Journal of Survey Statistics and Methodology (2024) **12**, 36–59

PANEL CONDITIONING IN A PROBABILITY-BASED LONGITUDINAL STUDY: A COMPARISON OF RESPONDENTS WITH DIFFERENT LEVELS OF SURVEY EXPERIENCE

FABIENNE KRAEMER*
HENNING SILBER
BELLA STRUMINSKAYA
MATTHIAS SAND
MICHAEL BOSNJAK
JOANNA KOßMANN
BERND WEIß

Learning effects due to repeated interviewing, also known as panel conditioning, are a major threat to response quality in later waves of a panel study. To date, research has not provided a clear picture regarding the circumstances, mechanisms, and dimensions of potential panel conditioning effects. In particular, the effects of conditioning frequency, that is, different levels of experience within a panel, on response quality are underexplored. Against this background, we investigated the effects of panel conditioning by using data from the GESIS Panel, a German mixed-mode probability-based panel study. Using two refreshment samples, we compared three panel cohorts with differing levels of experience on several response quality indicators related to the mechanisms of reflection, satisficing, and social desirability. Overall, we find evidence for both negative (i.e., disadvantageous for response quality) and positive (i.e., advantageous for response quality) panel conditioning. Highly experienced respondents were more likely to satisfice by speeding through the questionnaire. They also had a higher probability of refusing to answer sensitive questions than less experienced panel members. However, more experienced respondents were also more likely to optimize the response process by needing less time compared to panelists with lower experience levels (when controlling for speeding). In contrast, we did not find significant differences with respect to the number of "don't know" responses, nondifferentiation, the selection of first response categories and mid-responses, and the number of nontriggered filter questions. Of the observed differences, speeding showed the highest magnitude with an

https://doi.org/10.1093/jssam/smad004 Advance access publication 13 April 2023

average increase of 6.0 percentage points for highly experienced panel members compared to low experienced panelists.

KEYWORDS: Data quality; Longitudinal surveys; Panel conditioning; Probability-based panel study.

Statement of Significance This study investigates the susceptibility of panel data from a probability-based mixed-mode survey to panel conditioning effects, a potential measurement error induced by repeated interviewing. We explore the magnitude of bias in the data due to panel conditioning by examining the underlying mechanisms of panel conditioning (satisficing, social desirability, reflection) and its consequences on different aspects of response quality. The research compares respondents with three different levels of experience (low, medium, high) within the panel study regarding eight indicators of response quality. The results show the occurrence of positive and negative effects of panel experience on response quality. In sum, our study shows measurement differences between different levels of panel experience, which need to be considered when conducting panel data analyses.

Fabienne Kraemer is a Doctoral Researcher at the Department of Survey Design and Methodology, GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany. Henning Silber is a Senior Researcher at the Department of Survey Design and Methodology, GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany. Bella Struminskaya is an Assistant Professor at the Department of Methodology and Statistics, Utrecht University, Utrecht, Netherlands. Matthias Sand is a Senior Researcher at the Department of Survey Design and Methodology, GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany. Michael Bosnjak is a Full Professor at Trier University, Trier, Germany. Joanna Koßmann is a Research Assistant at ZPID—Leibniz Institute for Psychology, Trier, Germany, and Project Staff in the Section for Teacher Education and Research at Trier University, Trier, Germany. Bernd Weiß is a Senior Researcher at the Department of Survey Design and Methodology, GESIS—Leibniz Institute for the Social Sciences, Mannheim, Germany.

The GESIS Panel data are distributed by GESIS—Leibniz Institute for the Social Sciences and can be accessed by completing and signing a Data Use Agreement. For detailed information on how to access the data, see https://www.gesis.org/en/gesis-panel/data.

This study design and analysis were not preregistered.

The Stata do-file written for the analyses in this article and the R-file written for the calculation of calibration weights are available at: https://osf.io/d7b5w/.

This work was supported by the "German Research Foundation" (DFG) (Grant Number: 418316085).

This article uses data from the GESIS Panel, doi:10.4232/1.13785. The GESIS Panel is carried out by GESIS—Leibniz Institute for the Social Sciences as an open access panel infrastructure that enables the social science community to collect own survey data. As a part of GESIS—Leibniz Institute for the Social Sciences, the GESIS Panel is funded by the federal and state governments.

Address correspondence to Fabienne Kraemer, Survey Design and Methodology, GESIS—Leibniz Institute for the Social Sciences, B6 4-5, 68159 Mannheim, Germany; E-mail: fabienne.kraemer@gesis.org.

1. INTRODUCTION

Longitudinal studies are highly valued by researchers across disciplines due to the multitude of advantages they provide compared to cross-sectional studies. With longitudinal data, researchers can analyze intra-individual stability or change over time and are able to approach causal relationships (Lynn 2009; Andreß et al. 2013). However, longitudinal data also entail specific problems and challenges. When analyzing longitudinal data, researchers might implicitly assume that measuring respondents' attitudes and behavior in previous interviews does not alter responses in subsequent surveys. However, when working with panel data, researchers should consider the possibility that respondents' previous interviews influence their responses in subsequent waves of the study—a phenomenon referred to as *panel conditioning* (Kalton et al. 1989).

Various studies have demonstrated that previous survey participation influences respondents' answers in subsequent interviews (e.g., Das et al. 2011; Warren and Halpern-Manners 2012; Kroh et al. 2016; Halpern-Manners et al. 2017), pointing out the relevance of panel conditioning as a specific measurement error in longitudinal studies. Panel conditioning is caused by learning effects that take place over the course of a panel study causing changes in respondent's actual behavior, attitudes, and knowledge or solely affect the way in which answers are reported (Struminskaya 2020). Those artificial changes in actual attitudes and behavior represent a major threat to the validity of panel data since such changes are often indistinguishable from real changes in respondents' attitudes and behaviors over time and might lead to substantial misconceptions when investigating developments of social patterns across panel waves.

Following the concern of jeopardized validity of scientific conclusions based on panel data, researchers have investigated the presence and magnitude of panel conditioning effects in large-scale and widely-used longitudinal studies such as the US General Social Survey (Halpern-Manners et al. 2017; Sun et al. 2018), the LISS Panel (Toepoel et al. 2009; Schonlau and Toepoel 2015; Crossley et al. 2017), the Current Population Survey (Halpern-Manners and Warren 2012), and the German Socio-Economic Panel (Kroh et al. 2016). Despite numerous studies that have investigated panel conditioning, prior research provides mixed evidence of its existence and consequences for the quality of survey responses. Whereas some studies show evidence of negative learning effects (e.g., increased straightlining, misreporting to questions that trigger follow-ups) (Halpern-Manners and Warren 2012; Schonlau and Toepoel 2015), which result in lower response quality, others report positive effects of prior survey participation on response quality in later waves (e.g., increased reliability of survey responses, higher knowledge levels) (Waterton and Livesley 1989; Sturgis et al. 2009; Toepoel et al. 2009; Kroh et al. 2016). Yet, some studies do not find any evidence of changes in response quality due to repeated interviewing (Sun et al. 2018). Thus, it remains unclear whether respondents' experience with a survey is advantageous or disadvantageous for the overall quality of panel data, and if so, under which circumstances.

In this article, we examine the following research questions: (1) Does positive and/or negative panel conditioning exist within a probability-based mixed-mode panel of the general population? (2) What is the magnitude of panel conditioning effects and which indicators of response quality are affected most by panel conditioning? (3) Which theoretical mechanisms are responsible for the effects of panel experience on response quality?

