

Adjusting to the Survey: How Interviewer Experience Relates to Interview Duration

André Pirralha, Christian Haag & Jutta von Maurice
Center for Study Management (CSM), Leibniz Institute for Educational Trajectories (LifBi)

Abstract

Interview duration has been shown to become shorter as fieldwork progresses. This has been attributed to a learning effect interviewers go through as they gather experience. In this study, we expand on this knowledge by focusing on how two kinds of interviewer experience relate to interview duration in telephone surveys. Using data from the German National Educational Panel Study (NEPS), we employ multilevel models, accounting for the clustering of respondents within interviewers. The results strengthen previous findings associating within-survey interviewer experience with decreasing interview duration. On the other hand, countering previous work, we find evidence that interview duration also decreases with overall interviewer experience. The results add to our knowledge concerning the effect of interviewer experience in the telephone survey mode. The effects are robust to several model specifications and to different interviewer, respondent, and interview characteristics. We conclude with a discussion about how to manage interviewer experience during training and fieldwork.

Keywords: Interview Duration; Interviewer experience; CATI surveys; panel data; para-data



© The Author(s) 2023. This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Interviewers are important actors in the collection of standardized data in survey studies. For example, they set the pace of an interview, elicit respondent cooperation, or help and warn respondents of the cognitive effort they should put into their responses (Ackermann-Piek & Massing, 2014; Olson, Smyth, Dykema, et al., 2020b; West & Blom, 2017). Interview duration is a widely used and easy to measure indicator to monitor and evaluate fieldwork as it allows to detect deviations from the standardized interview protocol and because of its impact in determining survey costs in interviewer-administered studies (Jin et al., 2019; Lepkowski et al., 2010; Vandenplas et al., 2019).

Ever since the seminal work of Olson and Peytchev (2007), it is well-known that the time spent on administering a survey interview differs extensively not only between interviewers but also throughout the fieldwork phase, with a clear tendency for interviews to become shorter (Böhme & Stöhr, 2014; Kirchner & Olson, 2017; Kosyakova et al., 2021; Loosveldt & Beullens, 2013b). This pattern has been associated with a learning effect interviewers take on when conducting interviews within the same study. There is also evidence that overall interviewer experience is important, with more and less experienced interviewers varying in how they follow survey protocols (Fowler & Mangione, 1990; Kirchner & Olson, 2017). While there is extended empirical evidence showing that interviewers in face-to-face survey modes (Computer-Assisted Personal Interview – CAPI) tend to become faster as fieldwork progresses (Kosyakova et al., 2021; Loosveldt & Beullens, 2013b; Vandenplas et al., 2018), there are only very few attempts directed to uncover whether this same effect also holds in telephone surveys (Computer-Assisted Telephone Interview – CATI). In particular, the most relevant publications focusing on CATI surveys are Kirchner and Olson (2017) as well as Olson and Smyth (2015, 2020), which support the finding that interview duration decreases throughout fieldwork and that within-survey interviewer experience is associated with shorter interviews. Nonetheless, these papers use data and paradata from the same study and for showing the robustness of these results other empirical analyses should be sought.

In this paper, we focus on the effect of interviewer experience on interview duration. We analyze how interviewer experience is associated with interview duration in a probability-based telephone survey of an educational panel study, with a sample of parents of children in school-age in Germany. Our contribution to the literature is an empirical examination of the relation between two aspects of interviewer experience and interview duration in a telephone survey, while accounting for an extended set of interviewer, respondent, and interview characteristics.

Direct correspondence to

André Pirralha, Leibniz-Institut für Bildungsverläufe, Zentrum für Studienmanagement, Wilhelmsplatz 3, 96047 Bamberg
E-mail: andre.pirralha@lifbi.de

We add to previous work by exploring telephone interview data from a large-scale panel survey, consisting of a substantial number of both respondents and randomly assigned interviewers with high variability of characteristics at both sides. We employ a multilevel modeling strategy aiming to disentangle how different aspects of interviewer experience are associated with interview duration and to account for the clustering of respondents and interviewers. Finally, we discuss some of the implications of our results for interviewer training and fieldwork management.

The Role of Interviewer Experience for Interview Duration

Interviewer effects are expected to differ between survey modes to some extent because “the mode or device for the interaction changes the nature of the interaction between interviewers and respondents” (Olson, Smyth, Dykema, et al., 2020b, p. 6). This should be the case with face-to-face and telephone interviews due to particular aspects of field control and payment scheme. In particular, CATI interviewers are subjected to more direct quality controls, as interviewing takes place in a centralized location and the survey agency actively supervises the outcome and the interviewers’ actions (Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V [ADM] et al., 2021; Kosyakova et al., 2021; Stiegler & Biedinger, 2015).

