Erschienen in:
Callegaro, Mario/Baker, Reg/Bethlehem, Jelke/Göritz, Anja S./Krosnick, Jon A./Lavrakas, Paul J. (Hrsg.): Online Panel Research: A Data Quality Perspective. Chichester: Wiley, 2014, 238-262

The Impact of Speeding on Data Quality in Nonprobability and Freshly Recruited Probability-based Online Panels

Robert Greszki
Marco Meyer
Harald Schoen

University of Bamberg
Feldkirchenstrasse 21
D-96045 Bamberg

Manuscript (January 2014)

Introduction

Online surveys have become a widely used tool in survey research (see for example Baker et al., 2010; Bethlehem & Biffignandi, 2012; Tourangeau, 2004). They differ from traditional survey techniques in a number of respects. Some of them, including low costs and short field times, make them attractive. Other features are more ambivalent, including the absence of an interviewer (see for example Chang & Krosnick, 2010; Schaeffer, Dykema & Maynard, 2010). While the absence of an interviewer might reduce social desirability effects, it also implies that the interviewing process is uncontrolled and accordingly data quality might be low (Groves et al., 2009). In particular, measurement problems arising from inattentive respondents are likely to be quite pervasive in online surveys.

Analysts of data gleaned from online surveys thus face a threat to data quality arising from inattentive respondents especially due to the uncontrolled interviewing situation. Aiming at valid results, they have strong incentives to identify low-quality responses and to remove them from analyses. Identifying low-quality responses is not an easy task, however. To be sure, some low-quality responses might be detected using conventional indicators like item-specific nonresponse or "don't know" answers. But sophisticated respondents who want to get material rewards for completing surveys might not choose these responses because otherwise they risk being easily identified as inattentive. Rather, they might straightline in item batteries, choose frequently middle categories, or any (random) answer to quickly get through the questionnaire and receive rewards. One way to identify those low-quality responses is to utilize response time as an indicator of data quality. Building on the fact that reading questions and processing information requires time (e.g. Tourangeau, Rips & Rasinski, 2000), particularly quick responses, the so-called speeding, might indicate minor data quality. Accordingly, removing those responses from the data might improve the quality of the data and help avoiding biases in substantive results.

In this chapter, we thus explore whether removing data on the basis of response time affects substantive findings. If it turns out that response time is a valid indicator of data quality and removing “too fast” responses from the data set changes substantive findings, this might be a viable strategy to increase data quality in online surveys. Otherwise, i.e. if response time is a poor indicator of data quality or removing “speeders” does not alter findings, this strategy might be not applicable.

We will study the prevalence and impact of speeding in a probability-based and nonprobability online panel (see Chapter 1 in this handbook; Couper, 2000) to explore whether the nature of the sample makes a difference. It might be argued that inattentive respondents are more prevalent in nonprobability than in probability-based samples, because the former sampling is more self-selective and might attract more persons who are likely to satisfice. At the same time, self-selection might be driven by factors that are not closely tied to satisficing, e.g. profound interest in the survey topic. It is thus an empirical question whether the samples differ in the prevalence and the impact of speeding on substantive findings.

The remainder of the chapter is structured as follows. The next section outlines the theoretical framework underlying the notion that speeding might be a symptom of low data quality. After a short description of the data, we will demonstrate that response time is a reasonable indicator of data quality and we will describe the page-specific procedures to identify too fast responses. The analyses show that the nonprobability survey is somewhat more plagued by speeding than the probability-based survey. In both, however, removing too fast responses from the data sets does not alter substantive findings in terms of marginal distributions and multivariate models. The chapter concludes by summing up key findings and discussing implications.

Theoretical Framework

According to the total survey error framework (Groves et al., 2009), measurement error is one of the most serious sources of low data quality. Such "… a departure from the true value of the measurement as applied to a sample unit and the value provided" (Groves et al., 2009, p. 52) can be caused by respondents as well as external factors. Responses given by inattentive respondents, who do not pay close attention or give not a sufficient level of thought to their responses, might prove invalid (Baker et al., 2010). As we study just two cases, effects of external factors like questionnaire design or technical obstacles cannot be examined in this contribution. The empirical analysis in this chapter, rather, focuses on problems arising from inattentive respondents. Nevertheless, potential effects of external factors will be addressed if applicable and necessary.

Given the lack of control in self-administered online interviews, as compared to interviewer-administered surveys, inattentive respondents are likely to be particularly prevalent in online surveys. To identify inattentive respondents, research institutes pursue different strategies (see for example Baker & Downes-Le Guin, 2007; Balden, 2008; Knapton & Garlick, 2007). For example, red herring questions are widely employed. These control questions are implemented in item batteries as a single item like "mark the option 'Neither like nor dislike'" to test whether respondents have carefully read the survey questions (Berinsky, Margolis & Sances, 2012; Miller, 2006; Miller & Baker-Prewitt, 2009). In a study of 13 U.S. online panels, Miller (2006) reported about five to ten percent respondents, who answered red herring questions wrongly. Moreover, employing red herring questions early in a survey appears to increase completion time and to decrease the likelihood of straightlining (Miller & Baker-Prewitt, 2009). Using straightlining as indicator of inattentiveness, Smith and Brown (2005) identified just one percent of the respondents in 20 large surveys as inattentive. Finally, Meade and Craig (2011) utilized a battery of eleven indicators to capture inattentiveness and found that between five and 15 percent of the respondents in an undergraduate internet survey lacked sufficient attention.

Moreover, respondents, although initially motivated, might become fatigued or distracted in the course of the interview (Krosnick, 1991). In line with this notion, Galesic and Bosnjak (2009) as well as Puleston and Sleep (2008) have demonstrated that respondents in online surveys appear to speed up in the course of the interview. This acceleration effect is likely to be stronger in online surveys than in interviewer-administered surveys, as the interviewer might be able to keep the respondent at the interview and positively influence the respondent to optimize only through his presence (Baker et al., 2010).

While online surveys raise specific issues concerning data quality, they might also provide a device to overcome this problem. Collecting survey data via the internet enables the automatic detention of paradata such as response time (see Couper, 2000, 2005; Heerwegh, 2003, 2011; Kaczmirek, 2009).[1] As survey software automatically captures the time a respondent spends on a specific survey page, researchers are in a position to explore response times regarding different pages and compare within and across respondents (see for techniques to measure response time in web surveys e.g., Fraley, 2004; Heerwegh, 2003; Kaczmirek, 2009). This information could prove helpful in identifying low-quality data – provided response time is linked to data quality.

To establish this link, we build on the response process model (Tourangeau, Rips & Rasinski, 2000; Tourangeau, 1984, 1987; Tourangeau & Rasinski, 1988; see also Kahn & Cannell, 1957) which suggests that the process of answering a survey question comprises four major steps. A respondent who fills in a web survey, first of all has to read the whole question text to comprehend a question. Reading only response options is insufficient and may lead to invalid answers. After comprehension of the question the respondent has to access the relevant information in memory before forming a judgment with this accessible information. Finally, the respondent formulates and reports an answer by clicking a response option or writing his answer in a textbox. To be sure, some steps

[1] Prior to the use of automatically generated paradata in web surveys, response times were already used in CATI surveys to measure, e.g., attitude accessibility (see Bassili, 1993, 1996; Bassili & Fletcher, 1991; Bassili & Scott, 1996; Draisma & Dijkstra, 2004; Johnson, 2004). In contrast to paradata in web surveys, response times in CATI surveys are usually gathered manually by interviewers.

might be skipped depending on the kind of question or the accuracy of the respondent's answer (Tourangeau, Rips & Rasinski, 2000). But it is reasonable to assume that the first and the fourth step of the response process model – comprehension and reporting an answer – are indispensable for valid survey answers. Thus, the response process, by its very nature, takes some time.

Provided that the response process takes a certain period of time, extremely short response latencies of individual respondents compared to all respondents, can be considered as indicative of invalid response behavior and thus inattentiveness. In this line of reasoning, very quick answers result from a response process in which several necessary steps are skipped because respondents engage in satisficing, rather than attempt to give valid answers (Krosnick, 1991; Krosnick & Alwin, 1987; Krosnick et al., 2002; Narayan & Krosnick 1996).

This time-saving strategy might lead to different forms of behavior, depending on the questions asked. When dealing with single items, an inattentive respondent might somewhat satisfice by simply selecting the first (given) response option which she considers reasonable. Thus, steps two and three are done quite quickly or skipped completely. A stronger form of satisficing implies that a respondent does not read the question content properly. In the worst case, respondents perform only step four: the formation of a (random) answer, e.g., choosing always middle categories (Schuman & Presser, 1996; Tourangeau, Couper & Conrad, 2004) or selecting the first response option (Malhotra, 2008). When asked to answer item batteries using a grid pattern, some respondents may straightline, i.e. they answer the questions rapidly by clicking always the same response category in each 'line' of the grid without paying attention to the substantive meaning of response options. Provided with a 'don't know' category, some respondents perform all steps of the response process, requiring some time, and actually come to no answer. Others, however, might skip one or more steps of the response process, thereby strongly satisficing, and give a 'don't know' (for a more sophisticated view see Krosnick et al., 2002). Finally, in online surveys clicking 'next' without paying any attention to question content is a time-saving strategy, even if not all online surveys provide respondents with an opportunity to click ‘next’, which is equivalent to ‘no answer’ in interviewer-administered surveys. Clicking ‘next’, however, does not necessarily reflect a lack of motivation because some respondents might perform all steps of the response process and give no answer after thorough consideration.

