*Journal of Survey Statistics and Methodology (2022) **10**, 249–275*

POSITIVE LEARNING OR DEVIANT INTERVIEWING? MECHANISMS OF EXPERIENCE ON INTERVIEWER BEHAVIOR

YULIYA KOSYAKOVA *
LUKAS OLBRICH
JOSEPH W. SAKSHAUG
SILVIA SCHWANHÄUSER

Interviewer (mis)behavior has been shown to change with interviewers' professional experience (general experience) and experience gained during the field period (survey experience). We extend this study by using both types of experiences to analyze interviewer effects on a core quality indicator: interview duration. To understand whether the effect of

YULIYA KOSYAKOVA is a Post-Doctorate Senior Researcher and Associate Lecturer with the Institute for Employment Research (IAB), Nuremberg, Germany; University of Mannheim, Baden-Württemberg, Germany; and Otto-Friedrich University of Bamberg, Bamberg, Germany. LUKAS OLBRICH is PhD Student with the Institute for Employment Research (IAB), Nuremberg, Germany and Germany and Ludwig-Maximilian University of Munich, Germany. JOSEPH W. SAKSHAUG is Professor of Statistics with the Institute for Employment Research (IAB), Nuremberg, Germany; University of Mannheim, Baden-Württemberg, Germany; and Ludwig-Maximilian University of Munich, Munich, Germany. SILVIA SCHWANHÄUSER is PhD Candidate with the Institute for Employment Research (IAB), Nuremberg, Germany and University of Mannheim, Baden-Württemberg, Germany.

The authors thank Christina Reisinger for valuable input, Johannes Ludsteck, and Mark Trappmann for providing helpful advice on earlier versions of this article, Marieke Volkert for engaging in insightful discussions at various stages of the research process, Simon Huber and KANTAR Team for providing additional information on the data collection process, and all those who commented on the earlier presentations of this study at the following events: the 8th Conference of the European Survey Research Association (ESRA) in Zagreb (July 2019) and the Advanced Seminar of Statistics and Econometrics (summer term 2019) at the University of Bamberg. The authors also thank the Editor and two anonymous referees for helpful comments and suggestions. Data access was provided via researcher contacts at the Institute for Employment Research (IAB). External researchers may apply for access to these data by submitting a user-contract application to the SOEP Research Data Center (https://www.diw.de/en/diw_02.c.222836.en/data_access_and_order.html). The computer code for the analysis is available at https://osf.io/ptkxg/. This study design and analysis was not preregistered.

Financial support was obtained from the Charles Cannell Fund in Survey Methodology through the Grant awarded to Yuliya Kosyakova.

*Address correspondence to Yuliya Kosyakova, Institute for Employment Research (IAB), Regensburger Str. 104, DE-90478 Nuremberg, Germany. E-mail: Yuliya.Kosyakova@iab.de.

doi: 10.1093/jssam/smab003 Advance access publication 22 February 2021

interviewer experience on duration is driven by increased efficiency or deviant behavior—both mechanisms of shorter interview durations—we additionally examine the triggering rate of filter questions to avoid burdensome follow-up questions and response differentiation over the field period. Using multilevel models and data from a large-scale survey on a special and difficult-to-interview population of refugees in Germany, we find that interview duration decreases with increasing survey experience, particularly among the generally inexperienced interviewers. However, this effect is not found for the triggering rate and response differentiation. The results are robust to different sample and model specifications. We conclude that the underlying mechanism driving interview duration is related to increasing efficiency, and not deviant behavior.

KEYWORDS: Filter questions; Interview duration; Interviewer experience; Response differentiation; Triggering rate.

1. INTRODUCTION

Despite the increasing importance of self-administered (web or mail) surveys, interviewers continue to play a pivotal role in administering large-scale face-to-face surveys for scientific and policy-oriented research. Interviewers are tasked with many responsibilities, including eliciting cooperation from households, implementing respondent selection rules, recording observations about the sampled units and their neighborhoods, administering the questionnaire, assisting respondents with the interview, among other duties (Groves et al. 2009; Josten and Trappmann 2016). With these responsibilities, interviewers have a direct influence on data quality and, correspondingly, multiple components of survey error as defined within the Total Survey Error framework (Groves et al. 2009). One of the central aspects of data quality—and the focus of this study—is concerned with interviewers' ability to accurately follow survey guidelines and instructions.

A core indicator of whether interviewers accurately follow the prescribed guidelines and instructions during the survey interview is interview duration, and, in particular, short interview durations (Herzog and Bachman 1981; Kirchner and Olson 2017). There are at least two mechanisms by which interviewers can influence interview durations downward. First, if interviewers have strong incentives to reduce the length of the interview, they are more likely to deviate from the prescribed interviewing procedures to shorten the interview. This misbehavior emerges particularly when interviewers are paid per completed case rather than hourly (Matschinger, Bernert, and Angermeyer 2005; Kosyakova, Skopek, and Eckman 2015). Such deviant behavior could increase with professional experience. Specifically, research shows that as interviewers gain more experience they become less careful with the prescribed procedures, leading to decreasing interview durations and potentially

influencing data quality in a negative way (Olson and Peytchev 2007; Olson and Bilgen 2011; Loosveldt and Beullens 2013a). The second mechanism is that shorter interviews could be the result of positive learning behavior for the reason that interviewers learn to become more efficient with their tasks as they gain more familiarity with their role and acquire more interviewing experience (Olson and Smyth 2019).

In this article, we investigate the role of interviewer experience on these two distinct mechanisms by applying multilevel models to a probability-based survey conducted on a special (and difficult-to-interview) population of refugees in Germany. Because of the particular burden on interviewers owing to language challenges, long questionnaires with sensitive items, as well as special housing situations of the respondents, these data represent an ideal setting for studying possible interviewers' misbehavior. Hence, we deem this survey to be particularly suitable for our research purposes as compared to a survey with a more general target population and a more standard interviewer staff.

We focus on two types of interviewer experience: (1) interviewers' survey-specific experience gained during the field period for a given survey at hand (henceforth referred to as *survey experience*) and (2) interviewers' prior experience with the survey organization (henceforth referred to as *general experience*). To study the impact of interviewers' experience on interview duration, we additionally consider two data quality indicators: the triggering rate of filter questions and response differentiation. Although interview duration is a common indicator for assessing potentially deceptive behavior of interviewers (Hood and Bushery 1997; Bushery, Reichert, Albright, and Rossiter 1999), the literature is less clear-cut on whether the effect of interviewer experience on duration is driven by deviant behavior or arises owing to increasing efficiency of administering a survey. This is the reason why we additionally look at the triggering rate—a potential gateway to manipulate the number of follow-up questions asked to keep the interview short (Matschinger et al. 2005; Loosveldt and Beullens 2013b; Kosyakova et al. 2015). Response differentiation, that is, the (dis)similarity of responses to a set of adjacent same-scaled items (Yan 2008), works as an additional indicator of inadequate interview speeding. Looking at multiple indicators of interviewers' (mis)behavior is an advantage to disentangle the primary mechanism behind interviewers' effects and, correspondingly, potential data quality.

2. INTERVIEWERS' EXPERIENCE AND INTERVIEWERS' BEHAVIOR

2.1 Interviewers' (Mis)behavior and Associated Mechanisms

Deliberate interviewers' misbehavior addresses intended deviations from interviewing guidelines (AAPOR 2003). Deviant behavior may take several forms

such as departing from strictly standardized interviewing, showing detrimental behavior (such as omitting parts of the interview), completely falsifying the interview, or providing unscripted feedback to respondents (Fowler and Mangione 1990; Loosveldt 2008). The hierarchy of quality controls implemented between the survey organization and its interviewers can influence the occurrence of this deviant interviewer behavior. Particularly in face-to-face settings, there is a lack of direct control on the part of the survey organization, with the exceptions of Computer-Assisted Recorded Interviewing (CARI) (Hicks et al. 2010; Edwards, Sun, and Hubbard 2020) and GPS monitoring (Edwards, Maitland, and Connor 2017), which are rarely implemented. Based on this lack of direct control, interviewers are trusted to ensure that the interaction with respondents follows the rules set (van der Zouwen and Dijkstra 1988).

