wagner (Eds.), *Vom Interview zur Analyse* (Vol. 1, pp. 231-258).

Knäuper, B. (1999). The Impact of Age and Education on Response Order Effects in Attitude Measurement. *Public Opinion Quarterly, 63*, 347-370.

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

Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology, 50*, 537-567.

Kwak, N., & Radler, B. (2002). A Comparison Between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. *Journal of Official Statistics, 18*(2), 257-273.

Loosveldt, G., & Beullens, K. (2013). 'How Long Will it Take?' An Analysis of Interview Length in the Fifth Round of the European Social Survey. *Survey Research Methods, 7*(2), 69-78.

Lugtig, P., & Toepoel, V. (2016). The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error. *Social Science Computer Review, 34*(1), 78-94. doi:10.1177/0894439315574248

Luke, D. A. (2004). *Multilevel Modeling*. Thousand Oaks: SAGE.

Mavletova, A. (2013). Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review, 31*(6), 725-743. doi:10.1177/0894439313485201

Mayerl, J., Sellke, P., & Urban, D. (2005). *Analyzing Cognitive Processes in CATI-Surveys with Response Latencies: An Empirical Evaluation of the Consequences of Using Different Baseline Speed Measures*. Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart. Universität Stuttgart. Stuttgart.

Olson, K., & Peytchev, A. (2007). Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes. *Public Opinion Quarterly, 71*(2), 273-286. doi:10.1093/poq/nfm007

Peytchev, A., & Hill, C. A. (2010). Experiments in Mobile Web Survey Design: Similarities to Other Modes and Unique Considerations. *Social Science Computer Review, 28*(3), 319-335. doi:10.1177/0894439309353037

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*. College Station: Stata Press.

Rada, V. D. d., & Dominguez-Alvarez, J. A. (2013). Response Quality of Self-Administered Questionnaires: A Comparison Between Paper and Web Questionnaires. *Social Science Computer Review, 32*(2), 256-269. doi:10.1177/0894439313508516

Rattinger, H., Roßteutscher, S., Schmitt-Beck, R., Schoen, H., Weßels, B., & Wolf, C. (2009-2015). *Long-term Online Tracking, T1-T29 (ZA5334-ZA5351 and ZA5719-ZA5729)*. Cologne: GESIS Data Archive.

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-Ended vs. Close-Ended Questions in Web Questionnaires. *Developments in Applied Statistics, 19*, 159-177.

Revilla, M., & Ochoa, C. (2015a). Open Narrative Questions in PC and Smartphones: Is the Device Playing a Role? *Quality & Quantity*, 1-19.

Revilla, M., & Ochoa, C. (2015b). What are the Links in a Web Survey Among Response Time, Quality, and Auto-Evaluation of the Efforts Done? *Social Science Computer Review, 33*(1), 97-114. doi:10.1177/0894439314531214

Roßmann, J. (2017). *Satisficing in Befragungen. Theorie, Messung und Erklärung [Satisficing in Surveys. Theory, Measurement, and Explanation]*. Wiesbaden: Springer VS.

Roßmann, J., & Gummer, T. (2016). *PARSEUAS: Stata Module to Extract Detailed Information From User Agent Strings*. Chestnut Hill, MA: Boston College.

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. doi:10.1093/jssam/smx020

Rothman, K. J. (1990). No Adjustments are Needed for Multiple Comparisons. *Epidemiology*, 43-46.

Schuman, H., & Presser, S. (1996). *Questions and Answers in Attitude Surveys*. New York: Academic Press.

Schumann, S., & Schoen, H. (2009). Muster an Beständigkeit? Zur Stabilität politischer und persönlicher Prädispositionen. In S. Huber, H. Rattinger, & C. Wagner (Eds.), *Vom Interview zur Analyse* (Vol. 1, pp. 13-34).

Silber, H., Lischewski, J., & Leibold, J. (2013). Comparing Different Types of Web Surveys: Examining Drop-Outs, Non-Response and Social Desirability. *Metodološki zvezki, 10*(2), 121-143.

Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-Ended Questions in Web Surveys: Can Increasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Response Quality? *Public Opinion Quarterly, 73*(2), 325-337. doi:10.1093/poq/nfp029

Snijders, T., & Bosker, R. (1999). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. London: SAGE.

Stapleton, C. E. (2013). The Smartphone Way to Collect Survey Data. *Survey Practice, 6*(2), 1-7.

Toepoel, V., & Lugtig, P. (2014). What Happens if You Offer a Mobile Option to Your Web Panel? Evidence From a Probability-Based Panel of Internet Users. *Social Science Computer Review, 32*(4), 544-560. doi:10.1177/0894439313510482

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, Position, and Order. Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly, 68*(3), 368-393. doi:10.1093/poq/nfh035

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin, 103*(3), 299.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paulo: Cambridge University Press.

Turner, G., Sturgis, P., & Martin, D. (2015). Can Response Latencies Be Used to Detect Survey Satisficing on Cognitively Demanding Questions? *Journal of Survey Statistics and Methodology, 3*(1), 89-108. doi:10.1093/jssam/smu022

Yan, T., & Tourangeau, R. (2008). Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times. *Applied Cognitive Psychology, 22*(1), 51-68.