In effect, there is a, though not perfect, link between very quick responses and low data quality which is supported by evidence (Callegaro et al., 2009; Malhotra, 2008; Rossmann, 2010). Yet, we have to keep in mind that raw response time might not be an appropriate indicator because respondents might differ in the time it takes to perform the task of providing valid responses. To give just two examples, cognitive ability as well as training might make a difference. Accordingly, it might be appropriate to utilize somewhat adjusted rather than raw response time as indicator of satisficing.

Nevertheless, it is reasonable to consider too fast responses as low-quality answers that have to be removed from the data before conducting substantive analyses. This conclusion raises the question for the appropriate yardstick to distinguish too fast from regular responses. Survey analysts know from pretests how long it usually takes to complete the survey. Assuming a usual response time of, e.g., 25 minutes, experienced respondents might be able to complete the questionnaire in 15 to 20 minutes without any loss in data quality. If a respondent manages to complete this survey in five or ten minutes, however, concerns about data quality will arise, even if a respondent is highly skilled and well trained. The same reasoning applies to the completion of individual pages within a survey. In effect, there is some objective information on reasonable response times. At the same time, the choice of a specific threshold to identify speeding is somewhat arbitrary.

Irrespective of the threshold chosen, analysts might remove too fast responses from the data before performing substantive analyses to get valid results. This raises the question whether this procedure will change the substantive results. Quite obviously, the answer to this question depends on the proportion of speeders in a sample. If just a tiny fraction of the sample gives too quick responses, removing them will hardly change substantive findings. Moreover, the distribution of too quick answers plays a role.

Concerning marginal distributions, removing too quick answers will change results considerably if speeding is correlated with the variable in question. If speeders exhibit a low level of interest in politics, for example, removing speeders will increase the level of political interest in the sample. When it comes to correlations and multivariate models with such biased variables, a similar reasoning applies (see, e.g., Faas & Schoen, 2006; Schoen, 2004).

These factors are not completely independent of *how* respondents speed. If a respondent chooses a ‘don’t know’ answer or just clicked next (‘no answer’), removing too-fast answers will not make a difference because those answers are treated as missing values anyway. If respondents chose any (random) answer, thereby skipping steps two and three or even steps one to three of the response process model, marginal distributions will not be affected by the exclusion of speeders, whereas correlations might be altered. The impact of satisficing by always choosing middle categories or straight-lining might not easily be identified. Moreover, it is, by and large, an empirical question whether, and in which way, removing those speeders from a data set alters substantive findings.

The prevalence and effects of speeding might, inter alia, depend on the nature of samples. It might be argued that nonprobability and freshly probability-based recruited samples differ in several respects that may affect also the motivation of respondents (Bethlehem & Biffignandi, 2012; Coen, Lorch & Piekarski, 2005; Topoel, Das & Van Soest, 2008), and thus, may vary also in the proportion of speeders. Compared to freshly probability-based recruited respondents, nonprobability online panels are in most instances recruited via a highly self-selective process, and therefore, likely to attract other persons with other intentions in a first instance compared to the random recruitment methods. It is thus reasonable to compare nonprobability-based and probability-based samples in terms of the prevalence and effects of speeding on substantive findings. Yet, it remains an empirical question whether there are really relevant differences.

In sum, speeding through web surveys is a potential threat to data quality. What is more, it cannot be taken for granted that speeding is irrelevant for the substantive findings gleaned from online surveys. In the subsequent sections of this chapter, we will explore these questions using data from a nonprobability online survey and a freshly probability-based recruited online survey.

Data and Methodology

To explore speeding in a nonprobability online survey and a freshly recruited probability-based online sample, we use data from the *German Longitudinal Election Study (GLES)* and data from the *American National Election Study (ANES)*. The 2008-2009 ANES Panel Study was fielded from January 2008 through September 2009, including 21 monthly panel waves. For our analysis, we utilize data from the first panel wave conducted in January 2008 in order to study ‘freshly recruited’ respondents who have not been subject to panel conditioning and are most different from online panelists. The target population comprises a representative sample of U.S. citizens aged 18 or older as of Election Day in November 2008. Respondents were telephone-recruited using RDD. In the first contact, a short recruitment interview was conducted and respondents were offered $10 per month to complete the surveys on the Internet. Willing respondents without a computer and Internet connection were provided with a free web appliance for the duration of the study (see DeBell, Krosnick, & Lupia, 2010).

The German election study includes a series of online surveys with a nonprobability online panel as sampling frame. Online surveys were conducted in the run-up to and after the 2009 German federal election. As substantive findings do not vary across surveys, we report the results from the survey conducted from 18th to 26th September 2009 (results from the other surveys are available from the authors). The dataset comprises 1,153 respondents who were drawn in a quota sample, where the panelists were quoted in terms of gender, age and education via a nonprobability online panel (see Rattinger et al., 2009). The quotas were a mix of the distribution in the target population and the current online population. The target population comprises all German citizens who were eligible to vote and thus differs considerably from the frame population of the online panel. The online panel comprised about 65.000 active panelists in Germany in 2009. As defined by the panel provider, active panelists are those persons who completed the double-opt-in registration, completed the master questionnaire about basic personal information, and successfully participated in at least one survey within the last twelve months. The panel uses different on- and offline channels to recruit new panelists which include opinion portals, on-site surveys, search engines, and recruitment by telephone. For participating in the surveys, members of the online panel are offered incentives, namely 10 panel-points per minute, which, in 2009, was the equivalent of approximately 0.10 €. Having collected at least ten Euros, panelists may choose between cash payment, shopping coupons, or a donation (GESIS, 2009a, 2009b; Respondi, 2009a, 2009b).

Both surveys were designed to be completed within 30 minutes (GLES) and 25 to 30 minutes (ANES), respectively. Empirically, respondents on average spent 35 to 36 minutes to complete either survey. Due to the right-skewed distribution of response times (Ratcliff, 1993), the median is considerably lower than the mean with 33 minutes in the GLES study and about 27 minutes in the ANES surveys. The fastest respondents, however, managed to complete the GLES survey in two minutes and the ANES survey in six minutes. Given the design of the surveys and the empirical distribution of response times, these minimal response times are indicative of speeding by inattentive respondents. Leaving aside these extreme cases, it is an empirical question whether short response times indicate low data quality. This is the question we address in the next section.

Response Time as Indicator for Data Quality

Using response time in general, and speeding in particular, as indicator of data quality rests on the assumption that valid indicators of data quality are correlated with response time. We consider four well-known indicators that might reflect satisficing behavior which allows respondents to complete a survey particularly quickly (see codebook for a detailed description of the four indicators). First, it is quite straightforward to interpret giving ‘no answer’ as time-saving response strategy. A respondent who clicks ‘next’ instead of giving a substantive answer simply skips virtually all steps the response process model includes. It thus does not come as a surprise that item-nonresponse is often correlated with low data quality (see for example de Leeuw, 1992; Groves, 1991). Second, we calculated the proportion of ‘don’t know’ answers because choosing this response option is for some respondents a well-known satisficing strategy (Krosnick, 1991; Krosnick et al., 2002). Third, choosing the middle category from five or more response options on an ordinal scale is another satisficing strategy (Kaminska, McCutcheon & Billiet, 2010; Krosnick, Narayan & Smith, 1996). We thus calculated the proportion of cases in which respondents chose the middle category when offered at least five response options in ordinal scales. Finally, we measured straight-lining by capturing systematic response patterns that do not reflect substantive preferences (Kaminska, McCutcheon & Billiet, 2010; Krosnick et al., 1996). We considered item-batteries comprising at least five items with at least five response options each. In analyzing the data, we were careful not to mistake substantive responses, e.g. consistently high ratings for a politician, for straight-lining. In effect, we calculated for each respondent the proportion of pages with straight-lining as compared to the number of pages with opportunities for straightlining (for a detailed description of the procedure see the codebook). To be sure, these indicators are not perfect because these kinds of response behavior might in some instances reflect substantive answers, rather than satisficing. In light of prior research (Kaminska, Goeminne & Swyngedouw, 2006; Kaminska, McCutcheon & Billiet, 2010; Krosnick et al., 1996), they are reasonable choices, however.

These indicators are easily applicable to the GLES data but not to the ANES survey because question formats differed considerably across surveys. First of all, the ANES questionnaire does not include item batteries that are displayed in grid fashion as well as it rarely includes items with at least five response options. Moreover, when ANES respondents clicked ‘next’ without choosing any response option they did not get immediately to the next page. Instead, the question was displayed once again and the respondent was asked to provide his best answer even if he was not completely sure. Only if she clicked ‘next’ for a second time, the question was skipped and recorded as a ‘no answer’. Thus, the warning message automatically forced respondents to spend some additional time on the respective survey page. As a consequence, this part of the analysis is confined to GLES data.

We calculated the four quality indicators by dividing the number of actual satisficing by the number of opportunities to satisfice. In order to explore whether the quality measures are correlated with response time, we regressed the overall response time spent on the survey pages that are relevant for the respective quality measure on the quality indicators, using OLS regression (Table 1).