Therefore, the hierarchy of controls can be considered as a principal–agent problem with the survey research organization as principal and the interviewer as agent. Since the principal can only observe the agents' outcome but not their actions in detail, a so-called moral hazard problem may arise (Haller 1985; Kosyakova et al. 2015). Interviewers' incentives to increase their own productivity by reducing interview duration might be contrary to the survey organization's objective of collecting high-quality survey data (Groves et al. 2009). In line with that, the literature on interviewer falsification supports the notion that honest interviewers produce significantly longer interviews than falsifiers do (Schreiner, Pennie, and Newbrough 1988). Thus, interview duration is considered as a crucial indicator for identifying deceptive interviewer behavior (Loosveldt and Beullens 2013a, 2013b; Olson and Peytchev 2007).

2.2 Interview Duration and Interviewer Experience

Previous research indicates that standardized interviewing behaviors may be omitted or shortened with growing survey experience, leading to decreased interview length and potentially adverse data quality (Chromy, Hughes, Giacoletti, and Odom 2002; Olson and Smyth 2019). Hypothesized mechanisms to explain unstandardized interviewer behaviors include carelessness and fatigue but also general experience (Chromy et al. 2002; Olson and Peytchev 2007; Bilgen 2011). One possible explanation for experienced interviewers following protocols less closely than inexperienced interviewers is that they bring their own shortcutting habits learned in previous interviews (Bilgen 2011; Olson and Bilgen 2011). Another explanation put forward is that with growing experience interviewers work more efficiently by, for example, not engaging in unnecessary side-talk, resolving difficulties more easily, and implementing more efficient communicative strategies (Olson and Smyth 2019). Looking specifically at the type of interviewer experience and its effect on interview duration, prior results have shown that interviewers perform their

interviewing tasks faster with growing survey experience acquired during the field period (Olson and Peytchev 2007; Loosveldt and Beullens 2013a, 2013b; Böhme and Stöhr 2014; Kirchner and Olson 2017). Hence, we formulate the following testable hypothesis:

H1a: Interview duration decreases with growing survey (field) experience.

At the same time, previous research suggests that generally experienced interviewers (i.e. those who have prior experience working for the same survey organization) conduct, on average, shorter interviews compared to generally inexperienced interviewers (Olson and Bilgen 2011; Olson and Peytchev 2007; though see Kirchner and Olson 2017 for no differences in interview duration owing to general experience in a CATI setting). They are able to draw upon a greater level of prior knowledge instilled by the survey organization and previously acquired interviewing behaviors (Olson and Peytchev 2007; Bilgen 2011). Accordingly, we expect that:

H2a: Interview duration is shorter for generally experienced interviewers compared to generally inexperienced interviewers.

Moreover, generally inexperienced interviewers show greater learning effects with their growing survey experience; thus, the effect of survey experience on reduced interview duration is expected to be stronger for the generally inexperienced interviewers than for experienced ones (Olson and Peytchev 2007). In other words, generally experienced interviewers bring their (positive or negative) habits and change their behavior to a lesser extent with increasing survey experience, whereas inexperienced interviewers may realize and implement shortcutting possibilities to a greater extent with increasing survey experience. This mechanism would predict that:

H3a: Interview duration decreases with growing survey experience more strongly for generally inexperienced interviewers than for generally experienced interviewers.

At this stage, however, we are unable to distinguish whether both types of interviewer experience are associated with increasing efficiency or rather detrimental rational behavior to keep the interview short. We address this issue in the next section.

2.3 Triggering Rate of Filter Questions

Dropping or manipulating answers to filter questions to avoid follow-up questions illustrates a possible gateway to keep the interview short and to reduce effort (Hood and Bushery 1997). Following this premise, recent studies have assessed the manipulation of filter questions as a possible interviewer behavior to ease interviewer and respondent burden, but also to reduce interview

duration, particularly in face-to-face settings (Chromy et al. 2002; Japec 2008; Groves et al. 2009; Kosyakova et al. 2015; Josten and Trappmann 2016).

Following the above rationale that generally experienced interviewers bring their own habits when beginning a new field period, we can assume that they also bring knowledge about how to take shortcuts during the administration of filter questions (Menold, Winker, Storfinger, and Kemper 2013; Ruckdeschel, Sauer, and Naderi 2016). However, with growing survey experience, learning effects may arise for generally inexperienced interviewers as well. We assume these learning effects to be larger compared to learning effects over the field period for generally experienced interviewers (van Tilburg 1998; Olson and Peytchev 2007). Potential moderators of detrimental behavior of generally inexperienced interviewers are increasing interviewer burden and fatigue, which are likely to be more pronounced, compared to generally experienced interviewers having already established interviewing routines (Japec 2008).

Bringing everything together, we may presuppose that if the underlying mechanism behind the interviewer's (survey and general) experience was related to interviewer deviance, the hypotheses formulated above would also hold for the triggering rate. Accordingly, we expect that:

H1b: Triggering rate decreases with growing survey experience.

H2b: Triggering rate is lower for generally experienced interviewers compared to generally inexperienced interviewers.

H3b: Triggering rate decreases with growing survey experience more strongly for generally inexperienced interviewers than for generally experienced interviewers.

If we find no support for hypotheses *H1b-H3b*, but do find support for *H1a-H3a*, then we may conclude that the mechanism driving interview duration is rather related to increasing efficiency and not interviewer deviance.

2.4 Response Differentiation

Response differentiation refers to the level of variation between responses to a battery of same-scaled items (Yan 2008). Low differentiation implies that the respondent selected similar response options for all items. An extreme form of such patterns is straightlining (Herzog and Bachman 1981), where the same response is selected for all items. Low-response differentiation or straightlining has been classified as a form of strong satisficing behavior (Krosnick 1991) used by respondents to avoid the cognitive effort of carefully responding to each item (Yan 2008). Low-response differentiation is also a well-known predictor of deviant interviewer behavior (Menold et al. 2013; Menold and Kemper 2014). For instance, deviant interviewers may ask only the first item in a battery and falsify the remaining responses by using responses similar to the first response. With regard to the related interview duration, interviewers

who administer fast-paced interviews can challenge respondents to adequately respond to items and, hence, drive satisficing behavior and lower response differentiation (Yan 2008; Vandenplas, Loosveldt, Beullens, and Denies 2018). The notion of a fast interview pace was studied experimentally by Cannell, Miller, and Oksenberg (1981), who observed interviewers who: "ask questions at a rapid rate" and "permit no pause between response and the next question" (Cannell et al. 1981).

In this context, we posit that if shorter interview durations for generally experienced interviewers are driven by more efficient interviewing practices (Olson and Smyth 2019), then we would expect no differences in response differentiation between generally experienced and inexperienced interviewers. However, if the cause of shorter interview durations is driven by habits of deviant-speeding behaviors learned from previous work experiences (Kirchner and Olson 2017), then we would expect lower response differentiation for generally experienced interviewers. Similar arguments can be developed for survey experience. Shorter interview durations caused by increasing efficiency should result in no changes in response differentiation over the field period. In contrast, if shorter interview durations are a result of hurried interviews—a potential habit which is likely reinforced over the field period—response differentiation will decrease as survey experience increases. It should be noted that generally experienced interviewers may have already developed such routines in their previous work as interviewers, which is the reason why the effects of survey experience are expected to be stronger for generally inexperienced interviewers.

Thus, if the effects of general and survey experience on interview duration are driven by interviewers' misbehavior, the hypotheses derived for interview duration will also hold for response differentiation:

H1c: Response differentiation decreases with growing survey experience.

H2c: Response differentiation is lower for generally experienced interviewers compared to generally inexperienced interviewers.

H3c: Response differentiation decreases with growing survey experience more strongly for generally inexperienced interviewers than for generally experienced interviewers.