Our study uses data from the GESIS Panel, a German probability-based mixed-mode panel survey (Bosnjak et al. 2018), and contributes to the literature that uses large-scale longitudinal surveys to study panel conditioning. To examine the effects of panel conditioning on survey response quality, we assess a range of response quality indicators that are derived from theoretical mechanisms of prior survey participation: (1) reflection of prior survey content that leads to changes in respondents' attitudes, behavior, and knowledge, (2) acquired strategies to reduce survey burden, and (3) a growing trust to reveal socially undesirable information. This theory-driven analytical approach provides additional evidence regarding which specific indicators of response quality worsen or improve by repeated interviewing and which mechanisms might account for the positive or negative consequences. We use the GESIS Panel's specific study design, which incorporates two refreshment samples along with initial panelists (the first cohort) to analyze the consequences of repeated interviewing on data quality. Thus, we extend previous research by examining different levels of experience within a panel in greater detail using a comprehensive set of indicators related to positive and negative panel conditioning.

2. MECHANISMS OF PANEL CONDITIONING AND HYPOTHESES

The majority of the literature on panel conditioning effects can be grouped by distinguishing among the mechanisms that cause different changes in survey responses due to prior survey participation: reflection, survey satisficing, and social desirability bias (Struminskaya and Bosnjak 2021). The first mechanism—*reflection*—describes changes in survey responses which are based on respondents' increased awareness of the survey topic, possibly leading to more deliberation and an increased information search (Sturgis et al. 2009). The second mechanism—*survey satisficing* (Krosnick 1991)—describes a change that results from increasingly applying response strategies which aim to reduce survey burden by taking shortcuts during the response process. The third mechanism—*social desirability bias*—comprises changes in survey responses which are based on the need to provide answers that are in line with presumed social norms (Struminskaya 2016; Bergmann and Barth 2018).

Satisficing and socially desirable responding can affect response quality in both longitudinal surveys and one-time surveys (see, e.g., Duan et al. 2007). The adverse consequences of satisficing and socially desirable responding on response quality threaten the validity of longitudinal data because repeated exposure to a survey can cause changes in those unwanted effects across waves. The mechanism of reflection, however, is especially relevant for longitudinal studies because the repeated exposure to the survey content is a precondition for reflection processes to take place.

Each respondent undergoes four stages of the cognitive response generation process when presented with a survey question (Tourangeau et al. 2000): (1) comprehension of a survey question; (2) retrieval of relevant information that is needed to make a founded judgment; (3) the formation of this judgment; and (4) the provision of an answer to the respective survey question. The three underlying mechanisms of panel conditioning affect response formation at different stages of the response process and thus can affect the quality of the final response positively or negatively (see, e.g., Struminskaya and Bosnjak 2021).

2.1 Reflection

2.1.1 Process learning

During the comprehension stage, complexity and ambiguity of the survey questions are key factors that affect the potential occurrence of measurement error (Kroh et al. 2016). Previous participation in the survey can be beneficial for question comprehension since respondents' familiarity with the survey topic(s) and questions can decrease a question's complexity and ambiguity (Waterton and Livesley 1989). Specifically, by participating in previous interviews respondents may gain general experience with answering different types of survey questions as they become familiar with the overall response procedure, different survey instruments, and how to use different response scales (Basso et al. 2001; Nancarrow and Cartwright 2007; Kroh et al. 2016). Accordingly, one can assume that experienced respondents have an improved comprehension of both the general survey response process and specific questions due to learning effects from previous interviews. These learning effects, however, are more likely if respondents fill in previous questionnaires carefully and put effort into answering the questions.

Considering respondents' increased familiarity with the survey procedure, possibly improving their overall comprehension of administered question types, we assume that experienced respondents have optimized the response process, independent of the specific question content.

> *Hypothesis 1: Experienced respondents undergo the response process more efficiently than less experienced respondents.*

2.1.2 Content learning

The second stage of the cognitive response process is characterized by the retrieval of relevant information from memory. Again, the reflection process as one of the main mechanisms of panel conditioning might be relevant at this stage of the response process. Specifically, previous survey participation might raise respondents' awareness of topics that were addressed during a survey. Respondents may reflect on survey topics and may discuss these topics with others or seek out further information on these topics, by paying closer attention to the news, for example (Sturgis et al. 2009). As a result, respondents' attitudes can become more stable and reliable over the course of a study (Sturgis et al. 2009; Kroh et al. 2016) and their knowledge on surveyed topics may increase (Toepoel et al. 2009). Thus, respondents might provide fewer "don't know" answers (Binswanger et al. 2013). We expect these learning effects to be more likely for respondents who carefully answered the questions of previous surveys and assume that their question-specific knowledge increases with more panel experience.

Hypothesis 2: Experienced respondents show higher levels of question-specific knowledge than less experienced respondents.

2.2 Survey Satisficing

In the third stage of the cognitive response process, respondents form judgments based on their comprehension of the question and the information they considered relevant and retrieved from memory. Some respondents might resort to the response strategy of survey satisficing if their main interest does not lie in providing optimal responses, but instead in reducing the burden of answering demanding survey questions. According to Krosnick (1991), three factors affect the susceptibility to satisfice: (1) respondent's cognitive ability, (2) respondent's motivation, and (3) the task difficulty. Respondents' motivation—a key influence on response quality—might decrease across panel waves as respondents experience the tedious process of repeatedly answering identical survey questions (also known as panel fatigue; Lundmark and Gilljam 2013).

When aiming to produce a sufficient answer to a question, respondents can use several response strategies to shorten the survey response process: nondifferentiation of response options in matrix questions (e.g., Roßmann et al. 2018), selecting "don't know" instead of a substantive response option (e.g., Silber et al. 2021), selecting the first acceptable response option of a question (e.g., Holbrook et al. 2007), selecting mid-responses on rating scales (e.g., Kaminska et al. 2010), and speeding through the questionnaire (e.g., Zhang and Conrad 2014). Previous research investigating satisficing in a longitudinal context has shown that experienced respondents are indeed more likely to nondifferentiate their answers on matrix questions (Schonlau and Toepoel 2015) and speed through the questionnaire without taking the time to properly process the survey tasks and form substantive and accurate answers (Zhang and Conrad 2014).

An additional form of satisficing response behavior is motivated misreporting—a strategic underreporting of certain behaviors to prevent follow-up questions to reduce survey burden (Eckman et al. 2014; Eckman and Kreuter 2018). To shorten the survey, respondents with knowledge about the structure of the survey and the follow-up questions deliberately triggered fewer filter or looping questions (i.e., questions that are asked multiple times for different options) (Bach and Eckman 2018; Daikeler et al. 2020; Kreuter et al. 2020).

Within the context of repeated interviewing, motivated misreporting among respondents with an advanced knowledge of the questionnaire has been documented for unemployment (Bailar 1975; Halpern-Manners and Warren 2012), party membership (Warren and Halpern-Manners 2012), everyday hygiene product use (Nancarrow and Cartwright 2007), and functional limitations (Mathiowetz and Lair 1994). However, other studies analyzing reports of household purchases and social contacts did not find any evidence of an increase in motivated misreporting among experienced respondents (Silber et al. 2019; Bach and Eckman 2020; Eckman and Bach 2021).

In line with most previous research, we assume similar negative consequences for response quality due to increased satisficing and motivated misreporting with multiple interviews.

Hypothesis 3: Experienced respondents show higher levels of survey satisficing and motivated misreporting than less experienced respondents.