As the results reported in Table 1 show, indicators of satisficing are negatively correlated with response time. Straight-lining is most strongly correlated with response time. According to the evidence, a one-unit increase in straight-lining is accompanied by a decrease of 27 seconds on the respective survey sites. The correlations of response time with ‘no-answers’ and ‘don’t know’ answers prove also statistically significant, though they are somewhat smaller. The impact of the middle-category index on response time, however, is indistinguishable from zero. Accordingly, we might speculate that GLES respondents chose middle categories not primarily to satisfice. Notwithstanding the latter result, the evidence suggests that response time is correlated with – imperfect – measures of satisficing. It is thus reasonable to utilize response time as an indicator of the attention respondents paid to the survey and ultimately of data quality.

--- Table 1 about here ---

Due to data limitations, the above analysis was confined to GLES data. Yet, it is reasonable to consider the GLES results, by and large, applicable to the ANES data and to utilize response times as an indicator of data quality in both surveys. This conclusion rests on the fact that for a large number of comparable items the distributions of response times in both surveys resemble each other quite closely. Two give just two examples,

Figure 1 contains the distributions of response time in both surveys for political interest and the question of how often respondents talk about politics. The GLES and ANES distributions are quite similar. Moreover, the curves reach their peaks at rather short response times. The proportion of very fast respondents might be explained by the absence of an interviewer in self-administered online surveys. In interviewer-administered studies such short response times to similar questions are very unlikely, hence the question and response options had to be read first of all by the interviewer. Given the similarities in the distributions, we conclude that responses in the ANES and the GLES surveys are subject to similar regularities. It is thus warranted to utilize response time as – proxy – indicators of data quality in both surveys.

--- Figure 1 about here ---

Comparing the distributions of response times in both surveys, it turns out that the median is considerably lower in GLES rather than in ANES. This finding suggests that the former survey is somewhat more plagued by inattentive and thus speeding respondents. This pattern hints to the fact that nonprobability online surveys are conducive to larger problems arising from satisficing than probability-based surveys. Yet, the evidence is suggestive rather than conclusive because the two surveys under study differ in a number of respects, and not just in terms of sampling procedures.

How to measure ‘speeding’?

As response time is correlated with data quality it appears to be reasonable to exclude too quick answers from data before performing substantive analyses. This suggestion raises two related questions. First, we have to decide whether to identify speeding case- or page-wise. The case-wise procedure, as used by several research institutes (see for example GESIS, 2009b), rests on the assumption that some respondents answer too quickly to all questions. This notion is at odds with research on satisficing which suggests that respondents' attention varies in the course of a survey (Krosnick, 1991). What is more, previous research suggests that speeding is, by and large, not a stable characteristic of respondents but respondents respond regularly to some questions, whereas they satisfice when answering other items (Greszki, Meyer, & Schoen, 2012). We therefore use a page-specific strategy that aims at detecting too fast responses rather than too fast respondents. To this end, we calculate for each survey page, i.e. for all items on a page alike, the difference between a person's response time and the median response time (see also GESIS, 2009b; Rossmann, 2010).

Second, we have to identify a threshold to distinguish too fast responses from regular responses. One can easily imagine that different criteria lead to different substantive findings (see, e.g., Meyer & Schoen, in press). Moreover, despite some reasonable upper and lower limits, the ultimate choice of a specific threshold is somewhat arbitrary. To avoid results that crucially depend upon an arbitrarily chosen criterion, we employ three different thresholds. The most inclusive criterion identifies those responses as too fast who were given more than 30 percent faster than the median response time. According to the second measure, answers which were given more than 40% faster than the median response time are flagged as speeding. The most exclusive measure employs on the 50 percent criterion.[2] Being more than 50 percent faster than the median response is a rather hard criterion. The latter measure might thus be considered as being capable of excluding very fast respondents and being simultaneously insensitive to 'false positives' (see GESIS, 2009b; Rossmann, 2010).

--- Figure 2 and Figure 3 and Table 2 about here ---

[2] The ANES survey makes extensive use of automatic conditional branching on the same page of the parent item. That means, that dependent on the response choice, follow-up questions appear on the same screen below the original question. In terms of speeding it is warranted to take this into account because survey pages appear different depending on respondents' response option. For example when a respondent is asked whether he likes or dislikes the Democratic Party, a follow-up question is displayed on the same screen that asks if he likes or dislikes the Democratic Party a little, a moderate amount or a great deal. On the contrary respondents choosing 'neither nor' on the parent item, no follow-up question is displayed. Thus, such respondents are naturally faster because they only answer one question on that survey page. We take this into account and calculate separate page medians for those with and without follow-up questions.

Figures 2 and 3 show the percentage of flagged speeders for the first 50 survey pages according to these three criteria. Moreover, Table 2 reports the mean numbers of speeders across survey pages when applying the three criteria. First, the evidence demonstrates that the proportion of speeders is small to modest in the ANES and the GLES survey. Employing the most exclusive criterion results in less than ten percent, while the most inclusive measure suggests that roughly one in five respondents speeds through the survey. To be sure, twenty percent is a considerable proportion. But even this percentage indicates rather a minority and it is likely to include a considerable proportion of ‘false positives’.

Moreover, the evidence suggests that thresholds make a difference in the proportion of respondents flagged as speeders. As the criterion for speeding becomes more exclusive, the number of speeders declines considerably. To give just an example, in the GLES data applying the 30 percent criterion, on average some 22 percent of the respondents are flagged as speeders. Utilizing the 40 percent threshold, speeders approximate 15 percent of all respondents, whereas the 50 percent criterion leads to eight percent speeders. In effect, choosing a criterion might considerably affect conclusions about the pervasiveness of respondent inattentiveness.

Finally, comparing the findings from ANES and GLES, we find that for the 40 and 50 percent criteria, ANES data exhibit a significantly smaller proportion of speeders than the GLES survey. Yet, the differences are modest in substantive terms. Moreover, the 30 percent criterion yields virtually identical proportions of speeders in both surveys. So, the evidence supports the notion weakly that the freshly probability-based recruited respondents exhibit less speeding than self-selected respondents, if at all.

As already mentioned, raw response times (and the according medians) might be biased because respondents – due to differences in, e.g., cognitive ability and training – might differ in the time they need to provide a valid response. With respect to this kind of respondent characteristics, however, the GLES and ANES surveys differ considerably. To make our conclusion more convincing and to demonstrate that response times can be employed as quality indicator in both surveys alike, we tested two rivaling explanations that focus on the composition of the ANES and GLES samples.

First, GLES respondents are considerably older and better educated than ANES respondents. As cognitive ageing and cognitive abilities appear to have an impact on response behavior (Malhotra, 2008; Yan & Tourangeau, 2008), we might expect that younger, more internet-experienced, and better educated respondents handle online questionnaires more efficiently and thus answer faster. Provided the validity of these assumptions, the age and education differences between ANES and GLES samples might lead to between-samples differences in the prevalence of speeding.

In order to explore this hypothesis, we divided respondents into five age groups and three education groups and calculated page medians per group. Tables 3 and 4 first of all report how many respondents are defined as speeders according to the three criteria (as already reported in Table 2). Additionally, the tables report the proportion of respondents flagged within each group, i.e. the proportion of respondents who answer too quickly as compared to their fellow respondents in the respective age or education bracket (right-hand columns). For example, applying the 50 percent criterion to GLES data with no group correction, we see the expected pattern that twelve percent of the young respondents are flagged, but only two percent of those 60 years and older. A similar pattern applies to the ANES data, although the proportions of speeders are smaller.

--- Table 3 and table 4 about here ---

The within-group page medians yield a somewhat more balanced distribution of speeding flags across groups in the GLES and the ANES sample. In the GLES survey, for example, when the 50 percent group-specific measure is employed, the proportion of speeding flags ranges from roughly six percent among the oldest respondents to 8.6 percent in the 18-29 years group as compared to a two-to-twelve percent range when the 50 percent overall speeding measure is utilized. Similar patterns apply to education groups. But even when the group-based speeding measures are employed, we find a higher proportion of speeders in the GLES sample than in the ANES sample. As a result, the differences in the prevalence of speeding between the two samples cannot be explained by compositional differences in terms of education and age.

Second, whereas ANES respondents were recruited 'freshly' for the first panel-wave, the GLES respondents are regular members of an online panel and are more likely to have answered surveys before. As more experienced respondents are supposed to answer faster, this difference in sample composition might account for the differences in speeding between ANES and GLES. To explore this hypothesis, we once more utilized GLES data. In particular, we regressed the proportion of survey pages that a respondent completed faster than the respective median respondent on indicators of experience (e.g., number of completed surveys in the previous four weeks, duration of panel membership). Moreover, we included predictors capturing the recruitment process. As it turns out, these predictors explain just three percent of the variance in the dependent variable (Table 5). As concerns experience, panel membership exhibits a positive effect, suggesting that the longer a respondent is active in the panel, the more likely she is to have a higher amount of survey pages below the median. This effect, however, is far from being substantively relevant. Moreover, the number of surveys completed in the four weeks prior to the current survey does not exhibit a statistically significant effect. However, as respondent activity in online panels might vary over different time periods, considering only the number of completed surveys in the past four weeks might be a rather weak indicator of experience. Therefore, we also tested the number of completed surveys in larger time frames (last 12 weeks, last 12 months). These additional analyses exhibit no different results, suggesting that survey experience does not make a difference.[3]

--- Table 5 about here ---

[3] Yet, we have to keep in mind that we have only information about experience in the GLES panel. Respondents, however, might also be members of other panels on which we have no information.