Even if interviewer effects may be less pronounced than in CAPI surveys, research has shown that around 25% of the variance in interview duration in CATI surveys is due to interviewers (Kirchner & Olson, 2017; Olson & Smyth, 2015). What is less clear, though, is how this can be explained. Previous studies argue that, as the fieldwork advances, interviewers progressively collect task-related experience and/or implement prior interviewing experience, thus speeding up and reducing interview time (Kirchner & Olson, 2017; Kosyakova et al., 2021). The research discusses a variety of possibilities that can explain this finding: learning effects can lead to an increase in interviewers’ reading speed; a reduction of misreading, corrections, or use of filler words; reduced time that has to be invested in reading additional interviewer instructions in computer-based modes; the reduction of (unnecessary) side communications; more efficient prompting behavior and subsequently quicker and/or clearer responses of the interviewee (Ackermann-Piek & Massing, 2014; Kosyakova et al., 2021; Olson, Smyth, Dykema, et al., 2020a).

A less benevolent view argues that shorter interviews throughout fieldwork are the result of behaviors that are grossly inconsistent with the interview protocol’s standardized practices, such as shortening of introduction texts, deviations from instructions concerning how often answer schemes are to be read, or accepting interviewees’ nonresponse too quickly (for instance in case of items that are sensitive or might lead to longer discussions). In a more extreme form, interviewers might

even skip items or avoid entire loops in the instrument by influencing answers in filter questions, they might phrase items liberally, or even suggest answers (Kosyakova et al., 2021; Olson & Smyth, 2020). This ‘misbehavior’ can be coupled with the interviewer’s payment scheme and the main interest of finishing the interview as quickly as possible (Vandenplas et al., 2018).

Regardless of what specific behavior accounts for this effect, previous research has shown that interviewer experience impacts interview duration, albeit to a different degree and depending on the mode of data collection (Vandenplas et al., 2019). Olson and Smyth (2015) demonstrate that telephone interviews become shorter as the fieldwork progresses. Olson and Peytchev (2007), in addition to finding no difference between CAPI and CATI modes of data collection, also observe that inexperienced interviewers speed up faster which might be related to a more pronounced learning effect. Loosveldt and Beullens (2013b, p. 1429) report that “interviewers strongly determine the interview speed” and they do so to a greater degree than the respondents. Substantial evidence from the literature shows that within-survey interviewer experience is positively associated with declining interview length (Kirchner & Olson, 2017; Kosyakova et al., 2021; Loosveldt & Beullens, 2013a; Olson & Smyth, 2015; Vandenplas et al., 2019). Kirchner and Olson (2017) show that the behavior of telephone interviewers, which is derived from experience and resulting in decreasing interview duration over the field phase, remains of influence regardless of changes in the composition of the sample. All this previous research underscores that interviewer experience remains an important factor in understanding interview duration, with other aspects potentially adding nuance to this relationship. In light of this well-established understanding, this study seeks to revisit and build upon these findings, providing both a confirmation and potential new insights especially for CATI mode. Interviewer experience is usually defined through two specific facets of survey interviewing: within-survey interviewer experience and overall interviewer experience (Kirchner & Olson, 2017; Kosyakova et al., 2021). Within-survey interviewer experience relates to the experience gathered by the interviewer in one specific survey or wave. On the other hand, overall interviewer experience is defined as being independent of a given study or wave and the result of the total time working as a survey interviewer. While probably both facets of interviewer experience operate through the mechanism of learning, they might differ in actual interviewer behavior. Whereas within-survey interviewer experience could be more closely linked to the growing knowledge concerning the specific instrument currently in the field, overall interviewer experience acknowledges the possibility that some interviewer behaviors can originate from professional knowledge or experience unrelated to the current study. Thus, for within-survey interviewer experience, we expect the interview duration to decrease as the interviewer gathers more experience by conducting more interviews in the same survey.

On the other hand, overall experienced interviewers have a greater level of knowledge and routine about how to conduct interviews, as well as on the protocols of the specific survey institute, thus they operate differently with the specificities of the study currently in the field (Kirchner & Olson, 2017; Kosyakova et al., 2021). Based on this discussion, we expect that more overall interviewer experience will lead, on average, to shorter interview duration when compared to less experienced interviewers. We formulate this expectation even though overall interviewer experience was not previously found to affect interview duration in telephone survey interviews (Kirchner & Olson, 2017; Olson & Peytchev, 2007).

Additionally, the possibility that both types of interviewer experience interact should also be considered. It can well be that interviewers with less overall experience will have a more pronounced learning curve as they gather more within-survey interviewer experience compared to more experienced interviewers which start working on a new survey study already with shorter interview durations (Kirchner & Olson, 2017; Kosyakova et al., 2021).

In sum, the differentiation between overall and within-survey interviewer experience may be crucial, and thus call for different adjustments in training, feedback, and supervision.

The Role of Other Factors on Interview Duration

There are other influences on interview duration besides interviewer experience (Kosyakova et al., 2021; Loosveldt & Beullens, 2013b; Vandenplas et al., 2018). An alternative explanation for decreasing interview duration over the field phase is based on the changing composition of the sample over time (Kirchner & Olson, 2017). As the fieldwork progresses, harder-to-reach respondents become more common and the lower cooperativeness of the remaining sample could be what makes the interview duration shorter. These respondents could also have a greater tendency toward satisficing behaviors, leading to shorter interviews (Krosnick, 1991). On the other hand, harder-to-reach respondents could struggle with the answers and this would lead to longer interviews (Jin et al., 2019).