Similar to hypotheses *H1b-H3b*, if we find no support for hypotheses *H1c-H3c*, but do for hypotheses *H1a-H3a*, then we may conclude that the mechanism at work is rather related to the increasing efficiency of experienced interviewers and not interviewer deviance.

3. DATA AND METHOD

3.1 Data and Sample

This article uses factually anonymous data from the first wave of the German panel study "IAB-BAMF-SOEP Survey of Refugees in Germany" launched in 2016 (data version: V34) (Brücker, Rother, and Schupp 2017). The IAB-BAMF-SOEP Survey of Refugees in Germany is a longitudinal household survey conducted annually in Germany. The target population of refugees is based on the German Central Register of Foreigners (*Ausländerzentralregister*; AZR) and includes adult anchor persons and their adult household members who arrived in Germany between January 2013 and January 2016 and were recorded by the AZR no later than June 30, 2016 (Kroh, Kühne, Jacobsen, Siegert, and Siegers 2017; Jacobsen 2018). The first wave was launched in 2016 which yielded 3,289 responding households and 4,465 total adult respondents. In total, ninety-five interviewers completed between 1 and 282 interviews (mean, forty-seven) by means of computer-assisted personal interviewing (CAPI). The household-level response rate was 48.8 percent after the calculation of AAPOR Response Rate 2 (AAPOR 2016; Kroh et al. 2017).

Compared to general population surveys, a refugee survey is appropriate for studying interviewer effects, given that interviewers face potentially greater challenges in overcoming language barriers and carrying out standardized interviewing procedures with this vulnerable population (Bloch 1999, 2007). Owing to these burdens, the IAB-BAMF-SOEP Survey of Refugees in Germany might be particularly prone to deviant interviewer behavior. Together with language challenges, a long questionnaire with sensitive items on the refugee's way to Germany and their reason(s) for leaving their home country, as well as special housing situations of the respondents such as reception centers represents a setting, which provides multiple reasons for interviewers to deviate from the prescribed procedures. Thus, our analysis provides an extreme test of the set of hypotheses, as incentives for deviation are considerably larger than in general population surveys.

Around 50 percent of the adult respondents were born in Syria, followed by Iraq (13 percent), Afghanistan (12 percent), Eritrea and Somalia (6 percent), West Balkans (Albania, Serbia, Kosovo; 4 percent), Iran and Pakistan (4 percent), and around 12 percent from other countries. To facilitate interviewing respondents without German language proficiency, the questionnaires were provided in seven languages (Arabic, English, Farsi/Dari, German, Kurmanji, Pashtu, and Urdu). Audio files and access to an interpreter hotline supported respondents with reading difficulties. The person-level questionnaire's principal topics included migration and escape history, pre- and post-migration biographies on education, language acquisition, and employment, as well as satisfaction in different life domains, health, attitudes, and cognitive and

noncognitive skills. Overall, the person questionnaire contained around 450 possible questions (Brücker et al. 2017).

For our empirical analysis, we consider only person interviews conducted with adult (>17 years) household members. We trimmed the interview duration at the 99th percentile, as some interview durations were implausibly high and likely do not represent the true interview duration. Replication of our benchmark models with inclusion of excluded interviews provided similar results (table S4 of the online Supplementary Material). Correspondingly, exclusion of the implausible values as well as missing information for the dependent variables ($n = 98$; 2.1 percent) resulted in a final analysis sample of 4,367 interviews.

3.2 Analysis Variables

3.2.1 Dependent variables. The first dependent variable is *interview duration* in minutes, which was automatically captured in the CAPI instrument. The interview duration ranges from 7 up to 248 minutes and its distribution is slightly right-skewed: the mean interview duration is 89.6 minutes and the median is eighty-three minutes. The standard deviation is 37.7.

To calculate the *triggering rate*, the second dependent variable, we considered questionnaire sections containing at least ten filter questions. The corresponding sections include (1) escape route to Germany, (2) participation in integration courses, (3) current employment situation, (4) education and degrees obtained, and (5) family composition. Table S1 of the online Supplementary Material summarizes the distribution of presented filter questions, follow-up questions, and the total number of these questions per section. The triggering rate (T_i) is defined as the average share (in percent) of the total number of asked follow-up questions (f_{asked}) relative to the maximum possible number of follow-up questions ($f_{\max}$) across all sections (S) within a person-interview (i):

$$T_i = 100 * \frac{1}{S} \sum_{s=1}^{S} \frac{f_{i,s,\text{asked}}}{f_{i,s,\max}} \tag{1}$$

The corresponding variable ranges from 11.4 to 42.5 percent with mean of 22.7 percent (median, 22.1 percent) and standard deviation of 4.0.

To measure the third dependent variable, *response differentiation* (Kemper and Menold 2014), we first calculated the standard deviation of responses to eight sets of same-scaled questions containing at least five items (table S2 of the online Supplementary Material lists these item batteries). Second, we standardized each of the standard deviations to a Z-score since the response scales differ across the batteries. Third, we calculated the average standardized standard deviation for each interview. More technically, response differentiation (D_i) is calculated using the following formula:

$$D_i = \frac{1}{K}\sum_{k=1}^{K}\frac{SD_{i,k} - \bar{SD}_k}{\sigma_k}, \quad (2)$$

where K denotes the total number of answered item batteries, $SD_{i,k}$ is the standard deviation of responses in item battery k in interview i, $\bar{SD}_k$ is the average standard deviation of responses to item battery k across all interviews, and σ_k is the standard deviation of $SD_{i,k}$ across all interviews. The response differentiation ranges from −2.3 to 2.1 with mean 0.0 and standard deviation 0.5. A lower value for this variable implies less differentiation.

3.2.2 Independent variables. In this study, interviewer survey experience and general experience are focal predictors. *Survey experience* represents the sequential number of cumulative completed interviews derived from time-stamp data. It equals one for an interviewer's first interview and increments by one for each subsequent interview (Olson and Peytchev 2007; Kosyakova et al. 2015). Values for survey experience range from 1 to a maximum of 282 interviews across all interviewers. The variable was log-transformed in the multivariate analyses to account for nonlinear learning effects (Kirchner and Olson 2017). The second measure of interviewers' experience refers to their prior interviewing experience with the survey organization, that is, the German Socio-Economic Panel (GSOEP). *Generally inexperienced* interviewers are coded as one in the case of no prior experience with the GSOEP and zero otherwise (i.e. *generally experienced* interviewers). Roughly 71 percent of interviewers never worked for the GSOEP. This may be explained by the specifics of the target population and requirements of language knowledge relevant for the IAB-BAMF-SOEP Survey of Refugees in Germany (Kühne, Jacobsen, and Kroh 2019).

3.2.3 Control variables. We control for interviewer's *gender* (female, male), *age* (continuous variable, centered around the grand mean), and *educational level* (secondary education, intermediate education, upper secondary education, graduate degree) as explanatory variables at the interviewer level (Hox 1994; van Tilburg 1998; Loosveldt and Beullens 2013a). Thirty-six percent of the interviewers are female, the average age is approximately forty-nine years, and more than half of the interviewers have a graduate degree (Appendix table A1).

Regarding the respondents, we control for *gender* (female, male), *age* (continuous variable, quadratic specification, centered around the grand mean), *education level* (no degree/in school, secondary education, upper secondary education, college/university), and *accommodation* (reception center, private), since those are usually believed to affect data quality (Hox 1994; Flores-Macias and Lawson 2008). For similar reasons, we include the *country of origin* fixed effect and control for *interview language* (German/English, German/Arabic, German/Farsi, German/Pashto, German/Urdu, and German/Kurmanji). Prior research suggests that the language version of a questionnaire may

explain differences in interview duration between respondents from different linguistic regions (Loosveldt and Beullens 2013a; Jacobsen 2018).