In sum, survey experience, by and large, does not affect response time. Age and education make a difference in response time, but age and education differences between ANES and GLES do not account for the differences in speeding. Having shown that these alternative explanations do not account for the differences in speeding, we are in a better position to conclude that response time is a functionally equivalent indicator across surveys. Accordingly, we might conclude that the nature of the recruitment process – probability-based vs. nonprobability – makes a small difference in the proportion of speeders in the ANES and the GLES samples. We have to keep in mind, however, that the surveys differ also in questionnaire design for which we could not account for in our analysis. Given the evidence, we are thus cautious drawing strong conclusions about causal effects of the recruitment process.

Does speeding matter?

Having established that speeding, as measured above, occurs in the ANES and the GLES surveys, we now turn to the issue of whether it matters for substantive findings. As speeding is correlated with low-data quality, not removing ‘too fast responses’ from the data set might lead to biased results. To address this question, we examine whether excluding page-specific speeders from the analysis alters substantive findings both in terms of marginal distributions and parameters of interest in explanatory models. The results are gleaned from analyses using uncorrected speeding flags; using age and education corrected flags does not alter the substantive findings we report in the remainder.

We explore this question by analyzing two phenomena that are some kind of standard topics in research on public opinion and political behavior. For one thing, we analyze attitudes towards the head of government’s handling of the economy (e.g., Duch & Stevenson, 2008; Alesina & Rosenthal, 1995). For another, we address electoral participation (e.g., Blais, 2000; Hansen, 1975). Given the comparative nature of our analysis, we utilized explanatory models that include variables which were, by and large, covered both by ANES and GLES. More sophisticated models could not be used due to a lack of appropriate data.

Starting with evaluations of the president's and chancellor's handling of the economy, Tables 6 and 7 report the marginal distributions in both surveys. In each table, the left-hand column contains the results for the sample without any speeders removed. The remaining columns show the marginal distributions for the samples when speeders, as defined by the three above criteria, are excluded. The evidence clearly shows that marginal distributions do not considerably change when speeders are excluded, irrespective of the threshold. Although a somewhat larger proportion of the GLES respondents are flagged as speeders, neither in GLES nor ANES removing speeders does alter substantive findings.

--- Table 6 and table 7 about here ---

Turning to explanatory models, we model evaluations of handling of the economy as a function of party evaluations, ideological self-placement, and the respondent's perception of the national economy (all variables rescaled to run from 0 to 1). The results of the respective OLS regressions are reported in Tables 8 and 9. Once again, in each table the left-hand column reports the baseline model with no speeding respondents excluded; only those respondents are removed who exhibit missing values on at least one variable, i.e. 'don't know' or 'no answer' responses, in the analysis. The right-hand columns contain the models without those respondents being excluded as speeders according to the three different criteria. We additionally included columns that report the respective coefficient differences to the baseline model. Comparing models with and without speeders is warranted because alternative strategies, e.g., including interactions between speeding indicators and substantive predictors, imply that respondents who speed on the dependent variable cannot be adequately adressed.

--- Table 8 and table 9 about here ---

The evidence shows that in US model, the four independent variables exhibit statistically significant effects in the expected direction. Accordingly, approving of the Republicans and subscribing to conservative ideology, e.g., makes citizens more favorable of George W. Bush's handling of the economy. What is more, the substantive conclusions concerning the predictors of the President's handling of the economy are independent of whether speeders are excluded or not. In fact, the largest difference between coefficients in baseline model and a model with speeders removed amounts to a tiny 0.04.

Turning to GLES data, the evidence supports an identical conclusion. To be sure, there are some differences. For one thing, two predictors of economic evaluations, i.e. ideology and evaluations of the Social Democrats, prove statistically insignificant in the baseline model. For another thing, we find somewhat more sizable differences between the baseline model and the models with speeders excluded. In particular, ideology turns out to pass conventional levels of statistical significance when speeders are excluded according to the 30 or the 40 percent criterion. The differences between models and the baseline model are so small, however, that they prove neither statistically significant nor substantively relevant.

We thus conclude that both in ANES and GLES, the exclusion of speeders does not affect substantive conclusions about the distribution of attitudes toward the head of government's handling of the economy and about the determinants of these attitudes. Irrespective of whether speeders are removed, the substantive message is the same.

Next we address electoral participation. In both surveys, respondents were asked whether they intend to participate in the next federal / presidential election. Whereas the ANES survey offered two response options (yes/no), respondents had five response options in the German survey running from 'definitely' to 'definitely not'. To ease comparison, we dichotomized the German answer scheme ('definitely', 'probably' = 1; 'perhaps', 'probably not', definitely not'=0). Again, we address marginal distributions before exploring the determinants of (prospective) turnout.

The marginal distributions reported in Table 10 demonstrate that the exclusion of speeders does not make a difference in results on turnout intention. As the first column in Table 10 indicates, some 92 percent of the respondents were likely to vote. This percentage is much higher than the actual turnout in both elections and thus reflects sampling problems as well as measurement error (overreporting) in both surveys (see e.g., Bernstein, Chadha & Montjoy, 2001; Silver, Anderson & Abramson, 1986). Removing speeders from the sample does not alter the proportion of voters, irrespective of the criterion for speeding. The reason for this finding is the fact that speeders are similarly frequent among voters and nonvoters in the ANES and the GLES data.

--- Table 10 about here ---

Turning to multivariate models, we perform logistic regression analyses with turnout intention as dependent variable. Due to data limitations, we cannot perform models with exactly identical sets of predictor variables in ANES and GLES. But the ANES and GLES models alike build on the notion that electoral turnout is driven by political involvement, which includes interest in politics, feelings of political efficacy, and citizen duty. The results of these analyses are reported in Tables 11 and 12, with the left-hand columns including the baseline models.[4]

Starting with ANES results, political interest and internal efficacy appear to be particularly powerful in shaping electoral turnout. Comparing these findings to the results from the analyses with speeder corrections, we find that the coefficients on political interest, party identification, and – to a smaller extent – internal and external efficacy are

[4] In an attempt to explore whether cognitive ageing and cognitive abilities do matter (Malhotra, 2008; Yan & Tourangeau 2008), we studied the distribution of age and education in terms of valid cases according to listwise deletion in the regression models (see codebook). Comparing the baseline model with the 40 and 30 percent rule, we find patterns consistent with findings in prior studies. Accordingly, young and highly educated respondents are more likely to be flagged as speeders. For the 50 percent threshold, the differences to the baseline model are rather marginal. These patterns suggest, that using the 50 percent rule leads to the exclusion of rather ‘real’ speeding behavior, that is independent from ‘predispositions’, such as age and education. In this vein, the 40- and 30-variants appear to be sensitive, but not very specific. In effect, young and highly educated respondents who do not satisfice but quickly give valid responses are removed from the data set.

quite stable across models. The results concerning the impact of talking about politics on turnout, however, differ somewhat across models, in particular between the baseline and the 30 percent-criterion model. But the differences in coefficients do not pass conventional levels of statistical significance. Moreover, we calculated from the models reported in Table 11 predicted probabilities of turnout for different levels of the independent variable of interest, i.e. talking about politics, while setting the remaining variables in the model to their respective mean (not reported in tables). The results suggest that moving from respondents who never talk about politics to those who talk about politics everyday increases the likelihood of turnout slightly from 94 to 98 percent. In the model with the 30 percent-speeder criterion, the increase is from 91 to 99 percent, i.e. seemingly larger but indistinguishable from the findings in the model without corrections. The same pattern, which might also reflect some kind of ceiling effect, applies to the remaining models and variables. So in ANES data, speeders do not make a difference.

--- Table 11 and table 12 about here ---

Turning to the GLES baseline model in Table 12, citizen duty and again political interest are important determinants of turnout in Germany. When taking into account models with speeder corrections, the results concerning citizen duty and – to a smaller degree – political interest and party identification turn out to be remarkably stable across models. By contrast, point estimates for campaign interest and satisfaction with democracy appear to vary somewhat across models. Whereas the former's impact appears to decrease when excluding an increasing number of speeders, the latter's effect on turnout appears to increase. None of the differences in coefficients passes conventional levels of statistical significance, however. Calculating predicted probabilities from the regression estimates lends additional credence to the conclusion that speeder corrections do not make a difference (not reported in tables). To give just an example, in the baseline model moving from the low to high campaign interest increases turnout from 94 to 98 percent, in the model with the 30 percent correction from 97 to 98 percent. Neither difference attains statistical significance. As a result, removing speeders from the analysis does not affect substantive conclusions.

In summary, we conclude that both in ANES and GLES, the exclusion of speeders does not affect substantive conclusions about the distribution of the intention to turnout and its predictors. Irrespective of whether speeders are removed and how they are defined, the substantive findings do not differ. As the findings on turnout parallel those on attitudes toward the head of government's handling of the economy, we might conclude that speeding, though existent, does not pose severe problems for scholars interested in substantive results.