Further respondent characteristics that have been discussed as potentially accounting for interview duration, albeit with inconsistent evidence, are age, education, employment status, and family and work time demands (Loosveldt & Beullens, 2013a; Timbrook et al., 2018; Vandenplas et al., 2018).

These personal characteristics (particularly sex, age, and education), when extended to the interviewer level, have also been considered as explanatory factors for interview duration, both on their own and paired with respondent characteristics (Kirchner & Olson, 2017; Kosyakova et al., 2021; Olson & Peytchev, 2007). Similarly, there is also some inconsistency in the findings regarding their effect, as some

research fails to find significant results (Sturgis et al., 2021), while other findings show that older and male interviewers are associated with longer interviews (Timbrook et al., 2018).

Finally, some specific characteristics of the interview itself can impact interview duration and therefore should also be controlled. For example, using a mobile phone is associated with significantly longer interviews when compared to landline connections, because mobile communication tends to be more prone to interruptions and takes longer on average (Timbrook et al., 2018). The type of telephone connection could also confound the association between interviewer experience and interview duration as the interview situation might vary substantially. Additional interview characteristics can have similar impacts such as conducting interviews on a weekday vs. the weekend; having interviews conducted at first call; the time of day; or the number of contact attempts until a successful interview (Kirchner & Olson, 2017; West & Blom, 2017).

Data & Methods

Data

This paper uses data from the National Educational Panel Study (NEPS; see Blossfeld & Roßbach, 2019), Starting Cohort 4 – Grade 9 (doi:10.5157/NEPS:SC4:11.0.0; NEPS Network, 2020). The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide network and it is a multi-cohort longitudinal survey designed to find out more about how education is acquired in Germany and how competencies develop over time. Following a multi-informant perspective, the study is not restricted to student data but also includes data from relevant context persons such as parents, teachers, and school heads. For our analyses, we use the parents CATI interviews of wave 1 of a sample of students in grade 9, recruited from 545 randomly selected regular schools in Germany as well as 103 schools for students with special educational needs. The interview is directed to the parent primarily responsible for students' school aspects. From a total child sample of 16,425 cases, 11,097 (68%) parents gave their permission to be contacted for the parent interviews. Going along with the progression of the data collection within schools (and obtaining parents' permission) the addresses for the parent interviews were handed over to the responsible fieldwork agency in three tranches. From that parent sample, 9,180 (83%) CATI sessions were completed during fieldwork between January and July 2011 (for a description of the fieldwork for the parent interview see Aust et al., 2012). Additionally, some parents have more than one child involved in the study and some interviews were conducted in Turkish and Russian. As we are interested in keeping the interview dura-

tion as comparable as possible, we exclude incomplete interviews and interviews with parents with more than one child in the NEPS from the sample. Furthermore, as there is evidence of language accounting for shorter interviews (Vandenplas et al., 2018), we kept only those that were conducted in German. After excluding cases who asked for data deletion, the final sample consists of $N=8,622$ parent interviews (AAPOR Response Rate 1: 0.824; American Association for Public Opinion Research, 2016). The parent interviews were conducted by 180 interviewers. While children that participate in the NEPS do receive incentives, the parents only receive an advance letter with detailed accompanying information by regular mail.

Dependent Variable: Interview Duration

The dependent variable is interview duration. Within this paper, it is operationalized as core interview duration that measures the time in minutes passed from the start until the end of all content-related questions asked to the respondent. The contact module with information concerning the study is excluded as well as the verification of the status of the respondent as the child's legal guardian and main education contact parent. Also excluded are final questions concerning updates of contact data. The interview duration is calculated by using the time stamps which indicate the transition between questionnaire modules and ranges from 12.3 minutes minimum to 100.7 maximum, with a mean of 31.3 minutes and a standard deviation of 8.8. Following previous research (Garbarski et al., 2020; Olson & Smyth, 2015), we apply two transformations to the dependent variable: First, the response times with values below the 1st and higher than the 99th percentile were trimmed and replaced by those percentile values; and second, the variable was log-transformed to correct for skewness.

Explanatory Variables

We model two types of interviewer experience: within-survey interviewer experience and overall interviewer experience. We operationalize within-survey interviewer experience as the count of successful interviews per interviewer in the chronological order (time and date) as registered in the CATI software time-stamps. In terms of indicators for interviewer workload we register a range of interviews from 1 to 253, with a mean of 47.9, a standard deviation of 42.5, and a median of 38.5. The within-survey interviewer experience variable was also log-transformed to account for a possible nonlinear learning effect (Kirchner & Olson, 2017; Kosyakov et al., 2021).