2.3 Social Desirability

At the last stage of the cognitive response process, respondents provide an answer to the survey question. Even after undergoing the three steps of cognitive processing to arrive at a "true" answer, respondents might decide against reporting these opinions or behaviors. For questions on sensitive attitudes or behaviors, some respondents will bring their reported attitudes or behaviors in line with social norms instead of revealing socially undesirable information. This phenomenon is known as social desirability bias (Phillips and Clancy 1972).

Previous research has found inconsistent patterns on the effect of social desirability bias in longitudinal studies or across multiple survey waves. Whereas some studies found no change in socially desirable responding in a panel study (Pevalin 2000), other studies indicate an increase in social desirability bias (Fendrich and Vaughn 1994; Torche et al. 2012; Warren and Halpern-Manners 2012). An explanation for such an increase could be that respondents tend to bring their true behaviors and attitudes in line with societal norms to avoid experiencing cognitive dissonance and a repeated conflict between their true behaviors and attitudes and societal values in future waves (Warren and Halpern-Manners 2012). On the contrary, studies that found evidence for a decrease in social desirability bias over the course of a panel study have argued that respondents' increasing trust in the panel study and the confidentiality of their responses might account for the positive effects of prior participation in the panel on the reporting of sensitive information (Waterton and Livesley 1989; Wooden and Li 2014). Since respondents may grow familiar with the survey content and sponsor and learn that providing sensitive information does not lead to any adverse consequences, the reporting bias due to social desirability might decrease in later waves. Indeed, several studies have shown this positive effect of previous survey participation on response quality with respect to sensitive questions in subsequent waves. For example, respondents answered more truthfully to questions about their body weight (Uhrig 2012), life satisfaction (Van Landeghem 2012), were more likely to report racial prejudice (Waterton and Livesley 1989), and the receipt of unemployment benefits (Yan and Eckman 2012). Based on this evidence, we assume that previous participation within the panel will lead to more honest and less socially desirable responding, positively affecting the quality of the respective responses.

Hypothesis 4: Experienced respondents show less socially desirable responding than less experienced respondents.

3. METHODS

3.1 Data

To test our hypotheses, we use data from the GESIS Panel, a German probability-based mixed-mode access panel of about 5,000 panelists (GESIS 2021). The first cohort of the GESIS Panel ($n = 4{,}938$) was recruited in 2013, based on a random sample drawn from municipality registers targeting German-speaking persons aged 18–70 years who were permanently residing in private households in Germany. In 2016 and 2018, two refreshment samples ($n = 1{,}710$; $n = 1{,}607$) were recruited to counter the effects of panel attrition. In each survey wave, respondents receive an annually administered longitudinal core study on a specific issue and several studies on varying topics (for detailed information, see appendix A in the supplementary data online).

In this article, we use data from two survey waves of the GESIS Panel, the 17th and 29th waves, which have been administered in October–November 2016 and October–November 2018 (GESIS 2021). In the 17th ($n_{\text{1st cohort}} = 3{,}273$; $n_{\text{2nd cohort}} = 1{,}447$) and 29th waves of the GESIS Panel ($n_{\text{1st cohort}} = 2{,}759$; $n_{\text{2nd cohort}} = 1{,}149$; $n_{\text{3rd cohort}} = 1{,}313$), the annually repeated core questionnaire focused on the topics of media usage and work and leisure. As refreshment samples have been incorporated into the existing panel at different stages, we selected those waves comprising respondents of different panel cohorts and with different levels of experience regarding the survey content. The 29th wave of the GESIS Panel includes all three cohorts and thus represents our main data source for the analyses, whereas the 17th survey wave is a suitable data base to analyze the mechanism of social desirability due to the availability of specific items, which were not administered in wave 29.

In the 29th wave, completion rates (AAPOR COMR; AAPOR 2016) for the first, second, and third cohorts were 91.5, 90.7, and 88.2 percent, respectively. Cumulative response rates (CUMR2) varied from 12.8 percent for the first cohort to 10.9 percent for the second and 11.5 percent for the third cohort. Attrition rates were 39.4, 26.1, and 8.0 percent for the three cohorts, respectively (Bretschi et al. 2019). In the 17th wave, completion rates (AAPOR COMR) were 90.7 and 87.5 percent for the first and second cohorts. Cumulative response rates (CUMR2) varied from 15.1 to 13.7 percent and attrition rates were 27.7 and 3.9 percent for the first and second cohorts, respectively (Pötzschke et al. 2017) (for a detailed description and calculation of the outcome rates, see appendix E in the supplementary data online).

3.2 Measures

3.2.1 Survey experience

We compare respondents with different levels of survey experience with respect to the quality of their survey responses. This variation in panelists' experience with the survey content is given by the specific survey design of the GESIS Panel, which encompasses three different panel cohorts that differ in their frequency of participation and thus represent more experienced or less experienced respondents. In the 29th wave of the GESIS Panel—our main data source—the initial cohort has responded to the annual core survey questions on media usage and work and leisure for the fifth time (highly experienced cohort), the second cohort has answered the questions for the third time (medium experienced cohort), whereas the third—newly integrated refreshment sample—responded to the questions for the first time (low experienced cohort, see figure 1). Thus, membership in the different cohorts in the 29th panel wave serves as a measure of respondents' experience within a panel in the following analyses.

To ensure that nonresponse did not confound the different levels of experience across the three cohorts, we only included respondents of the first and second cohorts if they had participated in the previous annual core studies at least once more than the subsequent cohort, respectively (i.e., respondents of the first cohort were only included if they had participated in at least three of the previous core studies; $n = 2{,}736$; respondents of the second cohort were only included if they had participated in at least one of the previous core studies; $n = 1{,}139$). 93.0 percent of the first and 93.2 percent of the second cohort had participated in all previous target studies.

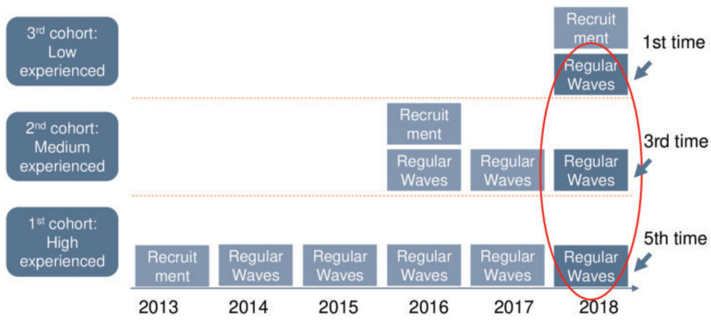

Figure 1. Cohort Comparison Across Survey Experience Levels within the GESIS Panel.

Note.—We used weights to control for sample composition bias due to panel attrition when comparing the three panel cohorts.

Our analyses follow a quasi-experimental design. We conduct pairwise comparisons across the three cohorts of the GESIS Panel using several response quality indicators to assess differences in response behavior of differently experienced respondents (analytic sample: $n = 3{,}756$).

3.2.2 Control of confounders

A major concern for the comparability of the three panel cohorts is confounding effects due to different sample composition. Initial differences in unit nonresponse across the cohorts or nonrandom dropout of respondents from the panel can lead to an imbalance of respondent characteristics among the cohorts that in turn can confound the relationship between experience within the panel and the provided survey responses. We compared the panel cohorts in waves 29 and 17 on selected covariates which were previously found to predict panel attrition (Lugtig 2014; Lugtig et al. 2014). Based on covariates for which we found significant differences between the cohorts, we calculated calibration weights using iterative proportional fitting (raking) (Battaglia et al. 2009) to adjust for the different distribution of the covariates across the cohorts (for detailed information on control of confounders, see appendix B in the supplementary data online).