Of course, the model should be very parsimonious in terms of including a variety of respondent-level variables. The measurement of a multitude of respondent characteristics like employment status and history, and information about family composition, is dependent on the actual interviewer–respondent performance that is under study. Thus, the availability and the quality of these variables (e.g. items related to language course participation) are conditional on the quality of the responses to the preceding filter questions. Since most of the socio-demographic and socio-economic variables are captured by exactly those filter questions, it is not advisable to rely on these data in the analysis; therefore, they are excluded here.

To account for potential regional differences, we include region fixed effects (38 *Regierungsbezirke*) (van Tilburg 1998) and control for the municipality size (0–1, 999, 2,000–4,999, 5,000–19,999, 20,000–49,999, 50,000–99,999, 100,000–499,999, and 500,000+). It should be note that regional differences of respondents in comparison to interviewer effects are assumed to be small owing to the national dispersal policies in Germany related to the asylum-seekers exogenous regional assignment (German Federal Office for Migration and Refugees (BAMF) 2018; Grote 2018).

Respondents' willingness to co-operate may also affect interviewers' incentives to linger over difficult cases until later in the field period (Kirchner and Olson 2017). To approximate the difficulty of households, we control for the number of household contacts by the interviewer and the number of interviewers who contacted each household. Appendix table A1 provides descriptive statistics for all variables used in the analysis.

3.3 Modeling Approach

To address the research hypotheses, a multilevel modeling framework is used. As with all observational and nonexperimental studies, it is challenging to entirely separate out the possible confounding effects of unobserved heterogeneity. However, a multilevel approach accounts for unobserved heterogeneity in the model and, thus, it is able to alleviate these concerns to some extent. The model explicitly accounts for the fact that standard errors and variances of estimates are underestimated when model residuals are not independent.

Since all dependent variables are treated as metric, we employed linear mixed-effects models. The model is formulated using matrix notation for simplicity:

$$y_{r,i} = \alpha + \mathbf{X}_r\boldsymbol{\beta}_1 + \mathbf{X}_i\boldsymbol{\beta}_2 + u_i + \varepsilon_{r,i} \quad (3)$$

The subscript notation follows the nesting logic of the data: respondents (r) are nested within interviewers (i). The estimation procedure yields coefficients for

the independent variables and a constant term (α) reflecting the grand mean (fixed parameters). Vector $\boldsymbol{\beta}_1$ contains coefficients of respondent-level variables ($\mathbf{X}_r$), which includes interviewer survey experience. Vector $\boldsymbol{\beta}_2$ denotes coefficients of interviewer-level variables ($\mathbf{X}_i$), such as general experience. The random-intercept model includes random effects for interviewers (u_i). The variance components for each of the slope parameters are fixed at zero and do not vary across levels by assumption. The random effects are assumed to be independent random variables following a normal distribution on each level with zero mean and a variance to be estimated from the sample. Finally, a residual error term at the level of measurement is denoted by $\varepsilon_{r,i}$.

The model is estimated in a stepwise fashion. First, only control variables and the sequence order of interviews within interviewers are included (Model 1), analyzing if and to what extent interview duration decreases with interviewers' survey experience (*H1a).* Second, the indicator for general interviewer experience is introduced in the model (Model 2) to assess whether general experience affects interview duration (*H2a*). The final model (Model 3) includes an interaction term between interviewers' survey experience and general experience and is used to test whether increasing survey (field) experience affects interview duration more negatively among generally inexperienced interviewers (*H3a*). These steps are then repeated for the triggering rate and response differentiation.

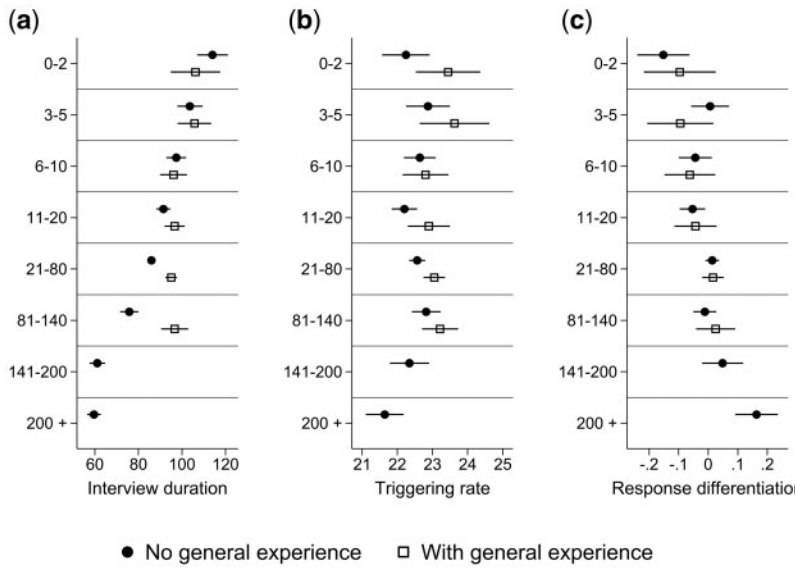

Figure 1. Means of Interview Duration, Triggering Rate, and Response Differentiation, by Interviewers' Survey Experience and General Experience.

4. RESULTS

4.1 Descriptive Analyses

Do experienced and inexperienced interviewers differ with regard to their interview duration over the course of the survey field period? Figure 1a–c shows the mean interview duration, triggering rate, and response differentiation by interviewers' survey experience and general experience. Overall, generally experienced interviewers have a ten-minute-longer mean interview duration than generally inexperienced ones (96.7 versus 86.7 minutes). Furthermore, as the number of interviews per interviewer increases, mean interview duration decreases for both experienced and inexperienced interviewers; however, this change is more pronounced for inexperienced interviewers. It should be noted that the last two categories provide little information as only three inexperienced interviewers conducted more than 140 interviewers.

Generally inexperienced interviewers tend to have a slightly lower rate of triggered filter questions than generally experienced interviewers (22.5 versus 23.1 percent). In addition, no systematic change in the triggering rates emerges throughout the field period for both groups of interviewers. Contrary to the results for interview duration, we do not find differences in response differentiation between generally experienced and inexperienced interviewers (-0.009 versus -0.005 units). Notably, response differentiation seems to increase with survey experience for generally experienced interviewers, whereas no such pattern emerges for generally inexperienced interviewers.

These descriptive findings suggest that the decreasing interview durations observed in the survey are mainly driven by interviewers who increase their interviewing efficiency, rather than interviewers who exhibit deviant behavior during the interview. However, this analysis does not consider interviewer-specific effects or other potential confounders of the relationship between the dependent variables and both types of experience. Therefore, in the next section, we implement the multilevel models and control for relevant interviewer- and respondent-level variables to formally test our hypotheses.

4.2 Multilevel Results—Interview Duration

Models 1.1–1.3 in table 1 display the results of the multilevel regression models of interview duration on survey and general experience. All models include the control variables described in Section 2.2.

Model 1.1 indicates that interviewer survey experience has a statistically significant impact on interview duration. For example, after holding constant all of the other predictors in the model, the fiftieth interview is $\sim$25.2 minutes $\big(= \ln(50) * (-6.4)\big)$ shorter than the first interview. As shown in Model 1.2, general interviewer experience has no statistically significant effect on interview duration. Hence, the sizeable gap between generally experienced and

Table 1. Hierarchical Linear Random Intercept Models Estimated Coefficients, Standard Errors, and Variance Components