Overall interviewer experience was operationalized as the number of years each interviewer has worked for the contracted fieldwork agency. Overall interviewer experience with the given fieldwork agency is aggregated into the categories

Table 1 Variables included in the analysis

Block	Variables
Main explanatory variables	Within-survey interviewer experience Overall interviewer experience
Interviewer-level controls	Gender Age Education
Respondent-level controls	Gender Age Education Employment status Net equivalent income Household size Type of child's school
Interview-level controls	Number of contact attempts before a successful interview Interview at the first call Days since advance letter Telephone connection Time of day Day of the week Item nonresponse (%) Number of questions

of below 2 years of employment, 2 to 3, 4 to 5, or more than 5 years of experience. This is notoriously different from previous studies where interviewer experience was operationalized as a dichotomous variable distinguishing between no previous experience and at least 1 year of experience (Kirchner & Olson, 2017; Kosyakova et al., 2021). Furthermore, we include three additional distinct sets of control variables related to the interviewer, respondent, and interview characteristics, detailed in Table 1.

Gender and age of interviewer and respondent are included as dichotomous variables, while education is operationalized as a three-level categorical variable consisting of lower, intermediate, and higher education. Also at the respondent level, employment status is included as dichotomous variable while income is modeled as net equivalent income (OECD, 2013) distinguishing between three income groups (risk of poverty, average income, high income) using the official national median income threshold for the year 2011 of 1,416 € (Statistische Ämter, 2021). Household size distinguishes between one to three persons, four persons, and more

than four persons in the household; the type of school the child attends is divided between “*Gymnasium*” (the school that leads to a university entrance certificate) and other German educational possibilities. Finally, under the interview characteristics block, we further introduce the number of contact attempts before a successful interview; whether the interview was conducted at the first realized telephone contact; the number of days between the posting of the advance letter requesting the parents’ participation (controlling for three tranches handed over from the school field) and the interview; the type of telephone connection (landline, mobile, undefined); the time of day and whether the interview was conducted during the week or at the weekend; the percentage of item nonresponse; and the total number of questions answered. As discussed in the previous section, the effect of several of these variables on interview duration is inconsistent throughout the literature. Consequently, we do not elaborate further on our theoretical expectations. The descriptives for all variables under study are given in Table A.1 in the appendix.

Method

The main interest in this paper is to study how interviewer experience is associated with interview duration. Given that each interviewer conducted several interviews, we follow a multilevel modeling strategy where the first level corresponds to respondents and the higher level to interviewers, under a two-level hierarchical linear model with random intercepts framework. The models are estimated using the *R* (R Core Team, 2021) environment and we fit several two-level hierarchical linear models with random intercepts using the package *lme4* (Bates et al., 2015). The model is formulated in the following way:

$$\log(\text{Interview Duration})_{i,j} = \beta_0 + \beta_1 \text{Var1}_i + \beta_2 \text{Var2}_j + u_j + \varepsilon_{i,j} \quad (1)$$

In this equation, the subscript formalizes the clustered nature of the data where respondents (*i*) are nested within interviewers (*j*). The different explanatory variables are represented by β , where β_1 (Var1_i) denotes regression coefficients for respondent-level variables, such as respondent age or gender, and β_2 (Var2_j) for the interviewer-level variables, such as the interviewer experience indicators. The parameter β_0 reflects the fixed overall effect and u_j the interviewer random-effects component. Finally, we assume that the individual unobserved heterogeneity is uncorrelated with the explanatory variables, following a normal distribution, and the residual error term is represented by $\varepsilon_{i,j}$.

The model is estimated in a stepwise approach starting with the unconditional or empty model (Model 0), which shows how much of the variance of interview duration is explained by the higher level (interviewer). The introduction of within-survey interviewer experience and overall interviewer experience follows in the

next step in Model 1. Each thematic block of control variables is then introduced sequentially (interviewer-level, respondent-level, and interview-level characteristics), respectively Model 2, Model 3, and Model 4. The last model (Model 5), adds the interaction term between within-survey and overall interviewer experience, introduced to test whether within-survey interviewer experience has a differential impact concerning each of the overall interviewer experience categories (<2 years, 2–3 years, 4–5 years, >5 years). In each of the steps, we will look closely at the intraclass correlation coefficient (ICC) as a measure indicating the variance due to the interviewer.

There was a total of less than 1% missing data, with no missing observations on the dependent variable (Figure S.1, online supplementary material). Most affected by item nonresponse is the income variable. Aiming to minimize bias due to missing data, we used the package *missForest* for multiple imputation (Stekhoven, 2022). The *missForest* is a nonparametric method of imputation in which the algorithm used is an iterative process that assigns initial values to the missing data, fits a random forest for each variable based on the observed values predicting new imputed observations until convergence (Stekhoven & Bühlmann, 2012). Also as a robustness check, Table S.3 (online supplementary material) replicated the main model without the imputation procedure, considering only the cases with complete information. The results are very close to the main model, indicating therefore that the imputation process is unlikely to be driving our main results.

Results

How is interview duration related to interviewer experience? Before the multilevel model results, Figure 1 shows a descriptive analysis comparing the mean interview duration as the fieldwork progressed, by overall interviewer experience and the number of interviews.