3.2.3 Reflection

To test our first hypothesis predicting optimized response processes among experienced respondents, we used response time measurements. For the comparison of response time measurements across cohorts, we only used data on those respondents who participated in the online mode since page-level *response latencies* were only collected when panelists completed the questionnaire online (see table 1). We focused on the core study questionnaire, which included 45 items displayed on eight pages administered to every respondent.

Table 1. Overview of Response Quality Indicators and Analytic Strategies

Type of conditioning		Indicator	Example References	Regression model	Corresponding hypothesis	Empirical result
Positive conditioning	Reflection	Response latencies (total response time in sec)	Höhne and Schlosser (2018) and Schnell (1994)	Linear regression	Hypothesis 1	Supported
		Given "don't know" answers (proportion)	Struminskaya (2016)	Binomial GLM with logit link function	Hypothesis 2	Not supported
	Decreased socially desirable responding	Item nonresponse in sensitive questions (proportion)	Tourangeau and Yan (2007)	Binomial GLM with logit link function	Hypothesis 4	Evidence for adverse effect
Negative conditioning	Satisficing	Nondifferentiation in grid questions (proportion)	Roßmann (2018)	Binomial GLM with logit link function	Hypothesis 3	Not supported
		Selection of mid-responses (proportion)	Chyung et al. (2017) and Nadler et al. (2015)	Binomial GLM with logit link function	Hypothesis 3	Not supported
		Selection of first response options (proportion)	Holbrook et al. (2007)	Binomial GLM with logit link function	Hypothesis 3	Not supported
		Speeding (threshold: 300 milliseconds per word × number of words in question; proportion)	Zhang and Conrad (2014)	Binomial GLM with logit link function	Hypothesis 3	Supported
	Motivated misreporting	Number of nontriggered filter questions (proportion)	Daikeler et al. (2020)	Binomial GLM with logit link function	Hypothesis 3	Not supported

NOTE.— GLM, generalized linear model.

To ensure comparability, we excluded the 21 remaining question pages of the core questionnaire for the calculation of the response time measurements as they were not administered to all respondents due to routing. Since outliers can distort the comparison, we excluded extreme response times at the lower and upper ends of the distribution. Extremely short response times are typically associated with satisficing, while overly long response times are indicative of difficulties with question comprehension, interruptions, and multitasking (Matjašič et al. 2018). For the exclusion of outliers at the upper end, we followed the boxplot criterion (Schnell 1994; Höhne and Schlosser 2018). We excluded response times higher than the median response time (q5) plus the interquartile range (IQR) multiplied by 1.5 (i.e., cases with response times longer than $q5 + 1.5 \times \text{IQR}$). We further excluded overly short response times following Zhang and Conrad (2014), who define response latencies under a threshold of 300 milliseconds per word as speeding, arguing that questions cannot be adequately processed and answered within such a short amount of time. We consequently excluded response times lower than 185.7 s, which equals the sum of the page-level speeding thresholds for the eight underlying question pages of the indicator response time calculated following Zhang and Conrad (2014).

To test our second hypothesis on an increase in question-specific knowledge, we used a 17-item battery that focused on the use of different media devices. The response options of the items were "yes," "no," and "don't know." We created a sum score of selected "don't know" options (Struminskaya 2016) across the item battery ranging from 0 to 17 and calculated the proportion of "don't know" answers based on the number of items a respondent had answered (see table 1).

3.2.4 Satisficing response behavior

To test our third hypothesis on an increase in satisficing behavior with higher levels of experience, we used five different indicators of satisficing response behavior: nondifferentiation, the selection of first response options, the selection of middle response options, speeding, and motivated misreporting (see table 1).

Nondifferentiation. To assess a difference in nondifferentiation, we used eight different matrix questions within the core study (see appendix F in the supplementary data online for question wordings and implementation). We operationalize nondifferentiation as providing the same (i.e., nondifferentiated) answers across a set of items (Roßmann et al. 2018). In the first step, we created a count variable that indicates to how many of the eight item batteries in the questionnaire a respondent had provided identical answers to the respective set of items. In the second step, we calculated the proportion of matrix questions answered with nondifferentiation over the total number of matrix questions a respondent answered.

Selection of first response options. To measure selecting the first response option, we used questions with a vertical response scale (nine questions in total) from the core study. We computed the mean proportion of selected first response options based on the number of the respective questions a respondent answered.

Selection of middle response categories. For this indicator, we used all questions from the core study that use Likert rating scales with a midpoint category (three questions in total) (Nadler et al. 2015; Chyung et al. 2017). We calculated the proportion of midpoint responses over the number of questions a respondent answered and compared mean proportions across the three cohorts.

Speeding. To examine differences in respondents' speeding behavior, we restricted our analytical sample to the online respondents for whom page-level response latencies were collected. Following Zhang and Conrad (2014), we defined speeding as those response latencies that lie under a threshold of 300 milliseconds per word. We first calculated page-level indicators of speeding that we then aggregated to composite scores, comprising the counts of question pages on which respondents showed speeding. We compared the mean proportion of pages on which a respondent sped to the number of pages respondents have seen. We included 45 items displayed on eight pages within the core study that were also used for the calculation of the response time indicator and were administered to all respondents.

Motivated misreporting. To examine differences across the cohorts in motivated misreporting, we used three different filter questions within the core study. The analyzed filter questions could trigger up to 21 follow-up questions. We created a count variable indicating how often respondents chose response options to filter questions that did *not* trigger follow-up questions (Daikeler et al. 2020). We compared the mean proportion of motivated misreporting over the number of received filter questions across the three panel cohorts.

3.2.5 Socially desirable responding

To examine our fourth hypothesis predicting a decrease in socially desirable responding with higher levels of experience, we used data from the 17th wave of the GESIS Panel that included a study on attitudes toward foreigners, refugees, and ethnic minorities (e.g., Muslims, Sinti and Roma). In wave 17, only two of the three panel cohorts were surveyed. Specifically, the wave included the first cohort, which answered the questions of the study for the second time (representing medium experienced respondents), and the second cohort, which was just newly integrated into the panel and answered those questions for the first time (representing low experienced respondents). We operationalize socially desirable responding as the extent of item nonresponse in several sensitive attitude questions (23 items in total) (Beatty and Herrmann 2002;

Tourangeau and Yan 2007; Müller and Schmiedeberg 2021). Item nonresponse here refers to either not giving any response to a sensitive question at all or explicitly refusing to answer by choosing an offered evasion category ("I do not want to answer."). We created a count variable indicating how many of the sensitive questions were not answered (see table 1) and compared the mean item nonresponse rate based on the number of received questions across the two cohorts.