APPENDIX A

Table A.1 Survey participation statistics

		N			Completion Rate
Survey	Field Period	accepted invitation	screened out & rejected	breakoff	in %
1	Apr 30 - May 05, ’09	4557	803	442	88.2
2	Mai 27 - Jun 05, ’09	2566	945	409	74.8
3	Jul 03 - Jul 13, ’09	1820	272	415	73.2
4	Jul 31 - Aug 11, ’09	1927	176	607	65.3
5	Aug 24 - Sep 01, ’09	1879	228	512	69.0
6	Sep 18 - Sep 27, ’09	1634	268	213	84.4
7	Sep 29 - Oct 08, ’09	2163	623	393	74.5
8	Dec 10 - Dec 20, ’09	1803	275	397	74.0
9	Apr 15 - Apr 23, ’10	1563	222	205	84.7
10	Jun 24 - Jul 05, ’10	1671	290	243	82.4
11	Sep 16 - Sep 26, ’10	1858	586	124	90.3
12	Dec 09 - Dec 19, ’10	1636	357	135	89.4
13	Mar 09 - Mar 19, ’11	1604	246	221	83.7
14	May 23 - Jun 03, ’11	1618	185	283	80.3
15	Aug 24 - Sep 03, ’11	1643	316	169	87.3
16	Dec 08 - Dec 18, ’11	1640	303	223	83.3
17	May 02 - May 15, ’12	1709	427	266	79.3
18	Sep 17 - Oct 01, ’12	1517	254	188	85.1
19	Jan 04 - Jan 19, ’13	1532	326	172	85.7
20	May 24 - Jun 08, ’13	1626	350	228	82.1
21	Sep 09 - Sep 21, ’13	1373	184	177	85.1
22	Nov 29 - Dec 13, ’13	1648	384	215	83.0
23	Feb 21 - Mar 07, ’14	1493	265	205	83.3
24	May 09 - May 23, ’14	1446	199	203	83.7
25	Aug 29 - Sep 13, ’14	1404	231	162	86.2
26	Nov 21 - Dec 05, ’14	1446	174	253	80.1
27	Feb 27 - Mar 13, ’15	1375	165	181	85.0
28	Jun 05 - Jun 19, ’15	1569	388	162	86.3
29	Sep 11 - Sep 25, ’15	1460	282	151	87.2

APPENDIX B

Operationalization of Respondent Level Variables

This appendix describes the operationalization of the respondent level variables that rely on the questions asked in 29 surveys. These variables include education, age, interest in politics, strength of a respondent's identification with a political party, (intended) turnout to vote at a federal election, sex, and region of residence.

Education

We categorized respondents' formal education as *low, intermediate*, and *high*. Since the response options to the open-ended question regarding respondents' formal level of education slightly changed throughout the 29 surveys, we relied on a standardized scheme of coding. The qualification that enabled students to enter a university was coded as *high* education while completing secondary/high school was considered to be an *intermediate* education. Anything less than completing secondary/high school was categorized as *low* education. For analytical purposes, we treated the variable as a categorical variable (0 = low / 1 = intermediate / 2 = high).

Age

According to their age, we coded respondents in five categories: 18–29 years, 30–39 years, 40–49 years, 50–59 years, and 60 years or older. Age was measured differently throughout the surveys. In surveys 1–7, respondents had to select one of the following categories in a close-ended question: 18–29 years, 30–39 years, 40–49 years, 50–59 years, and 60 years and above. Since survey 8, respondents have been asked about their date of birth in an open-ended question. For analytical purposes, we treated the variable as a categorical variable (0 = 18–29 / 1 = 30–39 / 2 = 40–49 / 3 = 50–59 / 4 = 60+).

Interest in the survey topic

We measured respondents' interest in the survey topic by a question on their political interest. This question used a 5-point scale that was labeled *very strong, fairly strong, moderately, fairly weak*, and *very weak*. We recoded the answers *very strong* and *fairly strong* as high political interest; the answer *moderately* as intermediate political interest; and the responses *fairly weak* and *very weak* as low political interest. Accordingly, respondents' interest in politics was coded as *low, intermediate*, and *high*. Again, for analytical purposes, we treated this variable as a categorical variable (0 = low / 1 = intermediate / 2 = high interest in politics).

Strength of party identification

We asked a question regarding the strength of a respondent's political party identification once they had stated they identified with a political party in a previous question. They had to answer the question on a 5-point scale that was labeled *very strong, fairly strong, moderately, fairly weak*, and *very weak*. If respondents did not identify with a political party, we coded their strength of party identification as *none*. If respondents identified with a party and reported the strength to be *fairly weak, very weak*, or *moderately*, we considered this as *moderate* strength. We coded respondents with a party identification of *fairly strong* or stronger as *strong*. For analytical purposes, we considered this variable as a categorical variable (0 = none / 1 = moderate / 2 = strong).

Intention to vote

To investigate respondents' motivation to participate in an election, we differentiated between the respondents who intended to turn out to vote at a federal election and those who did not. All surveys, except survey 7, featured a question on whether respondents would take part in the next German federal election. The five response options were *certain to vote, likely to vote, might vote, likely not to vote,* and *certain not to vote*. We coded respondents that reported to be *certain* or *likely* to vote as *yes*, while we considered the other respondents to have *no* intention to turn out. In survey 7, which was fielded in the aftermath of the German federal election 2009, a question on the actual turnout (*yes* or *no*) was asked. We used this question to code respondents of survey 7 either as *yes* or *no* with respect to their intention to vote at an election. For analytical purposes, we considered this variable as a binary variable (0 = no / 1 = yes).

Sex

We asked respondents about their sex with the response options *male* and *female*. For analytical purposes, we created a binary variable (0 = male / 1 = female).

Region of residency

We asked the respondents in which federal state of Germany they currently were residing. We coded the federal states Schleswig-Holstein, Hamburg, Lower Saxony, Bremen, North Rhine-Westphalia, Hesse, Rhineland-Palatinate, Baden-Wuerttemberg, Bavaria, and Saarland as *West Germany*; and we coded Berlin, Brandenburg, Mecklenburg-Vorpommern, Saxony, Saxony-Anhalt, and Thuringia as *East Germany*. For analytical purposes, we created a binary variable (0 = East Germany / 1 = West Germany).