	Model 1.1 Interview duration	Model 1.2 Interview duration	Model 1.3 Interview duration	Model 2.1 Triggering rate	Model 2.2 Triggering rate	Model 2.3 Triggering rate	Model 3.1 Response differentiation	Model 3.2 Response differentiation	Model 3.3 Response differentiation
ln(Survey experience)	−6.427***	−6.424***	−3.647***	−0.032	−0.031	−0.123	0.022**	0.023**	0.031*
	(0.504)	(0.504)	(0.931)	(0.059)	(0.059)	(0.110)	(0.007)	(0.007)	(0.014)
General experience: No		0.651	11.549		0.025	−0.351		0.066	0.099
		(6.083)	(6.786)		(0.353)	(0.519)		(0.049)	(0.068)
× ln(Survey experience)			−3.915***			0.128			−0.012
			(1.104)			(0.130)			(0.016)
Intercept	111.614***	111.105***	105.506***	20.088***	20.065***	20.312***	−0.039	−0.099	−0.121
	(12.612)	(13.478)	(13.517)	(1.067)	(1.116)	(1.143)	(0.139)	(0.145)	(0.148)
(Respondent and interviewer level controls omitted see table S3 of the online Supplementary Material)									
Between interviewer var.	428.308***	428.385***	423.202***	0.995	0.995	0.995	0.022***	0.022***	0.022***
	(71.222)	(71.222)	(70.876)	(0.245)	(0.245)	(0.245)	(0.005)	(0.005)	(0.005)
Within interviewer var.	745.278***	745.273***	743.274***	11.595***	11.595***	11.592***	0.173***	0.173***	0.173***
	(16.140)	(16.139)	(16.099)	(0.252)	(0.252)	(0.252)	(0.004)	(0.004)	(0.004)
LR test	1,018	994	964	84	84	84	150	145	146
Prob ≥ chibar2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ICC	36.50	36.50	36.28	7.90	7.90	7.91	11.30	11.05	11.07
Log Likelihood	−20,774	−20,774	−20,768	−11,611	−11,611	−11,610	−2,445	−2,444	−2,444
Dfm	97	98	99	97	98	99	97	98	99
Sample size	4,367	4,367	4,367	4,367	4,367	4,367	4,367	4,367	4,367

NOTE.—Significance level ***$p < .001$, **$p < .01$, *$p < .05$. Standard errors in parentheses. For the full list of model covariates refer to Section 2.2.

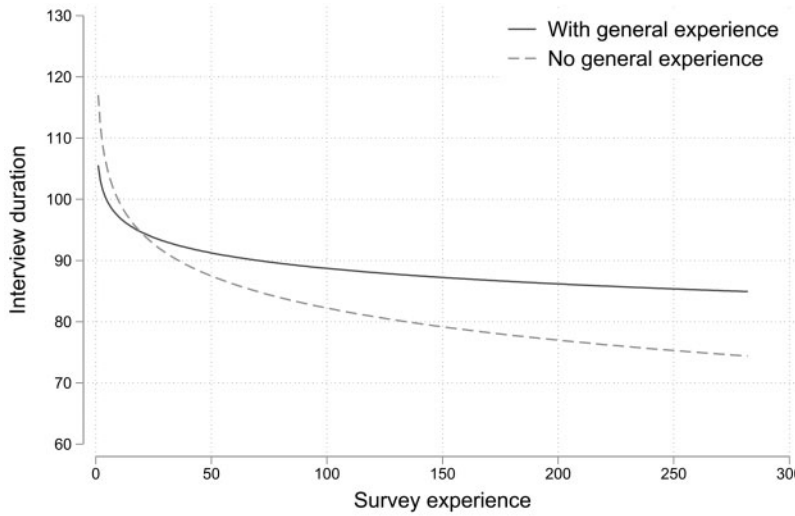

Figure 2. Predicted Conditional Interview Duration by Interviewers' Survey and General Experience (Based on Model 1.3 in table 1).

inexperienced interviewers observed in the descriptive analysis does not occur in the multilevel analysis after accounting for potential confounders.

Inclusion of the interaction effects between survey and general experience of interviewers in Model 1.3 improves model fit significantly (χ^2 of likelihood-ratio test: 12.5), indicating that the effect of survey experience varies by general experience of interviewers. There is no statistically significant difference between generally experienced and inexperienced interviewers in their first interviews. However, the significant negative interaction effect suggests that the interview duration decreases with every additional interview and it does so more steeply for the generally inexperienced interviewers. That is, all other predictors held constant, the fiftieth interview of the generally experienced interviewers is 14.3 minutes $(= \ln(50)*(-3.7))$ shorter than their first interview, whereas it is 29.5 minutes $\left(= \ln(50)*(-3.7 + (-3.9))\right)$ shorter for the generally inexperienced interviewers. This interaction effect is shown in figure 2.

It should be noted that the intra-interviewer correlation coefficient (Hox 1994), which denotes the proportion of variance in interview duration explained by the interviewers, varies between 36.3 and 36.5 percent. As we included fixed effects for region and accounted for municipality size and respondent characteristics, these values are unlikely to be driven by other factors. Hence, interviewers seem to have a substantial influence on interview duration. This result is in line with previous studies that found similarly large ICCs for interview duration (Loosveldt and Beullens 2013a; Kirchner and Olson 2017; West, Conrad, Kreuter, and Mittereder 2018).

4.3 Multilevel Results—Triggering Rate

Models 2.1–2.3 summarized in table 1 provide the estimation results for the triggering rate. Model 2.1 includes the natural logarithm of interviewers' survey experience and the control variables. We do not find a statistically significant effect of survey experience on the triggering rate, nor do we find a statistically significant effect of general experience on the triggering rate (Model 2.2). Inclusion of the interaction effect between both types of experience in Model 2.3 does not improve model fit ($\chi^2$of likelihood-ratio test: 1.0) and the corresponding coefficient is not statistically significant. Hence, neither survey experience nor general experience affects the triggering rate.

The ICCs vary around 7.9 percent, which is smaller than in other studies (van Tilburg 1998; Marsden 2003; Brüderl, Huyer-May, and Schmiedeberg 2013; Josten and Trappmann 2016), suggesting that interviewers generally do have an effect on the triggering rate. These interviewer effects are considerably smaller than the effects of interviewers on interview duration. This pattern is consistent with the study of Vandenplas et al. (2018), who argued that such differences could be the result of the more direct effect of interviewers on the interviewing process (e.g. speed of reading questions and engaging in side-talk) and less direct effect of interviewers on the quality of survey responses, which can also be influenced by respondent factors (e.g. motivation and cognitive burden).

4.4 Multilevel Results—Response Differentiation

The estimation results for response differentiation are presented in Models 3.1–3.3 (table 1). The effect of survey experience in Model 3.1 is positive and statistically significant, though the effect is negligible in size: holding everything else constant, response differentiation increases on average by 0.1 units after hundred interviews, which accounts for approximately one-fifth of a standard deviation. Model 3.2 indicates that general interviewer experience is not associated with response differentiation, whereas the survey experience effect remains robust. The interaction effect between survey and general experience in Model 3.3 is not statistically significant and indicates that the previously observed effect of survey experience does not differ between generally experienced and inexperienced interviewers.

As with the triggering rate, the estimated ICCs are smaller than the ICCs estimated for interview duration. Interviewers account for about 11 percent of the variance in response differentiation, which is similar in size to other studies (Loosveldt and Beullens 2017).

4.5 Robustness Checks

To test the robustness of our results to different model specifications, we performed a series of robustness checks on our benchmark models (table 1). In

the following paragraphs, we briefly discuss our findings; the results are presented in online Supplementary Material.

First, independent of their (general and survey) experience, interviewers often process easier cases (with co-operative respondents) at the beginning of the field period and linger over rather difficult cases for the remainder of the field period (Kirchner and Olson 2017). This mechanism would explain our findings on no differences between generally experienced and inexperienced interviewers at the beginning of the field period (figure 1). We have argued that speeding up the interviews over time could be attributed to survey experience. Another explanation could be, however, that interviewers change their interviewing tactics and tend to collect only key survey information from reluctant respondents. Shifting more difficult cases to a later time point in the field period presupposes interviewers' knowledge on case difficulty. At the interviewer's first household contact, such knowledge is naturally lacking. Correspondingly, if interviewers tend to postpone difficult cases, we would expect (1) no survey experience effect for the respondents interviewed at the first contact and (2) a survey experience effect for the respondents interviewed at later contacts because the interviewers may know whether it is a difficult case or not. To examine this mechanism, we replicated our analyses including an interaction effect between the (general and survey) experience of interviewers and an indicator for interviews conducted at the first household contact (equal to 1; and 0 otherwise). The interaction term turned out to be insignificant (in size and statistically) in either specification (table S5 of the online Supplementary Material).