Altogether, the mean interview duration for more overall experienced interviewers (>5 years) is 30.5 minutes and 32.9 minutes for less experienced interviewers (<2 years). Interviewers with two to three years of experience (2–3 years) have a mean of 32.1 minutes while in the remaining overall experience category (4–5 years) the average is 31.6 minutes. These differences indicate that more experienced interviewers are generally faster than the less experienced ones; a tendency also present in Figure 1, which also considers, in chronological order, the increasing number of interviews by a given interviewer within that specific survey. It should be noted, though, that in the highest within-survey interviewer experience category (“+200”), only the more overall experienced interviewers are included because there are no cases in the lower experience categories. Nevertheless, these initial descriptive findings indicate that, in disagreement with previous research on tele-

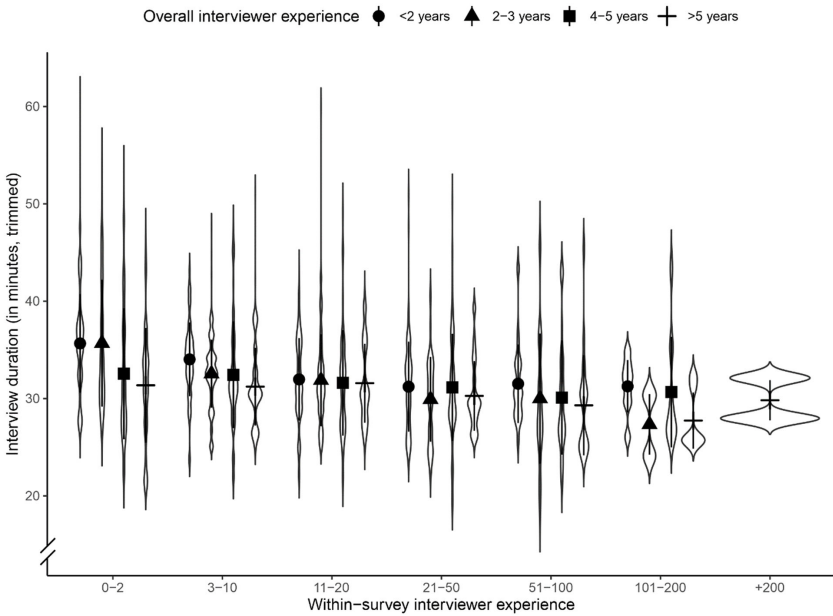


Figure 1 Average interview duration distribution by within-survey interviewer experience intervals and mean overall interviewer experience

phone surveys (Kirchner & Olson, 2017), both types of interviewer experience are likely associated with a tendency towards shorter interview durations. Next, we take the analysis forward and consider other factors that can affect the relationship of the experience variables and interview duration in the NEPS parent interviews.

Table 2 presents the coefficients and respective standard errors, variance components, and model comparison statistics for all the estimated multilevel regression models for interview duration. For reasons of clarity, we only present the estimates for the main explanatory variables. The complete table, including the estimates for the control variables, can be consulted in Table A.2 in the appendix.

The ICC for the null model shows that interviewers account for 29% of the variance of interview duration. While this proportion is higher than typically observed for substantive variables in telephone surveys, it aligns closely with previous studies examining face-to-face interviews (e.g. Olson & Peytchev, 2007; Loosveldt & Beullens, 2013a; Kosyakova et al., 2021). This indicates that the interviewer grouping variable significantly affects the mean interview duration. As we add more variables to the model, the null model is taken as the baseline.

The next step is to introduce the interviewer experience variables: within-survey and overall interviewer experience (Model 1). Introducing these variables

Table 2 Hierarchical linear random intercept models: estimated coefficients, standard errors, and variance components (reduced to main explanatory variables and interaction)

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	Null	+	+	+	+	+
		Experience	Interviewer	Respondent	Interview	Interaction
Main explanatory variables						
Within-survey interviewer experience (log)		-0.029*** (0.002)	-0.030*** (0.002)	-0.031*** (0.002)	-0.043*** (0.002)	-0.046*** (0.004)
Overall interviewer experience (ref. < 2 years)						
2–3 years		-0.029 (0.027)	-0.047 (0.026)	-0.045 (0.026)	-0.030 (0.022)	-0.003 (0.027)
4–5 years		-0.023 (0.027)	-0.053 (0.026)*	-0.052* (0.026)	-0.032 (0.022)	-0.049 (0.027)
> 5 years		-0.031 (0.032)	-0.083*** (0.032)	-0.081** (0.032)	-0.070** (0.027)	-0.122*** (0.032)
Interaction						
Within-survey * 2–3 years experience						-0.011 (0.006)
Within-survey * 4–5 years experience						0.006 (0.005)
Within-survey* > 5 years experience						0.018** (0.005)
Intercept	3.466*** (0.011)	3.564*** (0.021)	3.539*** (0.031)	3.575*** (0.032)	3.467*** (0.029)	3.477*** (0.030)

Table 2 (continued)