3.3 Analysis

We test our four hypotheses using pairwise comparisons across the three panel cohorts for each of the response quality indicators. For this, we estimate different regression models in which the dependent variables reflect the respective outcome measures. To investigate differences in response latencies across the cohorts, we use linear regression analyses. To assess the difference in the proportion of "don't know" answers, we used binomial generalized linear models with a logit link function to predict the probability to provide "don't knows" instead of substantial answers. Lastly, for the analyses of the remaining data quality indicators of satisficing response behavior and socially desirable responding, we also used binomial generalized linear models with a logit link function to model the relationship between the proportional outcome measures and cohort membership (see table 1 for an overview of the analytic methods). To assess the magnitude of bias introduced by prior experience in the panel, we calculated average marginal effects (AMEs) for the binomial generalized linear models. Specifically, AMEs describe the average change in probability for a less experienced cohort to show a respective outcome (i.e., speeding) compared to the highly experienced cohort. Given the multitude of comparisons we aimed to conduct across the panel cohorts, we performed multiple testing corrections on *p*-values following Scheffe's method (Wright 1992). Furthermore, accounting for the calibration weights in our models, we estimated robust standard errors based on the first-order Taylor linearization method (Wolter 2007) for each comparison, which is the standard procedure for the variance estimation of regression coefficients for complex survey data (i.e., surveys not based on a simple random sample).

4. RESULTS

4.1 Reflection

The first hypothesis predicted that experienced respondents take significantly less time to complete the questionnaire compared to their less experienced counterparts as a result of an optimized response process. In line with this hypothesis, results of the linear regression model show that both the first (highly experienced) cohort and the second (medium experienced) cohort show significantly lower response times than the least experienced cohort (tables C1 and C2 in the supplementary data online). As shown in figure 2, compared to respondents of the least experienced cohort (who showed an average total response time of about 5 minutes), highly experienced respondents took almost 24 seconds less to complete the questionnaire ($\beta = -23.9$, $t(2{,}325) = -6.32$, $p < .001$). The difference in response time between respondents from the medium experienced cohort and the low experienced cohort was also significant and similarly strong with medium experienced respondents taking about 20 seconds less than the least experienced cohort ($\beta = -20.3$, $t(2{,}325) = -4.53$, $p < .001$). However, we did not find a significant difference in response time between members of the highly and medium experienced cohort: medium experienced panelists only took about 4 seconds more to complete the questionnaire compared to the highly experienced cohort ($\beta = 3.6$, $t(2{,}325) = 0.95$, $p = .638$). The latter finding might indicate that respondents' familiarity with the survey process and their advanced knowledge about the survey content evolved after participation in a few more survey waves.

To exclude the possibility that increased levels of satisficing were the cause of the lower overall response latencies among more experienced respondents, we conducted a robustness check that included nondifferentiation and both the selection of first response options and mid-responses as control variables into the regression model. The results did not change substantially, still showing significantly higher response times for the low experienced cohort compared to members of both the highly and medium experienced cohort (see tables D1 and D2 in the supplementary data online).

The second hypothesis predicted that experienced respondents provide fewer "don't know" answers compared to less experienced ones. As shown in figure 2, we did not find any significant differences in the proportion of selected "don't know" answers across the three cohorts. Specifically, both medium experienced [$\beta = -0.09$, $p = .912$, OR = 0.92 (95 percent CI: [0.611, 1.372])] and low experienced [$\beta = 0.22$, $p = .508$, OR = 1.24 (95 percent CI: [0.862, 1.793])] respondents were not significantly more likely to provide "don't know" answers than members of the highly experienced cohort (for detailed results see tables C1 and C2 in the supplementary data online). Thus, our results did not support the second hypothesis of increased knowledge levels influencing response behavior.

4.2 Satisficing

The third hypothesis predicted that experienced respondents show more satisficing response behavior for which we analyzed several indicators.

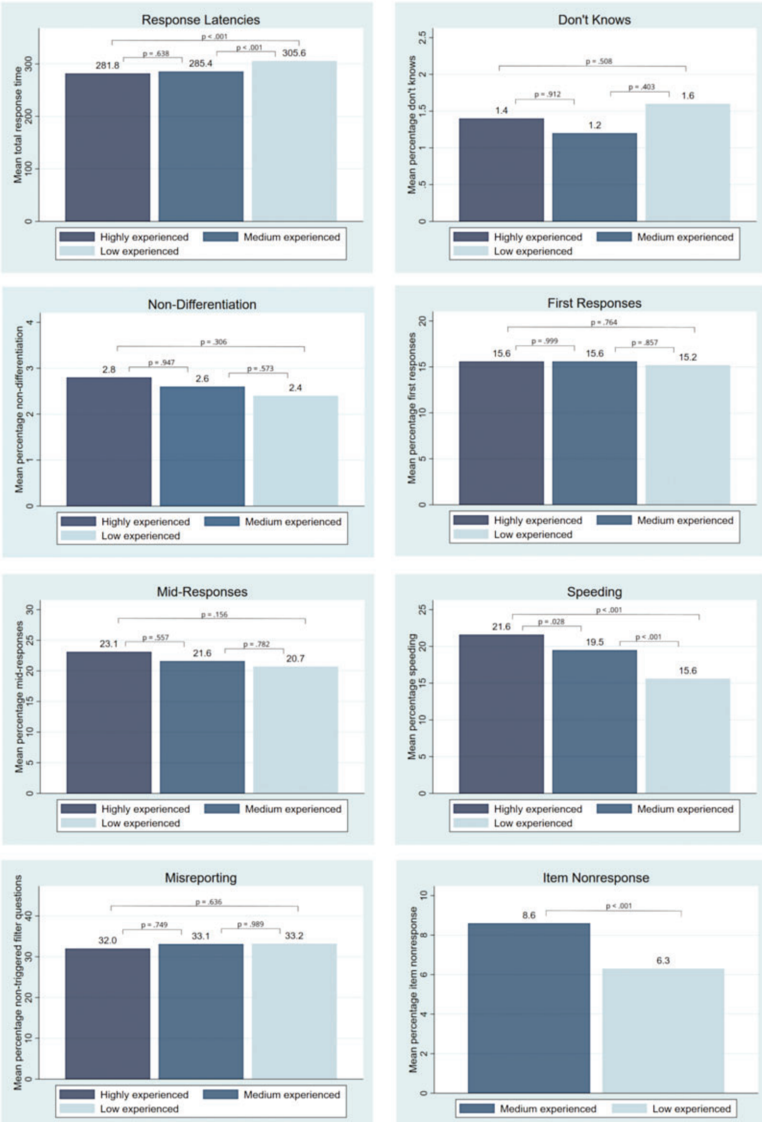

Figure 2. Pairwise Comparisons of Different Response Quality Indicators.
NOTE.—Weighted estimates; reported *p*-values are Scheffe-adjusted *p*-values based on generalized linear models; for response latencies Scheffe-adjusted *p*-values are based on linear regression models.

4.2.1 Nondifferentiation

We did not find significant differences in the extent of nondifferentiation across the three panel cohorts (figure 2). Respondents of both the medium experienced [$\beta = -0.04$, $p = .947$, OR = 0.96 (95 percent CI: [0.756, 1.221])] and low experienced [$\beta = -0.19$, $p = .306$, OR = 0.83 (95 percent CI: [0.652, 1.053])] cohorts were not significantly less likely to nondifferentiate on matrix questions compared to highly experienced panelists (see tables C1 and C2 in the supplementary data online).

4.2.2 Selection of first response option

We did not find any significant differences in the selection of first response options across the three cohorts (figure 2). Accordingly, medium experienced [$\beta = -0.00$, $p = .999$, OR = 0.10 (95 percent CI: [0.902, 1.103])] and low experienced panelists [$\beta = -0.04$, $p = .764$, OR = 0.97 (95 percent CI: [0.878, 1.061])] were not significantly less likely to select first response options as a strategy to satisfice than members of the highly experienced cohort (for detailed results see tables C1 and C2 in the supplementary data online).