Second, linguistic proximity between interviewers' and respondents might contribute to more efficient interviewing processes and, hence, shorter interviews. At the same time, general interviewer experience likely correlates with interviewers' migration background: to accommodate the target population, the survey organization recruited new interviewers with migration backgrounds similar to the pool of respondents (KANTAR 2020). Unfortunately, the information on the interviewers' country of origin and language skills is not available for the first wave of the survey. Following our request, the survey organization provided information on the mother tongue and country of birth for interviewers who were still part of their interviewer staff. This was, however, the case for only forty-six out of ninety-five interviewers. However, to check whether the interviewer's language skills may confound the relationship between interviewer experience and the dependent outcomes, we replicated the analyses on the sample restricted to forty-six interviewers and controlled for *linguistic proximity* between interviewer–respondent pairs. Linguistic proximity was developed by ethnolinguists and ethnobiologists and measures the similarities between a fixed number of words in a pair of languages (Melitz and Toubal 2014). The corresponding sample restriction and model specification did not show any substantial differences from the original findings that are summarized in table 1 (table S6 of the online Supplementary Material).

Third, the survey experience of the interviewer could affect both interview duration and response differentiation (or the triggering rate), whereas response differentiation (or the triggering rate) itself may simultaneously affect the interview duration. In such a case, response differentiation or the triggering rate could be important omitted confounders between interviewer experience and interview duration. To address this, we replicated the benchmark model for interview duration (table 1, Models 1.1–1.3) including the triggering rate and response differentiation as additional controls. The results remained fairly robust (table S7 of the online Supplementary Material). Although both indicators are correlated with the interview duration, the survey experience effect does not change.

Fourth, the IAB-BAMF-SOEP survey is a household survey and thus some respondents are nested within households. However, more than two-thirds of the households are one-person-households (69.3 percent). Moreover, households are not perfectly clustered within interviewers as members of eight households were interviewed by more than one interviewer. Accordingly, this limitation undermines the possibility to incorporate household as an additional level into the multilevel model. However, to alleviate potential concerns related to the household effect, we replicate the main analyses with the household level as an additional level while excluding eighteen respondents who were nested within households in which two interviewers conducted interviews (table S8 of the online Supplementary Material). Our results change only marginally.

Fifth, previous research has pointed to changes in the sociodemographic composition of the sample over the course of the field period (Kirchner and Olson 2017). These changes could challenge our substantive conclusion regarding the survey experience effects on the triggering rate if changes in the sociodemographic composition of the sample are collinear with the substantive content of the filter questions. Consider, for example, if larger households are generally more likely to be interviewed earlier in the field period, and thus are more likely to trigger the filter questions on family composition. These could countervail the negative effect of survey experience on the triggering rate and drive the observed non-significant effect (table 1, Models 2.1–2.3). However, if changes in sociodemographic composition resulted in an increasing triggering rate, we would also expect longer interview durations later in the field period, which was not the case. Nonetheless, we address these concerns with further robustness checks. To examine whether the composition of respondents within the interviewer's pool changes over the field period (Kirchner and Olson 2017), we split each interviewer's workload in half to obtain two samples of early and late respondents and compared sociodemographic characteristics (table S9 of the online Supplementary Material). This analytical exercise revealed some notable differences between early and late interviews, for example, respondents' age, education, language version of the questionnaire, municipality size, and accommodation. Thus, we replicated our analyses (a) without

any respondent or interviewer control variables and (b) without only respondent control variables to see whether the effects of interest are strongly mediated by the set of the included respondent controls. If the respondent variables countervailed a negative survey experience effect, we would find a positive learning effect in the models without respondent variables. However, our substantive conclusions were not affected (table S10 of the online Supplementary Material). In addition, we use the split samples described above to further investigate potential countervailing effects of the sample composition. Although those effects might be prevalent in the full sample, changes in composition within the first or second half of each interviewer's workload that increase the triggering rate are less likely. Thus, we estimate the main model for the split samples separately (table S11 of the online Supplementary Material). The estimated coefficients are close to the results of the main specification (table 1, Model 2.3) and not statistically significant. Hence, countervailing effects of the sociodemographic composition are unlikely to drive our results for the triggering rate.

4.6 Multilevel Results—Summary

To assist the interpretation of the findings in light of the hypotheses, table 2 provides an overview of the hypotheses, expected findings dependent on the mechanism behind interviewers' survey experience effect, and actual results from the multilevel analyses.

In line with hypothesis *H1a*, we find a negative effect of interviewer survey experience on interview duration (Section 3.2 and table 2). This pattern implies a learning effect over the interviewers' field period. Second, we do not find empirical support for hypothesis *H2a* that generally experienced interviewers' prior knowledge of how to keep an interview short may shape their subsequent interviewing. Third, the negative effect of survey experience is stronger among generally inexperienced interviewers than experienced ones; thus, hypothesis *H3a* is empirically supported.

To grasp whether these results are driven by increasing efficiency or deviant behavior, we examined the effects of general and survey experience on further data quality indicators (triggering rate in Section 3.3 and response differentiation in Section 3.4). If these indicators were not affected by general or survey experience, we may conclude that the mechanism at work is increasing efficiency and not deviant behavior. Following our results, we find no empirical evidence for hypotheses *H1b-H3b* and *H1c-H3c*. Neither survey experience nor the interviewers' general experience has a negative effect on the triggering rate or response differentiation. Altogether, our findings (table 2, column four) indicate that the mechanism causing the decrease in interview duration observed earlier is rather owing to increasing efficiency (table 2, column two) than deviant behavior (table 2, column three).

Table 2. Summary of Hypotheses and Findings

	Expected mechanism behind interviewers' survey experience effect		
Hypotheses	Increased efficiency	Deviant behavior	Results
H1a: Interview duration decreases with growing survey experience	✓	✓	✓
H2a: Interview duration is shorter for generally experienced interviewers compared to generally inexperienced interviewers	✓	✓	x
H3a: Interview duration decreases with growing survey experience more strongly for generally inexperienced interviewers than for generally experienced interviewers	✓	✓	✓
H1b: Triggering rate decreases with growing survey experience	x	✓	x
H2b: Triggering rate is lower for generally experienced interviewers compared to generally inexperienced interviewers	x	✓	x
H3b: Triggering rate decreases with growing survey experience more strongly for generally inexperienced interviewers than for generally experienced interviewers	x	✓	x
H1c: Response differentiation decreases with growing survey experience	x	✓	x
H2c: Response differentiation is lower for generally experienced interviewers compared to generally inexperienced interviewers	x	✓	x
H3c: Response differentiation decreases with growing survey experience more strongly for generally inexperienced interviewers than for generally experienced interviewers	x	✓	x

5. DISCUSSION

Relying on survey data and multilevel models, we analyzed interviewer effects on interview duration and focused on (1) interviewer survey experience, that is, experience gained during the field period for a given survey at hand and (2) interviewer general experience, that is, interviewers' prior experience with the survey organization. Both types of interviewer experience have been found to correlate with shorter interview duration (Olson and Peytchev 2007; Olson and Bilgen 2011; Loosveldt and Beullens 2013a, 2013b; Böhme and Stöhr 2014; Kirchner and Olson 2017). The previous literature is, however, less clear on mechanisms behind this correlation: changes in interview duration over interviewers' experience may be influenced by interviewers' growing positive efficiency or their detrimental rational behavior to keep an interview short. Assuming that misbehavior would prompt growing effort to cut down on asking follow-up questions and inadequately speed through interviews, we additionally looked at the development of the triggering rate to filter questions and response differentiation by interviewers' survey and general experience. Correspondingly, our study enriches the existing literature on interviewer effects and interviewer experience (Olson and Peytchev 2007; Olson and Bilgen 2011; Loosveldt and Beullens 2013a; Böhme and Stöhr 2014; Kirchner and Olson 2017), by disentangling the mechanisms driving accelerated interviewing observed over the field period.