Conclusion

This chapter addressed the phenomenon of speeding in web surveys as an indicator of inattentive respondents and explored the problems arising from it for data quality. Building on the response process model, we suggested that speeding implies skipping necessary steps in the process of providing a valid answer to a survey question. In this vein, response time might be used as an indicator of data quality in self-administered online surveys. Because of the absence of an interviewer, web surveys are notably encountered with the uncontrolled interview situation. Utilizing a page-wise procedure, we identified too fast responses and explored whether removing them from the data might affect substantive results in a nonprobability online panel and a freshly probability-based recruited sample.

Using data from these two surveys, we demonstrated that there is a considerable number of respondents who 'speed' through the pages of web surveys. However, our analysis demonstrates that the choice of criteria to identify speeders makes a difference in the proportion of speeders. Moving from the 30- to the 50-percent criterion decreases the proportion of speeders considerably. These differences suggest that the more lenient criteria are likely to identify 'false positives' as speeders whereas stiffer criteria might yield some ‘false negatives’. In substantial terms, however, the proportion of speeders is quite small unless the very lenient 30-percent criterion is employed. So, speeding does occur, but it is clearly a minority of our samples that exhibit speeding behavior.

We explored the prevalence and effects of speeding in a nonprobability sample and a probability-based sample. The findings did not differ systematically across samples. To be sure, relying on the more exclusive criteria the nonprobability GLES sample yielded a somewhat higher percentage of speeders than the probability-based ANES sample. But these differences were not so sizable that they led to different substantive conclusions about the prevalence of speeding. When it comes to the effect of speeding on substantive findings, both samples led to identical conclusions. Irrespective of whether studying marginal distributions or multivariate models, irrespective of which threshold is applied, excluding too fast responses does not make a considerable difference in substantive results. We may caution, however, that we have evidence from just two cases that, in addition, differ not only in terms of recruitment. These qualifications notwithstanding, we might conclude that the evidence does not support the notion that the non-/probability recruitment differs considerably or is of critical importance when it comes to speeding.

Nevertheless, this finding is good news for online-based survey research. There are indeed inattentive respondents who ‘speed’ through the pages, but speeding appears not to bias substantive findings. We might account for this pattern by two related arguments. On the one hand, obvious speeders are often already excluded from analysis because of invalid missing data (e.g. item nonresponse). On the other hand, for valid answers, speeding is not systematically linked to certain variables or parameters of interest in our analyses. In this line of reasoning, there appear to be some variables that are not vulnerable to biases resulting from speeding.

As already mentioned, this research is subject to several limitations. The most severe limitations stems from the limited number of data sets available. Future research should thus utilize more data from more diverse online surveys. This approach would permit scholars to explore the prevalence and substantive impact of speeding in different societies, in surveys on a diversity of topics, with different speeding measures, in samples that comprise professional and novice respondents who were recruited in different ways. This kind of evidence might lend additional credence to the notion that speeding, though existent, does not affect substantive results. Provided this finding, scholars might conclude that inattentive respondents are not a real problem for online surveys. It cannot be taken for granted, however, that the evidence will not suggest that speeding poses severe problems for analyses of online data that were collected on specific topics, in particular samples or societies.

A related approach does not aim at assembling a diversity of data but identifies potential predictors of speeding and its substantive relevance in the first place and then employs a (quasi-)experimental design to gauge the impact of the factor of interest. For example, scholars might study the impact of certain features of online surveys, e.g. batteries of grid questions, on respondent motivation and speeding behavior. In order to study these questions, scholars might find it convenient to not only analyze response times but might employ also other techniques like eye-tracking that permit to scrutinize respondent behavior in more depth. In identifying factors conducive to respondent attention, scholars might also find complex interactions between respondent characteristics and features, be they technical or substantive, of the survey in shaping respondent motivation. Here, a valuable strategy to study the role of dispositional traits in the interplay with survey features might be also employing experimental designs in panel studies.

Identifying dispositional traits and survey features that are conducive to respondent motivation and thus presumably valid responses might be considered a first step toward strategies that do not aim at detecting speeders but at avoiding speeding at all. In this vein, scholars might consider about how to design online surveys in order make or keep all respondents sufficiently attentive to give valid responses. This research agenda might lead to the conclusion that there is a single optimal strategy for all respondents and topics. Alternatively, scholars might learn that the suitability of survey features varies across topics or respondents, thereby raising additional issues of comparability. Probably, findings will also change over time as online surveys become more pervasive and respondents more acquainted to them. In any event, we consider respondent attentiveness, and speeding as an indicator thereof, as a topic that warrants scholarly attention in the future.

References

Alesina, A. & Rosenthal, H. (1995). *Partisan politics, divided government, and the economy*. New York: Cambridge University Press.

Baker, R., Blumberg, S.J., Brick, M., Couper, M.P., Courtright, M., Dennis, J.M., et al. (2010). Research Synthesis. AAPOR report on online panels. *Public Opinion Quarterly*, *74*, 711-781.

Baker, R. & Downes-Le Guin, T. (2007). Seperating the wheat from the chaff: Ensuring data quality in internet samples. In M. Trotman, T.Burrell, L. Gerrard, K. Anderton, G. Basi, M. Couper, K. Moris, D. Birks, A.J. Johnson, R. Baker, M. Rigg, S. Taylor & A. Westlake (Eds.), *The challenges of a changing world: Proceedings of the fifth international conference of the association for survey computing* (pp. 157-166). Berkeley, UK: ASC.

Balden, W. (2008). An unwanted impact. Retrieved June 28, 2013, from *http://www.quirks.com/articles/2008/20080103.aspx?searchID=765725748&sort=5&pg=1*

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*, 331-346.

Bassili, J.N. (1993). Response latency versus certainty as indexes of the strength of voting intentions in a CATI survey. *Public Opinion Quarterly, 57*, 54-61.

Bassili, J.N. (1996). The how and the why of response latency measurement in telephone surveys. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 319-346). San Francisco, CA: Jossey-Bass.

Bassili, J.N. & Scott, B.S. (1996). Response latency as a signal to question problems in survey research. *Public Opinion Quarterly, 60*, 390-399.

Berinsky, A.J., Margolis, M. & Sances, M.W. (2012). *Separating the shirkers from the workers? Making sure respondents pay attention on internet surveys.* Paper presented at the NYU CESS 5th Annual Experimental Political Science Conference, New York, NY.

Bernstein, R., Chadha, A. & Montjoy, R. (2001). Overreporting voting: Why it happens and why it matters. *Public Opinion Quarterly, 65*, 22-44.

Bethlehem, J. & Biffignandi, S. (2012). *Handbook of web surveys*. Hoboken, NJ: Wiley.

Blais, A. (2000). *To vote or not to vote. The merits and limits of rational choice theory*. Pittsburgh, PA: University of Pittsburgh Press.

Callegaro, M., Yang, Y., Bhola, D.S., Dillman, D.A. & Chin, T. (2009). Response latency as an indicator of optimizing in online questionnaires. *Bulletin de Méthodologie Sociologique, 103,* 5-25.

Chang, L. & Krosnick, J.A. (2010). Comparing oral interviewing with self-administered computerized questionnaires. An experiment. *Public Opinion Quarterly, 74,* 154-167.

Coen, T., Lorch, J. & Piekarski, L. (2005). *The effects of survey frequency on panelists' responses.* Paper presented at the worldwide panel research conference ESOMAR, Amsterdam, NL.

Cohen, J., Cohen, P., West, S.G. & Aiken L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral science*. New York, NY: Routledge.

Couper, M.P. (2000). Usability evaluation of computer-assisted survey instruments. *Social Science Computer Review, 18*, 384-396.

Couper, M.P. (2005). Technology trends in survey data collection. *Social Science Computer Review, 23*, 486-501.

DeBell, M., Krosnick, J.A., Lupia, A. (2010). *Methodology report and user's guide for the 2008-2009 ANES panel study*. Palo Alto, CA and Ann Arbor, MI: Stanford University and the University of Michigan.

De Leeuw, E.D. (1992). *Data quality in mail, telephone and face to face surveys*. Amsterdam, NL: TT-Publikaties.

Draisma, S. & Dijkstra, W. (2004). Response latency and (para)linguistic expressions as indicators of response error. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for testing and evaluating survey questions* (pp. 131-147). Hoboken, NJ: Wiley.

Duch, R.M. & Stevenson, R.T. (2008). *The economic vote. How political and economic institutions condition election results*. New York, NY: Cambridge University Press.

Faas, T. & Schoen, H. (2006). Putting a questionnaire on the web is not enough – a comparison of online and offline surveys conducted in the context of the German Federal Election 2002. *Journal of Official Statistics, 22,* 177-190.

Fraley, R.C. (2004). *How to conduct behavioral research over the internet. A beginner's guide to HTML and CGI/Perl.* New York, NY: The Guilford Press.

Galesic, M. & Bosnjak M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73,* 349-360.