4.2.3 Selection of middle response option

With respect to the proportion of selected middle response options, we did not find significant differences across the three cohorts (figure 2). Neither medium experienced [$\beta = -0.07$, $p = .557$, OR = 0.93 (95 percent CI: [0.819, 1.059])] nor low experienced panelists [$\beta = -0.13$, $p = .156$, OR = 0.88 (95 percent CI: [0.777, 1.002])] were significantly less likely to select middle responses compared to respondents of the highly experienced cohort (for detailed results see tables C1 and C2 in the supplementary data online).

4.2.4 Speeding

Our results show that less experienced respondents are significantly less likely to speed through the questionnaire than members of the highly experienced cohort (figure 2). Specifically, respondents of both the medium experienced cohort [$\beta = -0.13$, $p = .028$, OR = 0.88 (95 percent CI: [0.800, 0.966])] and the least experienced cohort [$\beta = -0.40$, $p < .001$, OR = 0.68 (95 percent CI: [0.618, 0.738])] are significantly less likely to speed than highly experienced panelists. The latter difference shows to be relatively strong with an increase in average speeding behavior of 6.0 percentage points for the highly experienced cohort compared to low experienced panelists (figure 2). Furthermore, respondents of the least experienced cohort were significantly less likely to speed compared to their medium experienced counterparts [$\beta = -0.27$, $p < .001$, OR = 0.76 (95 percent CI: [0.681, 0.855])] (see tables C1 and C2 in the supplementary data online).

4.2.5 Motivated misreporting

With respect to motivated misreporting, we expected that respondents with lower or no experience with the survey would trigger follow-up questions more often. Our results shown in figure 2, however, did not support this assumption, since we did not find significantly lower levels of misreporting among medium experienced respondents [$\beta = 0.04$, $p = .749$, OR = 1.04 (95 percent CI: [0.941, 1.149])] or members of the least experienced cohort [$\beta = 0.05$, $p = .636$, OR = 1.05 (95 percent CI: [0.951, 1.157])] compared to highly experienced respondents (see tables C1 and C2 in the supplementary data online for detailed results).

4.3 Socially desirable responding

To investigate hypothesis 4 that predicted a decrease in socially desirable responses among experienced respondents, we analyzed the prevalence of item nonresponse in sensitive questions addressing attitudes toward ethnic minorities. Since we had to use a different survey wave (i.e., the 17th wave) for this analysis, we could only compare item nonresponse in sensitive attitude questions across two panel cohorts. We found a significant difference in item nonresponse to sensitive questions between the two groups [$\beta = -0.34$, $p < .001$, OR = 0.71 (95 percent CI: [0.601, 0.845])] (figure 2; see table C2 the supplementary data online for detailed results). However, contrary to our hypothesis, medium experienced panelists did not provide fewer nonresponse answers to sensitive questions compared to their less experienced counterparts. Conversely, we observed that respondents of the low experienced cohort are less likely to provide nonresponse answers to sensitive questions than medium experienced panelists, indicating that respondents with higher experience levels may have been more prone to socially desirable responding than low experienced respondents.

4.4 Magnitude of effects

The results show distinct effects regarding the significant differences in speeding and item nonresponse across the cohorts with the average change in the probability to show one of these response patterns ranging from 17 to 60 percent: compared to highly experienced panelists, medium experienced respondents have on average a 17 percent lower probability to speed (AME $= -0.17$, $p = .028$), whereas the average probability to speed is even about 48 percent lower (AME $= -0.48$, $p < .001$) for the least experienced cohort. Compared to medium experienced respondents, respondents of the low experienced cohort have on average a 31 percent (AME $= -0.31$, $p < .001$) lower probability to speed. We find the largest effect for the extent of item nonresponse in sensitive questions with low experienced respondents having a on average 60 percent lower probability (AME = −0.60, $p < .001$) to show item nonresponse than their medium experienced counterparts. Regarding the differences in response latencies across cohorts, we find that the least experienced panelists take on average 3.0 seconds longer than highly experienced panelists to complete a question ($\beta = 23.9$ for 8 questions). Compared to the medium experienced cohort, low experienced respondents take on average 2.5 seconds more to answer a question ($\beta = 20.3$ for 8 questions). Compared to these significant findings, our results on the nonsignificant differences in given "don't knows," nondifferentiation, selection of first and middle response categories, and motivated misreporting show not unexpectedly overall small average differences between the cohorts. The effects on average range from 0.2 percent for the probability to select first response categories to a maximum of 5 percent for the probability of selecting middle responses.

5. DISCUSSION

In this article, we examined the effects of prior survey participation on the quality of survey responses provided by members of a large-scale probability-based panel. We compared three panel cohorts with differing levels of experience across the panel on a variety of response quality indicators related to the mechanisms of reflection, satisficing, and social desirability. The results provided evidence for a positive effect of panel conditioning on response quality. More experienced respondents showed significantly lower response times (when excluding the possibility of short response times due to satisficing) than members of less experienced panel cohorts, indicating an optimized response process. However, we did not find a similar effect for the other two indicators of positive panel conditioning: "don't know" responses and item nonresponse to sensitive questions. In fact, we did not find differences with respect to "don't knows," but experienced respondents provided higher item nonresponse rates than less experienced panel members. The latter finding contradicts previous literature (Waterton and Livesley 1989; Wooden and Li 2014) which suggested that experienced panelists are more likely to provide confidential information.

A possible explanation for the finding regarding item nonresponse could be that experienced respondents might have tried to appear consistent in their survey responses across waves (Waterton and Livesley 1989) and their desire to avoid cognitive dissonance arising from the conflict of their own actual behavior or attitudes and society's norms (Warren and Halpern-Manners 2012). These cognitive processes might be especially prevalent after repeatedly answering identical sensitive questions causing discomfort and embarrassment (Uhrig 2012). Considering that we did not observe significant differences in the extent of "don't know" answers across the cohorts, it is also possible that respondents learn that they can simply skip or refuse to answer a question without providing any response. Low experience respondents might not be aware of this possibility and therefore show lower item nonresponse than the more experienced cohort. However, the item nonresponse indicator we used in the analyses includes both leaving the question blank and explicitly refusing to answer by selecting "I do not want to answer." We also compared both types of item nonresponse separately across the cohorts to test the hypothesis of learning to leave a question blank. Our results support this hypothesis (see table D3 in the supplementary data online). Previous research also supports this hypothesis by showing that "don't know" answers and item refusals are indeed different types of item nonresponse that result from different disruptions in the cognitive response process (Silber et al. 2021).

With respect to negative panel conditioning, we found evidence for a decrease in response quality as a result of speeding. Specifically, highly experienced respondents showed higher levels of speeding compared to less experienced respondents. We did not find, however, a significant conditioning effect on nondifferentiation, the selection of first or middle response options, and misreporting to filter questions. Comparing those results to previous research using the Dutch probability-based LISS Panel (Schonlau and Toepoel 2015), our study did not replicate their findings of higher levels of nondifferentiation among more experienced respondents. Our findings on motivated misreporting, however, support some of the previous research that did not provide evidence for an increase in misreporting to filter questions with higher panel experience (Silber et al. 2019; Bach and Eckman 2020; Eckman and Bach 2021).