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
	Null	+ Experience	+ Interviewer	+ Respondent	+ Interview	+ Interaction
Residual variance (interviewer)	0.020	0.017	0.015	0.014	0.010	0.010
Residual variance (respondent)	0.047	0.046	0.046	0.043	0.032	0.032
ICC	0.294	0.264	0.238	0.245	0.240	0.241
Marginal R ²	-	0.020	0.050	0.103	0.319	0.318
Conditional R	0.294	0.276	0.275	0.323	0.482	0.482
LogLikelihood	703.09	770.51	783.57	1,091.53	2,377.67	2,388.44
Pr (>Chisq)		0.000***	0.000***	0.000***	0.000***	0.000***
Sample size (respondents/interviewers)	8,622/180					

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Effects at the interviewer-level, respondent-level, and interview-level controls are omitted; for complete results see appendix, Table A.2.

results in an improvement of model fit, as indicated by the likelihood χ^2 test, and the ICC is reduced to 26%. The effect of within-survey interviewer experience itself is negative and significant, thus giving support to the argument that study-specific experience explains the reduction of interview duration. On the other hand, Model 1 shows no significant effect of overall interviewer experience on interview duration. However, this result changes as more explanatory variables are included in the model.

The next steps introduce the respective blocks of control variables: Model 2—when including the interviewer-level controls—shows that while within-survey interviewer experience still has a negative significant effect on interview duration, the impact of overall interviewer experience categories is negative and statistically significant (at 5% level) on the highest experience categories (4–5 years and >5 years). The remaining blocks of controls for respondent and interview characteristics are sequentially introduced in Model 3 and Model 4. Every time, model fit improved significantly, but the ICC is only reduced slightly to 24%.

Figure 2 shows the predicted conditional effect of within-survey interviewer experience by overall interviewer experience on interview duration. Less overall experienced interviewers start with higher interview durations and this difference holds throughout the fieldwork.

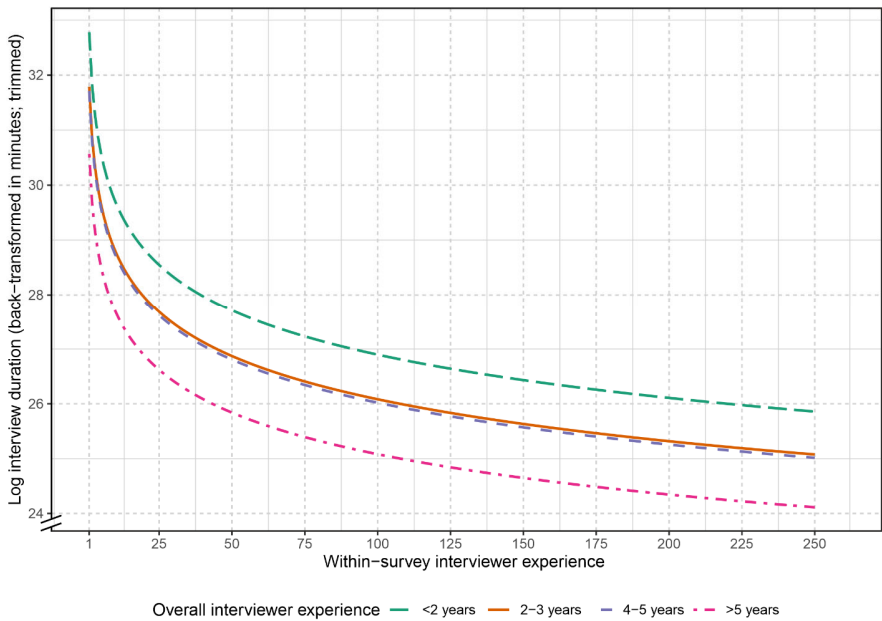


Figure 2 Predicted conditional interview duration by within-survey interviewer experience and overall interviewer experience (based on Model 4)

This indicates that interviewers with more than 2 years of experience on average start the fieldwork with a shorter interview duration and are consistently faster over the whole field phase. Furthermore, Model 4 shows that within-survey interviewer experience has a statistically significant negative effect on interview duration, even after including all control variables in the model. On the other hand, Model 4 also hints to a more nuanced interpretation regarding the effect of overall interviewer experience. In Table 2, we can see that after all blocks of controls are introduced, only the most overall experience category (>5 years) is still statistically associated with shorter interviews. While in other experience categories the significant negative effect on duration is explained away, only in the more experienced category of interviewers we see the persistence of the negative effect on the dependent variable. As for within-survey interviewer experience, the tendency of interviewers reducing interview duration as they conduct more interviews is clear.

We checked the effect of the control variables (see Model 4 with all controls in Table A.2 in the appendix): Older interviewers conduct longer interviews but there is no effect of gender and education on the interviewers' side. On the respondent level, those under risk of poverty in terms of net equivalent income have also significantly longer interviews than parents in the intermediate category of net equivalent income. Female respondents and households with more or fewer than four members also take significantly less time to be interviewed; respondent age and educational level other than intermediate go along with longer interviews. Moreover, being unemployed is not a significant predictor of interview duration. In contrast, the child enrolled in a "Gymnasium" is a significant predictor of a shorter interview. As the NEPS is dedicated to studying the German educational system, this result is particularly relevant as it is consistent across models and even after controlling for the number of questions asked (Model 4). Regarding the effects of interview characteristics, we see that a higher number of contact attempts before a successful interview is associated with longer interviews which could be due to the characteristics of harder-to-reach interviewees. Furthermore, respondents with higher rates of item nonresponse also have longer interviews. Another interesting result is that interviews conducted at the first realized contact are faster. It seems that if the respondent agrees to answer the survey immediately, the time used to complete the questionnaire is significantly shorter. Having the interview on the weekend is not significantly associated with interview duration, whereas interviews in the afternoon are shorter than those in the morning. In contrast, the number of days since the sending of the advance letter to the respondents, with the request to participate in the survey, is significant albeit with a very close to zero effect on interview duration. Also, as discussed in the literature, we find that respondents who use a mobile phone connection take significantly longer than respondents who use a landline. Finally, the number of questions is positively associated with interview duration.