GESIS – Leibniz-Institute for the Social Science (2009a). *German Longitudinal Election Study 2009: Langfrist-Online-Tracking, T6: ZA5339, Version 2.0.0, Studienbeschreibung*. Retrieved August 20, 2012, from http://info1.gesis.org/dbksearch19/SDesc2.asp?no=5339&tab=3&ll=10&notabs=&af=&nf=1&search=gles&search2=&db=e

GESIS – Leibniz-Institute for the Social Science (2009b). *German Longitudinal Election Study 2009: Short-term Campaign Panel: ZA5305, Version 3.0.0, Study Description* (ZA5305_sd_v3-0-0.pdf). Retrieved August 20, 2012, from http://info1.gesis.org/dbksearch19/SDesc2.asp?no=5305&tab=3&ll=10&notabs=&af=&nf=1&search=gles&search2=&db=E

Greszki, R.P., Meyer, M. & Schoen, H. (2012). *Speeding in web surveys - a serious threat to data quality?* Paper presented at the annual European Political Science Association Conference EPSA. Berlin, DE.

Groves, R.M. (1991). *Survey errors and survey costs.* New York, NY: Wiley.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. & Tourangeau, R. (2009). *Survey methodology*. Hoboken, NJ: Wiley.

Hansen, S.B. (1975). Participation, political structure, and concurrence. *The American Political Science Review, 69,* 1181-1199.

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review, 21,* 360-373.

Heerwegh, D. (2011). Internet survey paradata. In M. Das, P. Ester & L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet* (pp. 325-348). New York, NY: Routledge.

Johnson, M. (2004). Timepieces: Components of survey question response latencies. *Political Psychology, 25*, 679-702.

Kaczmirek, L. (2009). *Human-survey interaction. Usability and nonresponse in online surveys.* Cologne, DE: Herbert von Halem Verlag.

Kahn, R.L. & Cannell, C.F. (1957). *The dynamics of interviewing*. New York, NY: Wiley.

Kaminska, O., Goeminne, B. & Swyngedouw, M. (2006). *Satisficing in early versus late responses to a mail survey*. Leuven, NL: Katholieke Universiteit Leuven.

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

Knapton, K. & Garlick, R. (2007). *Catch me if you can*. Retrieved August 28, 2013, from http://www.quirks.com/articles/2007/20071107.aspx?searchID=765736159&sort=5&pg=1

Krosnick, J.A. & Alwin, D.F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*, 201-219.

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., Narayan, S. & Smith, W.R. (1996). Satisficing in surveys: Initial evidence. In M.T. Braverman & J.K. Slater (Eds.), *Advances in Survey Research (*pp. 29-44). San Francisco, CA: Jossey-Bass.

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. Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly, 66*, 371-403.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly, 72*, 914-934.

Meade, A.W. & Craig S.B. (2011). *Identifying careless responses in survey data*. Paper presented at the 26th Annual Meeting of the Society for Industrial and Organizational Psychology. Chicago, IL.

Meyer, M. & Schoen, H. (in press). Response latencies and attitude-behavior consistency in a direct democratic setting: Evidence from a subnational referendum in Germany. *Political Psychology*.

Miller, J. (2006). *Research reveals alarming incidence of "undesirable" online panelists*. Retrieved June 28, 2013, from http://www.burke.com/Library/Articles/Jeff%20Miller%20RCR%20PDF.pdf

Miller, J. & Baker-Prewitt, J. (2009). *Beyond 'trapping' the undesirable panelist: The use of red herrings to reduce satisficing*. Retrieved June 28, 2013, from http://www.burke.com/Library/Conference/Beyond%20Trapping%20the%20Undesirable%20Panelist_FINAL.pdf

Narayan, S. & Krosnick, J.A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly, 60*, 58-88.

Puleston, J. & Sleep D. (2008). *A look at the impact of boredom on the respondent experience.* Retrieved June 28, 2013, from http://www.quirks.com/articles/2008/20081106.aspx?searchID=765759022&sort=5&pg=1

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510-532.

Rattinger, H., Rossteutscher, S., Schmitt-Beck, R. & Wessels, B. (2009). *German Longitudinal Election Study – Long-term Online Tracking, T6 (GLES 2009)*. Cologne, DE: GESIS Leibnitz-Institute for the Social Science.

Respondi AG (2009a). *Our answers to the 26 Esomar Questions for Determining the Quality of Online Samples and Online Panels*. Retrieved August 20, 2012, from http://www.respondi.com.

Respondi AG (2009b). *Panel book*. Retrieved August 20, 2012, from http://www.respondi.com.

Rossmann, J. (2010). *Data quality in web surveys of the German Longitudinal Election Study 2009.* Paper presented at the 3rd ECPR Graduate Conference, Dublin, IE.

Schaeffer, N.C., Dykema, J. & Maynard, D.W. (2010). Interviewers and interviewing. In P.T. Marsden & J.D. Wright (Eds.), *Handbook of survey research* (pp. 437-470). Bingley, UK: Emerald.

Schoen, H. (2004). Online-Umfragen - schnell, billig, aber auch valide? Ein Vergleich zweier Internetbefragungen mit persönlichen Interviews zur Bundestagswahl 2002. *ZA-Information, 54,* 27-52.

Schuman, H. & Presser, S. (1996). *Question and answers in attitude surveys.* Thousand Oaks, CA: Sage.

Silver, B.D., Anderson, B.A. & Abramson, P.R. (1986). Who overreports voting? *The American Political Science Review, 80*, 613-624.

Smith, R. & Brown, H.H. (2005). *Assessing the quality of data from online panels: Moving forward with confidence.* Retrieved November 14, 2012, from http://www.harrisinteractive.com/servicebureau/pubs/HI_Quality_of_Data_White_Paper.pdf

Topoel, V., Das, M. & Van Soest, A. (2008). Effects of design in web surveys. Comparing trained and fresh respondents. *Public Opinion Quarterly, 72,* 985-1007.

Tourangeau, R. (1984). Cognitive science and survey methods. In T.B. Jabine, M.L. Straf, J.M. Tanur & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73-100). Washington, DC: National Academy Press.

Tourangeau, R. (1987). Attitude measurement: A cognitive perspective. In H.J. Hippler, N. Schwarz & S. Sudman (Eds.), *Social information processing and survey methodology* (pp.149-162). New York, NY: Springer.

Tourangeau, R. & Rasinski, K. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103,* 299-314.

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

Tourangeau, R., Couper, M.P. & Conrad, F. (2004). Spacing, position, and order. Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly, 68,* 368-393.

Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology, 55,* 775-801.

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,* 51-68.

Tables and figures

Table 1: Response time as a function of satisficing behavior (OLS)

	Constant	B	β (r)	Adj. R^2	N
No answer	1380.4***	-22.3***	-0.21	0.05	1,153
	(16.6)	(3.0)			
Don't know	235.4***	-8.3***	-0.20	0.04	1,153
	(4.0)	(1.2)			
Middle category	291.3***	-2.0	-0.05	0.002	1,153
	(8.2)	(1.2)			
Straightlining	372.9***	-26.8***	-0.33	0.11	1,153
	(5.1)	(2.3)			

Note: Cell entries are b-coefficients and standardized β-coefficients; standard errors in parentheses; Significance levels: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

Table 2: Number of excluded respondents and percentage of speeding flags across all survey pages (for three median-based criteria)

	GLES	ANES	Difference in percent points
50 percent faster	92 (8.0 %)	66 (4.1 %)	3.9***
40 percent faster	169 (14.7 %)	187 (11.6 %)	3.1*
30 percent faster	250 (21.7 %)	335 (20.8 %)	0.9
N	1144 - 1153	1607 - 1609	

Note: Significance levels: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. In order to avoid biases resulting from varying numbers of observations, we included only survey pages which have been seen by 99 percent or more respondents.

Table 3: Proportion of speeding flags by age group for the overall and the group-wise speeding measure (total and for different age groups)

	GLES		ANES	
	raw (%)	corrected (%)	raw (%)	corrected (%)
50 percent faster (total)	8.0	7.2	4.1	4.2
18 - 29 years	12.1	8.6	10.9	3.6
30 - 39 years	11.3	7.8	8.2	1.9
40 - 49 years	7.1	6.4	4.2	3.1
50 - 59 years	5.2	7.4	2.5	3.2
60 years +	2.3	5.7	0.9	3.9
40 percent faster (total)	14.7	13.0	11.6	10.3
18 - 29 years	21.2	14.4	24.8	10.9
30 - 39 years	20.6	14.2	21.8	8.2
40 - 49 years	14.0	11.7	12.5	10.3
50 - 59 years	6.8	13.2	8.9	10.4
60 years +	4.4	11.4	3.3	11.1
30 percent faster (total)	21.7	21.2	20.8	19.5
18 - 29 years	30.2	22.8	37.2	19.6
30 - 39 years	29.2	22.4	35.3	17.1
40 - 49 years	21.7	19.4	23.8	19.0
50 - 59 years	14.9	21.5	18.1	19.8
60 years +	7.0	19.8	8.1	20.9

Note: 'Raw' entries indicate the proportion of flags when speeding is measured by comparing an individual's response time to the response times in the whole sample. 'Corrected' entries indicate the proportion of flags when speeding is measured by comparing an individual's response time to the response times in the respective age group.