Considering the three respondent groups with differing levels of experience, the main differences we found were between panel members with the highest and lowest levels of experience. In fact, we only found significant differences between members of the highly and medium experienced cohort for one indicator (speeding). Apart from that, we did not find significant differences between highly and medium experienced respondents, potentially indicating that learning processes within a panel take place at an early stage after initial participation. However, the effects between members of the low and medium experienced cohorts were usually weaker than the effects between low and highly experienced panelists. This, on the other hand, suggests that a certain amount of experience might be necessary to evoke a positive or negative learning process within a panel study. Yet, due to the variety of findings for the different indicators and the multiple group comparisons, it was not possible to conclude on the amount of experience that causes the strongest learning effects. In this study, the difference in experience between the low and highly experienced panel members was about five years (or 28 waves), whereas the difference between the low and medium experienced respondents was only about two years (or 14 and 16 waves, respectively).

Although the effects of survey experience on most indicators of response quality appeared to be relatively small, larger effects of speeding exist for panel members with higher levels of experience (i.e., on average 6.0 percentage points difference between the highest and least experienced cohort). Speeding indicates lower response quality because respondents do not take enough time to properly process questions and to provide optimal answers. Thus, future research should examine this considerable increase in speeding behavior in greater detail by studying potential explanations such as respondent motivation, survey enjoyment, topic interest, or panel fatigue. Overall, our results suggest a dose–response relationship by showing graded response patterns for the different levels of experience within the panel, although many of the differences between the cohorts were not significant.

Several limitations need to be considered when generalizing the findings of the present research. As is typical for studies of panel conditioning that use a quasi-experimental design, the amount of experience and the duration of membership in the panel cannot be disentangled (Struminskaya 2016). Thus, we do not know whether one or the other may have caused the group differences that we observed. Also, due to the study design, there is the danger of confounding effects of panel attrition. To mitigate the attrition problem, we followed previous research using adjustment weighting (Dennis 2001; Nancarrow and Cartwright 2007; Struminskaya 2016). Furthermore, the selection of indicators was driven by the available data which were not specifically designed to identify panel conditioning effects. The reflection indicator of shorter total response times across more experienced panelists could also (partly) reflect less engaged response behaviors instead of optimized response processes. Although we excluded speeders from the indicator and differences remained significant after controlling for several satisficing response behaviors, we invite future studies to test the robustness of these findings.

SUPPLEMENTARY MATERIALS

Supplementary materials are available online at academic.oup.com/jssam.

REFERENCES

American Association for Public Opinion Research (2016), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th ed.), AAPOR.

Andreß, H.-J., Golsch, K., and Schmidt, A. W. (2013), *Applied Panel Data Analysis for Economic and Social Surveys*, Berlin, Heidelberg: Springer Science & Business Media.

Bach, R. L., and Eckman, S. (2018), "Motivated Misreporting in Web Panels," *Journal of Survey Statistics and Methodology*, 6, 418–430.

———. (2020), "Rotation Group Bias in Reporting of Household Purchases in the US Consumer Expenditure Survey," *Economics Letters*, 187, 108889.

Bailar, B. A. (1975), "The Effects of Rotation Group Bias on Estimates from Panel Surveys," *Journal of the American Statistical Association*, 70, 23–30.

Basso, M. R., Lowery, N., Ghormley, C., and Bornstein, R. A. (2001), "Practice Effects on the Wisconsin Card Sorting Test—64 Card Version across 12 Months," *The Clinical Neuropsychologist*, 15, 471–478.

Battaglia, M. P., Hoaglin, D. C., and Frankel, M. R. (2009), "Practical Considerations in Raking Survey Data," *Survey Practice*, 2, 2953.

Beatty, P., and Herrmann, D. (2002), "To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse," *Survey Nonresponse*, 71, 86.

Bergmann, M., and Barth, A. (2018), "What Was I Thinking? A Theoretical Framework for Analysing Panel Conditioning in Attitudes and (Response) Behaviour," *International Journal of Social Research Methodology*, 21, 333–345.

Binswanger, J., Schunk, D., and Toepoel, V. (2013), "Panel Conditioning in Difficult Attitudinal Questions," *Public Opinion Quarterly*, 77, 783–797.

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., and Weyandt, K. W. (2018), "Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel," *Social Science Computer Review*, 36, 103–115.

Bretschi, D., Schmidt, K., and Schulz, M. (2019), *GESIS Panel Wave Report: Wave fe*, Cologne: GESIS—Leibniz Institute for the Social Sciences.

Chyung, S. Y., Roberts, K., Swanson, I., and Hankinson, A. (2017), "Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale," *Performance Improvement*, 56, 15–23.

Crossley, T. F., de Bresser, J., Delaney, L., and Winter, J. (2017), "Can Survey Participation Alter Household Saving Behaviour?," *The Economic Journal*, 127, 2332–2357.

Daikeler, J., Bach, R. L., Silber, H., and Eckman, S. (2020), "Motivated Misreporting in Smartphone Surveys," *Social Science Computer Review*, 40, 95–107.

Das, M., Toepoel, V., and van Soest, A. (2011), "Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys," *Sociological Methods & Research*, 40, 32–56.

Dennis, J. M. (2001), "Are Internet Panels Creating Professional Respondents," *Marketing Research*, 13, 34−38.

Duan, N., Alegria, M., Canino, G., McGuire, T. G., and Takeuchi, D. (2007), "Survey Conditioning in Self-Reported Mental Health Service Use: Randomized Comparison of Alternative Instrument Formats," *Health Research and Educational Trust*, 42, 890–907.

Eckman, S., and Bach, R. (2021), "Panel Conditioning in the US Consumer Expenditure Survey," *Journal of Official Statistics (JOS)*, 37, 53−69.

Eckman, S., and Kreuter, F. (2018), "Misreporting to Looping Questions in Surveys: Recall, Motivation and Burden," *Survey Research Methods*, 12, 59–74.

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., and Presser, S. (2014), "Assessing the Mechanisms of Misreporting to Filter Questions in Surveys," *Public Opinion Quarterly*, 78, 721–733.

Fendrich, M., and Vaughn, C. M. (1994), "Diminished Lifetime Substance Use over Time: An Inquiry into Differential Underreporting," *Public Opinion Quarterly*, 58, 96–123.

GESIS (2021), "GESIS Panel—Standard Edition," GESIS Data Archive, Cologne, ZA5665 Data file Version 41.0.0. Available at https://doi.org/10.4232/1.13785.

Halpern-Manners, A., and Warren, J. R. (2012), "Panel Conditioning in Longitudinal Studies: Evidence from Labor Force Items in the Current Population Survey," *Demography*, 49, 1499–1519.

Halpern-Manners, A., Warren, J. R., and Torche, F. (2017), "Panel Conditioning in the General Social Survey," *Sociological Methods & Research*, 46, 103–124.

Höhne, J. K., and Schlosser, S. (2018), "Investigating the Adequacy of Response Time Outlier Definitions in Computer-Based Web Surveys using Paradata SurveyFocus," *Social Science Computer Review*, 36, 369–378.

Holbrook, A. L., Krosnick, J. A., Moore, D., and Tourangeau, R. (2007), "Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes," *Public opinion Quarterly*, 71, 325–348.

Kalton, G., Kasprzyk, D., and McMillen, D. (1989), *Panel Surveys*, New York: Wiley.

Kaminska, O., McCutcheon, A. L., and Billiet, J. (2010), "Satisficing among Reluctant Respondents in a Cross-National Context," *Public Opinion Quarterly*, 74, 956–984.