If shorter interview durations are attributed to within-survey interviewer experience, estimating an interaction between within-survey and overall interviewer experience lets us examine if conducting more interviews within one survey affects interview duration differently for interviewers with more or less overall professional experience. Including the interaction (Model 5) between both interviewer experience measures improves model fit significantly, as shown in Table 2. Figure 3 shows the predicted conditional effect on interview duration by different levels of the main explanatory variables within-survey interviewer experience and overall interviewer experience. While there are no significant differences across overall experience categories in the first interviews, this eventually changes. When examining Model 5 (Figure 3), which allows the within-survey coefficient to differ across overall interviewer experience-groups, we observe nuances suggesting varied learning trajectories across experience brackets. Figure 3 reveals that from the start, the “2–3 years” overall interviewer experience-group has a steeper decrease in interview duration than the “>5 years” experience group. By the 50th interview, the two durations intersect. This observation aligns with Table 2, where the “<2 years”, “2–3 years”, and “4–5 years” groups show a sharper decline in duration compared to the “>5 years” group. Keeping all other variables constant, by the 50th interview the interviewers belonging to the overall experience category of “2–3

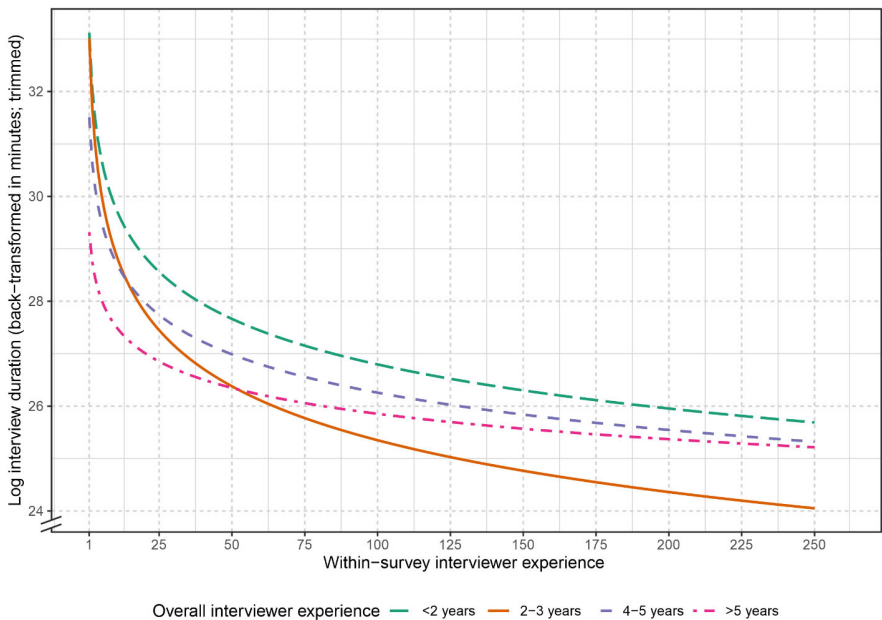


Figure 3 Predicted conditional interview duration by within-survey interviewer experience and overall interviewer experience (based on Model 5)

years” and “>5 years” take on average approximately 26.2 minutes to conduct an interview. This means that by the fiftieth interview, on average, the “2–3 years” overall experience group take less 6.6 minutes than their first interview while the “>5 years” overall experience group take 3.0 minutes less than their first interview.

The robustness of all these results was tested by estimating alternative model specifications. The results of these estimations can be found in the online supplementary material. First, our main model was replicated without trimming the dependent variable at the 1st and 99th percentile, without log transforming the dependent variable and, finally, without imputation of the missing variables. The results are very similar and can be found respectively in Table S.1, Table S.2 and Table S.3 (online supplementary material). Second, as discussed by previous research, an alternative explanation to survey experience driving the tendency for interviews to become shorter as the fieldwork develops is related to socio-demographic changes within the respondent sample (Kirchner & Olson, 2017). Throughout the field period, the characteristics of the respondents may change, making it more likely for interviews to take longer because of older and less educated respondents, for example. Even though we control for several respondent characteristics, in order to rule out a “compositional aspect” effect over the field period (Kirchner & Olson, 2017, p. 86), we divide the respondents of the first (and by sample size largest) tranche into three different samples of early, middle, and late respondents and compared socio-demographic characteristics. This effort showed some differences between early, middle, and late responders namely in terms of respondent age, employment status, household size, and the type of school attended by the child (Table S.4, online supplementary material). However, our main results have shown that the survey experience effect is robust even with respondent-level socio-demographic characteristics in the model. Following this line of thought, Table S.5 (online supplementary material) repeats the main model for the first tranche of 5,975 respondents only. The estimated coefficients for the first tranche are very similar to the main model results. Thus, it is not likely that the effect of interviewer experience on interview duration is being driven by the socio-demographic composition of the sample.