Our results showed a negative effect of survey experience on interview duration. Although we do not find a difference between generally experienced and inexperienced interviewers at the beginning of the field period, both types of interviewers do speed up their interviews as they gain more survey experience. Moreover, there seems to be a stronger negative effect of survey experience for inexperienced interviewers than for their generally experienced counterparts. These patterns follow our theoretical predictions that generally experienced interviewers would be less inclined to change their behavior over the fieldwork period because of their greater prior knowledge.

Either mechanism: – increased interviewer efficiency or introduction of detrimental interviewing behavior – could be responsible for these results. To investigate the underlying mechanism, we replicated the analytical exercise on interview duration for the triggering rate and response differentiation. The results imply that neither general nor survey experience negatively affects the triggering rate and response differentiation. What can we infer from these patterns for the mechanism behind the effect of interviewers' experience on interview duration? In our view, these corresponding patterns reflect rather a learning effect owing to improving interviewing efficiency and not owing to learning about and implementing shortcutting possibilities or inadequately speeding through the interview.

By employing multilevel models, we accounted for unobserved heterogeneity between interviewers and respondents under typical assumptions implicit to the multilevel modeling framework. A major strength of this approach is that it sheds light on the role of interviewer experience and its potential effect on data

quality in a real-world survey context and, hence, on the relevance of the problem overall. However, this study analyzed interviewer effects in a German survey context for a specific population interviewed. Although the IAB-BAMF-SOEP Survey of Refugees is appropriate for studying interviewer effects, given its vulnerable and difficult-to-interview study population, using different study populations would be advantageous for assessing the generalizability of these findings and further advancing the study of interviewer behavior on data quality.

6. CONCLUSION

From a survey practice perspective, the results suggest that using solely interview duration as a proxy for potential deviant interviewer behavior risks being a short-sighted view. Only in combination with the triggering rate (a potential instrument for shortcutting interviews) and response differentiation (indicative of inadequate speeding), we were able to come closer to the mechanism behind the effect of interviewer experience. Although previous literature suggests that shorter interview durations are indicative of deviant behavior (Bushery et al. 1999; Murphy et al. 2016), we found that decreasing durations over the field period are not associated with lower data quality. Future research might broaden this view by using more differentiated measures of interviewers' experience and recording interviews and coding interviewer behavior to gain further insights.

Supplementary Materials

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

Appendix

Table A.1. Descriptive Statistics for Dependent, Independent, and Control Variables

	Mean (SD)/ Proportion	Min	Max	Sample size
Dependent variables				
Interview duration	89.557 (37.680)	7	248	4,367
Triggering rate	22.662 (4.029)	11.931	42.516	4,367
Response differentiation	−0.006 (0.459)	−2.254	2.116	4,367

Continued

Table A.1. *Continued*

	Mean (SD)/ Proportion	Min	Max	Sample size
Independent variables				
Survey experience	49.691 (52.507)	1	282	4,367
General Experience: No	0.747	0	1	95
Control variables at the respondent level				
Respondent gender (1 = female)	0.379	0	1	4,367
Respondent age	33.574 (10.386)			4,365
Respondent education:				
No degree/in school	0.420	0	1	4,027
Secondary education	0.208	0	1	4,027
Upper secondary education	0.186	0	1	4,027
College/university	0.185	0	1	4,027
Interview language:				
German/English	0.163	0	1	4,367
German/Arabic	0.651	0	1	4,367
German/Farsi	0.125	0	1	4,367
German/Pashtu	0.010	0	1	4,367
German/Urdu	0.017	0	1	4,367
German/Kurmanji	0.034	0	1	4,367
Accommodation				
Reception center	0.325	0	1	4,367
Private	0.671	0	1	4,367
No Information	0.044	0	1	4,367
Municipality size				
−1,999	0.020	0	1	4,367
2,000–4,999	0.068	0	1	4,367
5,000–19,999	0.245	0	1	4,367
20,000–49,999	0.220	0	1	4,367
50,000–99,999	0.139	0	1	4,367
100,000–499,999	0.161	0	1	4,367
500,000+	0.148	0	1	4,367
Number of HH contacts				
1	0.151	0	1	4,367
2	0.274	0	1	4,367
3	0.259	0	1	4,367
4	0.154	0	1	4,367
5	0.078	0	1	4,367
6	0.037	0	1	4,367
7	0.024	0	1	4,367
8	0.007	0	1	4,367

Continued

Table A.1. *Continued*

	Mean (SD)/ Proportion	Min	Max	Sample size
9	0.016	0	1	4,367
Number of different interviewers				
1	0.939	0	1	4,367
2	0.061	0	1	4,367
3	0.005	0	1	4,367
Control variables at the interviewer level				
Interviewer gender (1 = female)	0.360	0	1	89
Interviewer age	48.853 (15.730)			89
Interviewer education:				
Secondary education	0.086	0	1	81
Intermediate education	0.148	0	1	81
Upper secondary education	0.235	0	1	81
Graduate degree	0.531	0	1	81

NOTE.—Variation in the sample size (column 3) is owing to differences in missing data across variables. In the multivariate model, missing values are controlled for by including a dummy variable for a corresponding variable. SD, Standard deviation.

REFERENCES

AAPOR (2003), "Interviewer Falsification in Survey Research: Current Best Methods for Prevention, Detection, and Repair of its Effects," available at https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/falsification.pdf.

The American Association for Public Opinion Research (AAPOR) (2016), Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition, AAPOR, available at https://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx

Bilgen, I. (2011), "Is Less More & More Less. . .? The Effect of Two Types of Interviewer Experience on 'Don't Know' Responses in Calendar and Standardized Interviews," *Survey Research and Methodology Program (SRAM)*, *Dissertations & Theses*, 4.

Bloch, A. (1999), "Carrying out a Survey of Refugees: Some Methodological Considerations and Guidelines," *Journal of Refugee Studies*, 12, 367–383.

———. (2007), "Methodological Challenges for National and Multi-Sited Comparative Survey Research," *Journal of Refugee Studies*, 20, 230–247.

Böhme, M., and T. Stöhr (2014), "Household Interview Duration Analysis in CAPI Survey Management," *Field Methods*, 26, 390–405.

Brücker, H., N. Rother, and J. Schupp (eds.) (2017), "*IAB-BAMF-SOEP-Befragung Von Geflüchteten 2016: Studiendesign, Feldergebnisse Sowie Analysen zu Schulischer Wie Beruf l Icher Qualifikation, Sprachkenntnissen Sowie Kognitiven Potenzialen*," *IAB-Forschungsbericht*, Berlin: DIW Berlin, German Institute for Economic Research.

Brüderl, J., B. Huyer-May, and C. Schmiedeberg (2013). "Interviewer Behavior and the Quality of Social Network Data," in *Interviewers' Deviations in Surveys. Impact, Reasons, Detection and Prevention*, eds. P. Winker, R. Porst, and N. Menold, pp. 147–160, Frankfurt: Peter Lang.

Bushery, J. M., J. W. Reichert, K. A. Albright, and J. C. Rossiter (1999). "Using Date and Time Stamps to Detect Interviewer Falsification," Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 316–320.

Cannell, C. F., P. V. Miller, and L. Oksenberg (1981). "Research on Interviewing Techniques," *Sociological Methodology*, 12, 389–437.

Chromy, J., A. Hughes, K. Giacoletti, and D. Odom (2002). "Impact of Interviewer Experience on Respondent Reports of Substance Use," in *Redesigning an Ongoing National Household Survey: Methodological Issues*, eds. J. Groerer, J. Eyerman, and J. Chromy, pp. 161–184, Washington, DC: Substance Abuse and Mental Health Services Administration.

Edwards, B., A. Maitland, and S. Connor (2017). "Measurement Error in Survey Operations Management," in *Total Survey Error in Practice*, eds. P. P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, and B. T. West, pp. 253–277, Hoboken, NJ: John Wiley & Sons, Inc..