Table 4: Proportion of speeding flags by education group for the overall and the group-wise speeding measure (total and for different education groups)

	GLES		ANES	
	raw (%)	corrected (%)	raw (%)	corrected (%)
50 percent faster (total)	7.9	8.0	4.1	4.2
education low	6.6	8.6	3.3	7.3
education medium	7.7	8.1	3.6	4.1
education high	9.9	7.1	4.9	2.8
40 percent faster (total)	14.6	14.2	11.6	11.5
education low	11.5	14.9	8.5	15.5
education medium	14.0	14.7	10.6	11.3
education high	19.0	12.6	14.1	9.8
30 percent faster (total)	21.6	21.5	20.8	20.5
education low	16.8	22.1	14.8	24.0
education medium	20.7	21.5	19.2	20.5
education high	28.3	20.8	25.0	18.8

Note: 'Raw' entries indicate the proportion of flags when speeding is measured by comparing an individual's response time to the response times in the whole sample. 'Corrected' entries indicate the proportion of flags when speeding is measured by comparing an individual's response time to the response times in the respective education group.

Table 5: Response time as a function of recruitment and respondent behavior (GLES; OLS)

	b	β
Constant	0.48***	
	(0.03)	
Number of survey completions (last 4 weeks)	-0.004	-0.04
	(0.003)	
Panel membership (weeks)	0.0004***	0.11
	(0.00)	
Self-recruitment	0.003	0.01
	(0.03)	
Recruitment through extern link	-0.083**	-0.14
	(0.03)	
Adj. R^2	0.03	
N	1140	

Note: Cell entries are b-coefficients and standardized β-coefficients; standard errors in parentheses; Significance levels: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. Reference group for recruitment: friend advertising

Table 6: Evaluations of Merkel's handling of the economy (marginal distributions across different thresholds)

	Without exclusion	50% faster than median	40% faster than median	30 % faster than median
	%	%	%	%
1 not at all suitable	18.1	17.8	17.4	17.3
2	16.0	17.1	17.4	17.3
3	26.3	24.4	25.0	25.6
4	25.7	26.9	26.8	26.6
5 very suitable	13.9	13.9	13.4	13.2
N valid	1091	972	891	836
N speeders of valid (%)	-	119 (11%)	200 (18%)	255 (23%)
N speeders total	-	134	219	278

Note: None of the differences between the proportions is statistically significant ($p > 0.05$). Question text: 'Angela Merkel has suitable concepts for stimulating the economy'.

Table 7: Evaluations of Bush's handling of the economy (marginal distributions across different thresholds)

	Without exclusion	50% faster than median	40% faster than median	30 % faster than median
	%	%	%	%
extremely disapprove	31.2	31.5	30.5	29.9
moderately disapprove	20.7	20.8	20.9	21.6
slightly disapprove	4.8	4.8	4.9	5.0
neither nor	24.9	24.3	24.8	24.2
slightly approve	6.1	6.2	6.3	6.7
moderately approve	8.1	8.1	8.3	8.6
extremely approve	4.2	4.2	4.3	4.0
N valid	1610	1569	1452	1321
N speeders of valid (%)	-	41(3%)	158 (10%)	289(18%)
N speeders total	-	43	160	292

Note: None of the differences between the proportions is statistically significant ($p > 0.05$). Question text: 'Do you approve, disapprove, or neither approve nor disapprove of the way George W. Bush is handling the economy?'

Table 8: Determinants of evaluations of Merkel's handling of the economy (across different speeding thresholds; OLS)

	baseline model	50 % faster than median		40% faster than median		30% faster than median	
	B (s.e.)	B (s.e.)	Diff. to baseline	B (s.e.)	Diff. to baseline	B (s.e.)	Diff. to baseline
Constant	0.15*** (0.02)	0.13*** (0.03)	0.02	0.14*** (0.03)	0.01	0.14*** (0.03)	0.01
Evaluation CDU	0.66*** (0.03)	0.65*** (0.03)	0.01	0.61*** (0.03)	0.05	0.60*** (0.04)	0.06
Evaluation SPD	-0.01 (0.03)	0.02 (0.03)	-0.03	0.01 (0.03)	-0.02	0.00 (0.03)	-0.01
Ideology (high=conserv.)	0.03 (0.04)	0.06 (0.04)	-0.03	0.09* (0.04)	-0.06	0.10* (0.05)	-0.07
Economy better than 1 year ago	0.14*** (0.03)	0.16*** (0.04)	-0.02	0.18*** (0.04)	-0.04	0.18*** (0.05)	-0.04
Adjusted R^2	0.49	0.50		0.49		0.47	
N	926	806		693		605	

Note: Cell entries are b-coefficients; Standard errors in parentheses; Significance levels: *: $p < 0.05$; **: $p < 0.01$; ***: $p< 0.001$. Difference tests for regression coefficients for different samples were applied following Cohen et al. (2003).

Table 9: Determinants of evaluations of Bush's handling of the economy across different speeding thresholds (OLS)

	baseline model	50 % faster than median		40% faster than median		30% faster than median	
	B (s.e.)	B (s.e.)	Diff. to baseline	B (s.e.)	Diff. to baseline	B (s.e.)	Diff. to baseline
Constant	0.14*** (0.03)	0.14*** (0.03)	0.00	0.16*** (0.03)	-0.02	0.16*** (0.04)	-0.02
Evaluation Democrats	-0.21*** (0.02)	-0.21*** (0.03)	0.00	-0.24*** (0.03)	-0.03	-0.24*** (0.03)	-0.03
Evaluation Republicans	0.28*** (0.03)	0.31*** (0.03)	-0.03	0.29*** (0.03)	-0.01	0.27 (0.03)	0.01
Ideology (high=conserv.)	0.12*** (0.03)	0.10*** (0.03)	0.02	0.11*** (0.03)	0.01	0.12** (0.04)	0.00
Economy better than 1 year ago	0.43*** (0.03)	0.41*** (0.03)	0.02	0.39*** (0.04)	0.04	0.40*** (0.04)	0.03
Adjusted R^2	0.46	0.47		0.45		0.44	
N	1598	1336		1067		862	

Note: Cell entries are b-coefficients; Standard errors in parentheses; Significance levels: *: $p < 0.05$; **: $p < 0.01$; ***: $p< 0.001$. Difference tests for regression coefficients for different samples were applied following Cohen et al. (2003).

Table 10: Marginal distribution of turnout intention in GLES and ANES

	without exclusion	50% faster than median	40% faster than median	30% faster than median
GLES	%	%	%	%
yes	91.5	92.2	92.6	92.8
no	8.5	7.8	7.4	7.2
N valid	1123	912	850	737
N speeders of valid (%)	-	211 (19%)	273 (24%)	386 (34%)
N speeders total	-	221	283	399
ANES	%	%	%	%
yes	92.7	93.0	93.1	92.9
no	7.3	7.0	6.9	7.1
N valid	1620	1588	1488	1362
N speeders of valid (%)	-	32 (2%)	132 (8%)	258 (16%)
N speeders total	-	32	132	258

Note: None of the differences between the proportions is statistically significant ($p > 0.05$).

Table 11: Determinants of turnout intention (ANES) with different speeding thresholds (logistic regression)

	Baseline model	50% faster than median		40% faster than median		30% faster than median	
	B	B	Diff. to baseline	B	Diff. to baseline	B	Diff. to baseline
Constant	-1.48*** (0.30)	-1.43*** (0.32)	-0.05	-1.42*** (0.37)	-0.06	-1.55*** (0.43)	0.07
Political Interest	4.84*** (0.59)	4.60*** (0.62)	0.24	4.90*** (0.73)	-0.06	4.70*** (0.86)	0.14
Efficacy external	0.93 (0.62)	0.94 (0.66)	-0.01	0.08 (0.73)	0.85	-0.30 (0.84)	1.23
Efficacy internal	1.73** (0.55)	1.59** (0.57)	0.14	2.06** (0.66)	-0.33	2.19** (0.74)	-0.46
Talking Politics	1.24* (0.59)	1.52* (0.63)	-0.28	1.50* (0.70)	-0.26	2.39** (0.88)	-1.15
PID yes	0.65** (0.23)	0.74** (0.24)	-0.09	0.72** (0.28)	-0.07	0.74* (0.31)	-0.09
Nagelkerke R^2	0.36	0.34		0.33		0.35	
N	1603	1472		1196		886	

Note: Cell entries are logit coefficients; Standard errors in parentheses; Significance levels: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. Difference tests for logit coefficients for different samples were applied following Cohen et al. (2003).

Table 12: Determinants of turnout intention (GLES) with different speeding thresholds (logistic regression)

	Baseline model	50% faster than median		40% faster than median		30% faster than median	
	B	B	Diff. to baseline	B	Diff. to baseline	B	Diff. to baseline
Constant	-2.50*** (0.39)	-2.37*** (0.51)	0.13	-2.24*** (0.59)	-0.26	-2.59*** (0.72)	0.09
Political Interest	2.85*** (0.67)	2.71** (0.87)	0.14	2.96** (0.99)	-0.11	3.46*** (1.16)	-0.61
Campaign Interest	1.38* (0.66)	1.14 (0.80)	0.24	0.81 (0.89)	0.57	0.44 (1.01)	0.94
Citizen duty	3.15*** (0.47)	3.24*** (0.54)	-0.09	3.11*** (0.57)	0.04	3.17*** (0.66)	-0.02
Satisf. w. Democracy	1.56** (0.58)	1.76* (0.75)	-0.20	1.72* (0.79)	0.16	3.09** (0.96)	-1.53
PID strength	0.54 (0.37)	0.69 (0.47)	-0.15	0.52 (0.59)	0.02	-0.03 (0.60)	0.57
Nagelkerke R^2	0.42	0.39		0.34		0.36	
N	1077	781		633		491	

Note: Cell entries are logit coefficients; Standard errors in parentheses; Significance levels: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. Difference tests for logit coefficients for different samples were applied following Cohen et al. (2003).