Kreuter, F., Eckman, S., and Tourangeau, R. (2020), "The Salience of Survey Burden and Its Effect on Response Behavior to Skip Questions: Experimental Results from Telephone and Web Surveys," in *Advances in Questionnaire Design, Development, Evaluation and Testing*, eds. P. Beatty, D. Collins, L. Kaye, J. Padilla, G. Willis, and A. Wilmot, Hoboken, NJ: Wiley, pp. 213–227.

Kroh, M., Winter, F., and Schupp, J. (2016), "Using Person-Fit Measures to Assess the Impact of Panel Conditioning on Reliability," *Public Opinion Quarterly*, 80, 914–942.

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

Lugtig, P. (2014), "Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers," *Sociological Methods & Research*, 43, 699–723.

Lugtig, P., Das, J. W. M., and Scherpenzeel, A. C. (2014), "Nonresponse and Attrition in a Probability-Based Online Panel for the General Population," in *Online Panel Research: A Data Quality Perspective*, eds. M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, and P. J. Lavrakas, New York: Wiley, pp. 135−153.

Lundmark, S., and Gilljam, M. (2013), "Participation Effects in Panel Surveys: Evidence from a Seven-Wave Randomized Experiment," in Annual Conference of American Association for Public Opinion Research, pp. 16–19, Boston, May.

Lynn, P. (2009), "Methods for Longitudinal Surveys," in *Methodology of Longitudinal Surveys*, ed. P. Lynn, Chichester, UK: Wiley, pp. 1–19.

Mathiowetz, N. A., and Lair, T. J. (1994), "Getting Better? Changes or Errors in the Measurement of Functional Limitations," *Journal of Economic and Social Measurement*, 20, 237–262.

Matjašič, M., Vehovar, V., and Manfreda, K. L. (2018), "Web Survey Paradata on Response Time Outliers: A Systematic Literature Review," *Advances in Methodology and Statistics*, 15, 23–41.

Müller, B., and Schmiedeberg, C. (2021), "Do Respondents Get Used to Answering Sensitive Questions?: Refusal of Items on Sexuality and Fertility in a Panel Survey," *Public Opinion Quarterly*, 84, 654–674. https://doi.org/10.1093/poq/nfaa041.

Nadler, J. T., Weston, R., and Voyles, E. C. (2015), "Stuck in the Middle: The Use and Interpretation of Mid-Points in Items on Questionnaires," *The Journal of General Psychology*, 142, 7189.

Nancarrow, C., and Cartwright, T. (2007), "Online Access Panels and Tracking Research. The Conditioning Issue," *International Journal of Market Research*, 49, 573–594.

Pevalin, D. J. (2000), "Multiple Applications of the GHQ-12 in a General Population Sample: An Investigation of Long-Term Retest Effects," *Social Psychiatry and Psychiatric Epidemiology*, 35, 508–512.

Pötzschke, S., Schaurer, I., Bretschi, D., and Bauer, R. (2017), *GESIS Panel Wave Report: Wave de*, Cologne: GESIS—Leibniz Institute for the Social Sciences.

Phillips, D. L., and Clancy, K. J. (1972), "Some Effects of "Social Desirability" in Survey Studies," *American journal of Sociology*, 77, 921–940.

Roßmann, J., Gummer, T., and Silber, H. (2018), "Mitigating Satisficing in Cognitively Demanding Grid Questions: Evidence from Two Web-Based Experiments," *Journal of Survey Statistics and Methodology*, 6, 376–400.

Schnell, R. (1994), *Graphisch Gestützte Datenanalyse*, Berlin, Boston: Oldenbourg Wissenschaftsverlag.

Schonlau, M., and Toepoel, V. (2015), "Straightlining in Web Survey Panels over Time," *Survey Research Methods*, 9, 125–137.

Silber, H., Roßmann, J., Gummer, T., Zins, S., and Weyandt, K. W. (2021), "The Effects of Question, Respondent and Interviewer Characteristics on Two Types of Item Nonresponse," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184, 1052–1069.

Silber, H., Schröder, J., Struminskaya, B., Stocké, V., and Bosnjak, M. (2019), "Does Panel Conditioning Affect Data Quality in Ego-Centered Social Network Questions?," *Social Networks*, 56, 45–54.

Struminskaya, B. (2016), "Respondent Conditioning in Online Panel Surveys: Results of Two Field Experiments," *Social Science Computer Review*, 34, 95–115.

———. (2020), *Panel Conditioning*, SAGE Publications Limited.

Struminskaya, B., and Bosnjak, M. (2021), “Panel Conditioning: Types, Causes, and Empirical Evidence of What We Know So Far,” in *Advances in Longitudinal Survey Methodology*, ed. P. Lynn, Hoboken, NY: Wiley, pp. 272–301.

Sturgis, P., Allum, N., and Brunton-Smith, I. (2009), “Attitudes Over Time: The Psychology of Panel Conditioning,” in *Methodology of Longitudinal Surveys*, ed. P. Lynn, Wiley, pp. 113–126.

Sun, H., Tourangeau, R., and Presser, S. (2018), “Panel Effects: Do the Reports of Panel Respondents Get Better or Worse over Time?,” *Journal of Survey Statistics and Methodology*, 7, 572–588.

Toepoel, V., Das, M., and van Soest, A. (2009), “Relating Question Type to Panel Conditioning: Comparing Trained and Fresh Respondents,” *Survey Research Methods*, 3, 73–80.

Torche, F., Warren, J. R., Halpern-Manners, A., and Valenzuela, E. (2012), “Panel Conditioning in a Longitudinal Study of Adolescents’ Substance Use: Evidence from an Experiment,” *Social Forces*, 90, 891–918.

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

Tourangeau, R., and Yan, T. (2007), “Sensitive Questions in Surveys,” *Psychological Bulletin*, 133, 859–883.

Uhrig, S. N. (2012), “Understanding Panel Conditioning: An Examination of Social Desirability Bias in Self-Reported Height and Weight in Panel Surveys using Experimental Data,” *Longitudinal and Life Course Studies*, 3, 120–136.

Van Landeghem, B. (2012), “A Test for the Convexity of Human Well-Being over the Life Cycle: Longitudinal Evidence from a 20-Year Panel,” *Journal of Economic Behavior & Organization, Elsevier*, 81, 571–582.

Warren, J. R., and Halpern-Manners, A. (2012), “Panel Conditioning in Longitudinal Social Science Surveys,” *Sociological Methods & Research*, 41, 491–534.

Waterton, J., and Livesley, D. (1989), “Evidence of Conditioning Effects in the British Social Attitudes Panel,” in *Panel Surveys*, eds. D. Kasprzyk, G. Duncan, G. Kalton, and M. P. Singh, New York: Wiley, pp. 319–339.

Wolter, K. M. (2007), “Taylor Series Methods,” in *Introduction to Variance Estimation. Statistics for Social and Behavioral Sciences*, New York, NY: Springer, pp. 226–271.

Wooden, M., and Li, N. (2014), “Panel Conditioning and Subjective Well-Being,” *Social Indicators Research*, 117, 235–255.

Wright, S. P. (1992), “Adjusted p-Values for Simultaneous Inference,” *Biometrics*, 48, 1005–1013.

Yan, T., and Eckman, S. (2012), “Panel Conditioning: Change in True Value versus Change in Self-Report,” in Proceedings of the Survey Methods Research Section, Alexandria, VA: ASA, pp. 4726–4736.

Zhang, C., and Conrad, F. (2014), “Speeding in Web Surveys: The Tendency to Answer Very Fast and Its Association with Straightlining,” *Survey Research Methods*, 8, 127–135.