Edwards, B., H. Sun, and R. Hubbard (2020). "Behavior Change Techniques for Reducing Interviewer Contributions to Total Survey Error," in *Interviewer Effects from a Total Survey Error Perspective*, eds. K. Olson, J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, and B. T. West, pp. 77–90, New York: Chapman and Hall.

Flores-Macias, F., and C. Lawson (2008). "Effects of Interviewer Gender on Survey Responses: Findings from a Household Survey in Mexico," *International Journal of Public Opinion Research*, 20, 100–110.

Fowler, F., and T. Mangione (1990). *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*, Thousand Oaks, CA: SAGE Publications, Inc.

German Federal Office for Migration and Refugees (BAMF) (2018). "Initial Distribution of Asylum-Seekers (EASY)," available at https://www.bamf.de/EN/Themen/AsylFluechtlingsschutz/AblaufAsylverfahrens/Erstverteilung/erstverteilung-node.html (accessed January 24, 2020).

Grote, J. (2018). "The Changing Influx of Asylum Seekers in 2014-2016: Responses in Germany: Focussed study by Germany Contact Point for the European Migration Network (EMN)," *Working Paper79, BAMF-EMN, Nuremberg*.

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

Haller, H. (1985). "The Principal-Agent Problem with a Satisficing Agent," *Journal of Economic Behavior and Organization*, 6, 359–379.

Herzog, A. R., and J. G. Bachman (1981). "Effects of Questionnaire Length on Response Quality," *Public Opinion Quarterly*, 45, 549.

Hicks, W. D., B. Edwards, K. Tourangeau, B. McBride, L. D. Harris-Kojetin, and A. J. Moss (2010). "Using CARI Tools to Understand Measurement Error," *Public Opinion Quarterly*, 74, 985–1003.

Hood, C. C., and J. M. Bushery (1997). "Getting More Bang from the Reinterview Buck: Identifying 'At Risk' Interviewers," Proceedings of the American Statistical Association (Section on Survey Research Methods), American Statistical Association, pp. 820–824.

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

Jacobsen, J. (2018). "Language Barriers during the Fieldwork of the IAB-BAMF-SOEP Survey of Refugees in Germany," in *Surveying the Migrant Population: Consideration of Linguistic and Cultural Issues*, ed. D. Behr, pp. 75–84, Köln: GESIS - Leibniz-Institut für Sozialwissenschaften.

Japec, L. (2008). "Interviewer Error and Interviewer Burden," in *Advances in Telephone Survey Methodology*, eds. C. Tucker, J. M. Brick, E. de Leeuw, L. Japec, P. J. Lavrakas, and M. W. Link, pp. 187–211, Hoboken, NJ: John Wiley & Sons.

Josten, M., and M. Trappmann (2016). "Interviewer Effects on a Network-Size Filter Question," *Journal of Official Statistics*, 32, 349–373.

KANTAR (2020). "Personal and e-mail communication with the Public Division of Kantar in Germany (July, 2020)."

Kemper, C. J., and N. Menold (2014). "Nuisance or Remedy? The Utility of Stylistic Responding as an Indicator of Data Fabrication in Surveys," *Methodology*, 10, 92–99.

Kirchner, A., and K. Olson (2017). "Examining Changes of Interview Length over the Course of the Field Period," *Journal of Survey Statistics and Methodology*, 5, 84–108.

Kosyakova, Y., J. Skopek, and S. Eckman (2015). "Do Interviewers Manipulate Responses to Filter Questions? Evidence from a Multilevel Approach," *International Journal of Public Opinion Research*, 27, 417–431.

Kroh, M., S. Kühne, J. Jacobsen, M. Siegert, and R. Siegers (2017). "Sampling, Nonresponse, and Integrated Weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4)," *SOEP Survey Papers 477. Berlin: DIW.*

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

Kühne, S., J. Jacobsen, and M. Kroh (2019). "Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees," *Survey Methods: Insights from the Field*.

Loosveldt, G. (2008). "Face-to-Face Interviews," in *The International Handbook of Survey Methodology*, eds. E. D. de Leeuw, J. J. Hox, and D. A. Dillman, pp. 201–220, New York: Taylor & Francis.

Loosveldt, G., and K. Beullens (2013a). "How Long Will It Take?' an Analysis of Interview Length in the Fifth Round of the European Social Survey," *Survey Research Methods*, 7, 69–78.

———. (2013b). "The Impact of Respondents and Interviewers on Interview Speed in Face-to-Face Interviews," *Social Science Research*, 42, 1422–1430.

———. (2017). "Interviewer Effects on Non-Differentiation and Straightlining in the European Social Survey," *Journal of Official Statistics*, 33, 409–426.

Marsden, P. V. (2003). "Interviewer Effects in Measuring Network Size Using a Single Name Generator," *Social Networks*, 25, 1–16.

Matschinger, H., S. Bernert, and M. C. Angermeyer (2005). "An Analysis of Interviewer Effects on Screening Questions in a Computer Assisted Personal Mental Health Interview," *Journal of Official Statistics*, 21, 657–674.

Melitz, J., and F. Toubal (2014). "Native Language, Spoken Language, Translation and Trade," *Journal of International Economics*, 93, 351–363.

Menold, N., and C. J. Kemper (2014). "How Do Real and Falsified Data Differ? Psychology of Survey Response as a Source of Falsification Indicators in Face-to-Face Surveys," *International Journal of Public Opinion Research*, 26, 41–65.

Menold, N., P. Winker, N. Storfinger, and C. J. Kemper (2013). "A Method for Ex-Post Identification of Falsifications in Survey Data," in *Interviewers' Deviations in Surveys—Impact, Reasons, Detection and Prevention*, eds. P. Winker, N. Menold, and R. Porst, pp. 25–47, Frankfurt am Main: Peter Lang, Academic Research.

Murphy, J., P. Biemer, C. Stringer, R. Thissen, O. Day, and Y. P. Hsieh (2016). "Interviewer Falsification: Current and Best Practices for Prevention, Detection, and Mitigation," *Statistical Journal of the IAOS*, 32, 313–326.

Olson, K., and I. Bilgen (2011). "The Role of Interviewer Experience on Acquiescence," *Public Opinion Quarterly*, 75, 99–114.

Olson, K., and A. Peytchev (2007). "Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes," *Public Opinion Quarterly*, 71, 273–286.

Olson, K., and J. Smyth (2019). "What do Interviewers Learn? An Examination of Interview Length and Interviewer Behaviors," in *Interviewers and Their Effects from a Total Survey Error Perspective Workshop*, University of Nebraska-Lincoln.

Ruckdeschel, K., L. Sauer, and R. Naderi (2016). "Reliability of Retrospective Event Histories within the German Generations and Gender Survey: The Role of Interviewer and Survey Design Factors," *Demographic Research*, 34, 321–358.

Schreiner, I., K. Pennie, and J. Newbrough (1988). "Interviewer Falsification in Census Bureau Surveys," Proceedings of the American Statistical Association (Survey Research Methods Section), pp. 491–496.

van Tilburg, T. (1998). "Interviewer Effects in the Measurement of Personal Network Size: A Nonexperimental Study," *Sociological Methods & Research*, 26, 300–328.

Vandenplas, C., G. Loosveldt, K. Beullens, and K. Denies (2018). "Are Interviewer Effects on Interview Speed Related to Interviewer Effects on Straight-Lining Tendency in the European Social Survey? An Interviewer-Related Analysis," *Journal of Survey Statistics and Methodology*, 6, 516–538.

West, B. T., F. G. Conrad, F. Kreuter, and F. Mittereder (2018). "Can Conversational Interviewing Improve Survey Response Quality without Increasing Interviewer Effects?," *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 181, 181–203.

Yan, T. (2008). "Nondifferentiation," in *Encyclopedia of Survey Research Methods*, ed. P. J. Lavrakas, pp. 520–521, Thousand Oaks, CA: SAGE Publications, Inc..

van der Zouwen, J., and W. Dijkstra (1988). "Types of Inadequate Interviewer Behaviour in Survey Interviews: Their Causes and Effects," *Bulletin de Méthodologie Sociologique*, 18, 5–20.