Figure 1: Distribution of response times for ‘political interest’ and ‘talking about politics’ in ANES and GLES

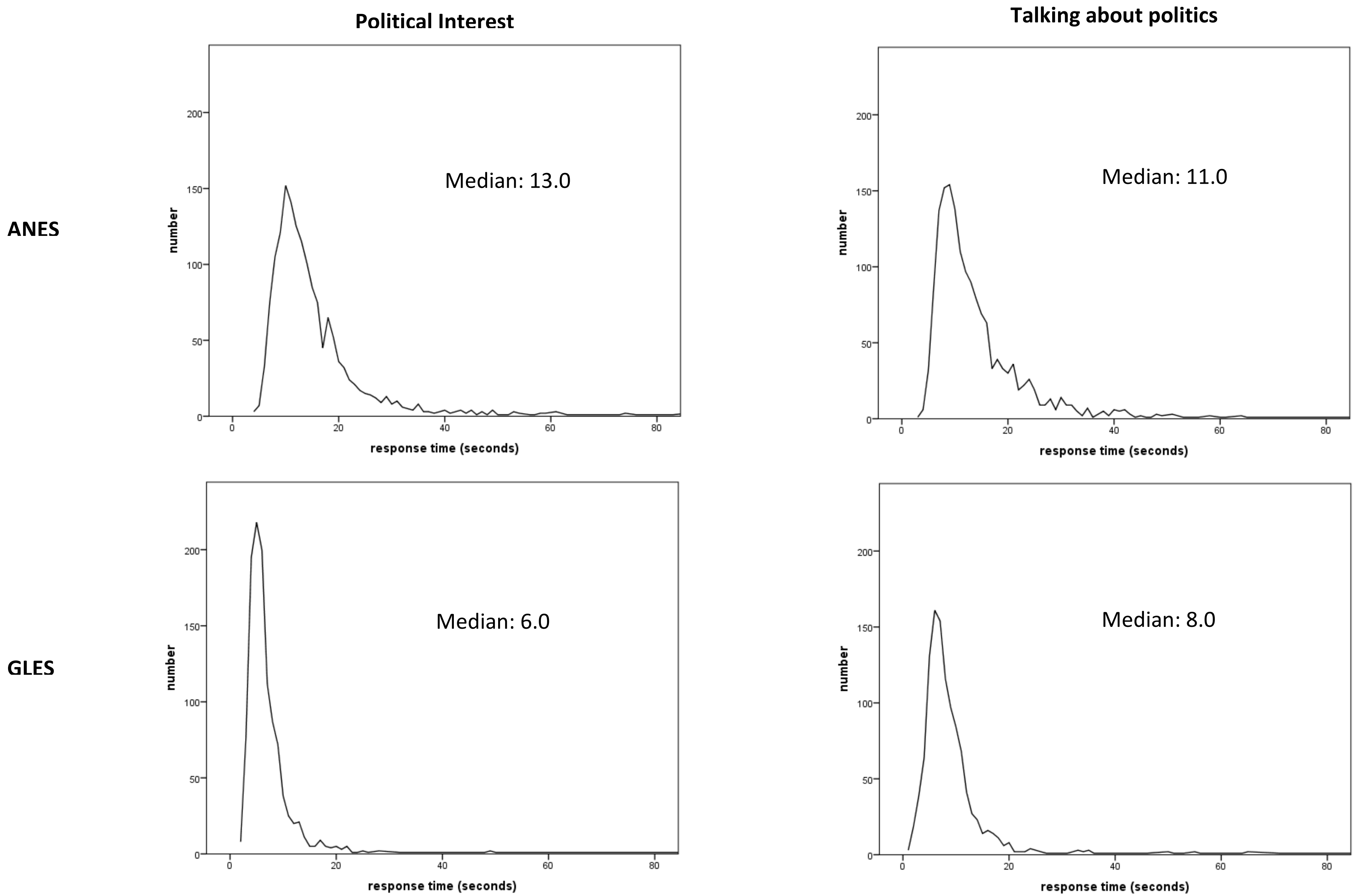
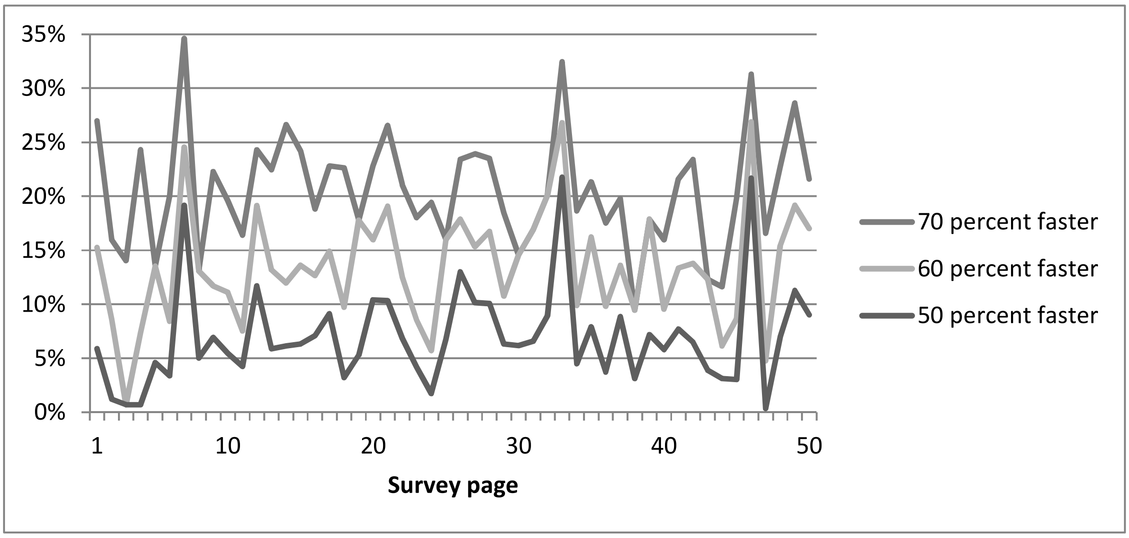
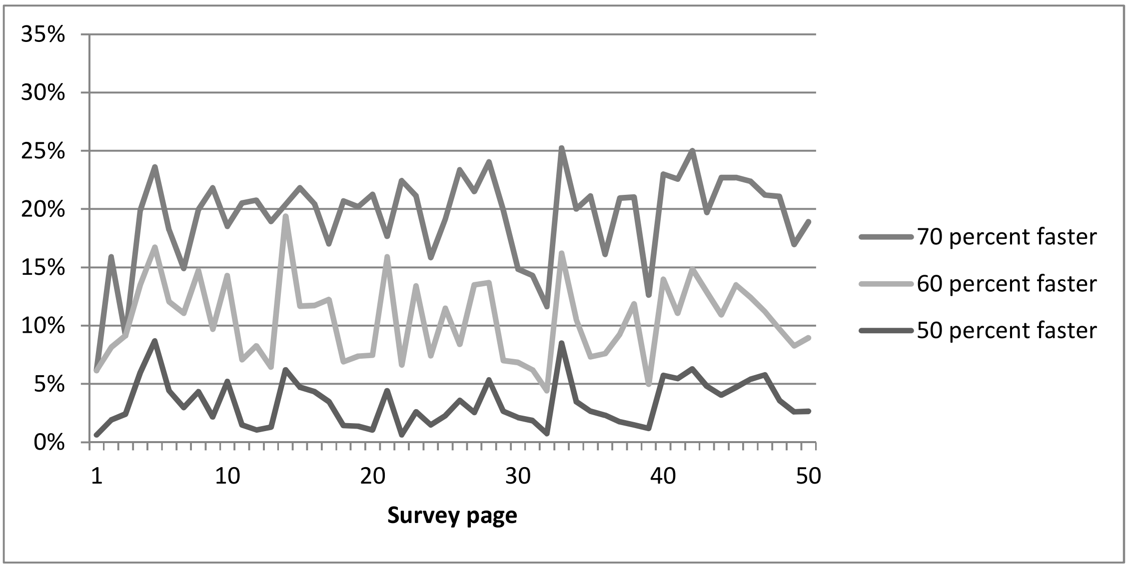


Figure 2: Percentage of flagged speeders for the first 50 survey pages (GLES)



Note: In order to avoid biases resulting from varying numbers of observations, we included only survey pages which have been seen by 99 percent or more respondents.

Figure 3: Percentage of flagged speeders for the first 50 survey pages (ANES)



Note: In order to avoid biases resulting from varying numbers of observations, we included only survey pages which have been seen by 99 percent or more respondents.