Overall, the proportion of variance explained by the interviewers varies between 29% in the null model (Model 0) and 24% for the complete model (Model 4). The introduction of within-survey and overall interviewer experience, as well as further control variables, did impact the ICC, but only reducing it by 5 percentage points. Given that several other potential confounders are included in the model, this indicates that interviewers have a large impact on how long the survey interview lasts.

Discussion

This paper aimed to investigate how interviewer experience impacts interview duration in a CATI-based large-scale panel study. First, as in previous research on face-to-face interviewing, our results show that interviewers are an important source of variation for interview duration also in telephone surveys, expanding the available empirical evidence to other modes of data collection. The variance explained by the interviewer level in this study is large and slightly higher when compared to other telephone surveys (Kirchner & Olson, 2017). Second, our results give further support to the findings that within-survey interviewer experience impacts interview duration (Kirchner & Olson, 2017; Kosyakova et al., 2021). This effect is stable and robust to the introduction of control variables for interviewer, respondent, and interview characteristics. On the other hand, contrasting previous findings using CATI survey data, we find that overall interviewer experience does have a significant negative impact on interview duration but only for the more experienced interviewers. This indicates that the effect of overall interviewer experience on time duration is not really continuous as interviewers gain experience. Instead, it appears more likely that interviewers working for more than 5 years (>5 years) in the survey fieldwork agency conduct interviews faster.

Third, the effect of within-survey experience on interview duration differs between categories of overall interviewer experience. While the difference becomes evident as interviews progress during fieldwork, interviewers with up to five years of experience tend to speed up at a faster rate than those with an experience of more than five years. It is particularly telling that the same effect was not found for the more inexperienced interviewers. For interviewers, it appears to be necessary to have some previous experience and knowledge to change their conduct in order to achieve shorter interview duration.

Furthermore, we also find some of the controls with important effects. Namely, the demographics of the interviewer, characteristics of the respondent, their socio-economic conditions, and the child's school situation as well as several interview characteristics impact interview duration. Most notably, parents whose child attends a "Gymnasium" have a shorter interview duration compared to children from other school types. A possible explanation for this result could be that the main aim of the parent interview is to talk about their children and, in the German context, "Gymnasium" children have a somewhat more streamlined and easier to explain educational trajectory. On side of the interview characteristics, the picture is a bit more mixed. Conforming with previous findings, a higher item non-response rate is associated with a longer interview duration, suggesting that interviewers might invest additional time to evoke an answer from the respondent—and not quickly accept a nonresponse and jump to the next question. This positive but also partly

counter-intuitive result has also been found in other studies (Kirchner & Olson, 2017).

Our analyses go along with a set of methodological limitations: (1) Whereas the measurement of within-survey interviewer experience is automatically recorded within the interviews and available in a fine-grained manner, the result concerning overall interviewer experience should be seen with caution as the inexperienced category is below 2 years of experience. It can be argued that this interviewer overall experience category is a measure too blunt to distinguish between experienced and not experienced interviewers. (2) Another limitation of this study is that we were not able to include the characteristics of the questions as item-based timestamps are not available. Previous research has shown that response times are also related to question type, question length, response format, presence of instructions, or the labeling of the response scale (Garbarski et al., 2020; Olson, Smyth, & Kirchner, 2020). (3) Our study can also be considered limited due to the lack of information regarding interviewer behavior and the interaction between the respondent and the interviewer. While we uncovered some patterns about how interviewers and respondents interact, we are still some distance away from unveiling the actual dynamics in each interview. More measures of this adaptive relationship between interviewers and respondents are necessary to link more closely how both of these agents' behaviors differentially impact interview duration. On a final note, while we look at the percentage of item missings, interview duration is an indirect measure and any further steps should include additional indicators of interviewer performance and data quality.

Nonetheless, the results of this paper can be used for optimizing interviewer training and supervision as well as for more adequate cost-forecasting within large-scale panel studies.

An early transition to shorter interviews due to learning effects and routine with the instrument would be desirable concerning survey costs. This could also impact the forecast of cost-aspects as CATI interview time in the NEPS is billed by the minute. Extension of practical sessions could be introduced in interviewer training and this way reduce interview duration earlier, even though that might not always be desirable. Nevertheless, more detailed research is needed to distinguish whether the decline in interview duration is a general learning effect or due to special interviewer behaviors, such as deviating from the standard protocols or reducing unnecessary conversations during the first interviews.

References

- Ackermann-Piek, D., & Massing, N. (2014). Interviewer Behavior and Interviewer Characteristics in PIAAC Germany. *Methods, Data, Analyses*, 8(2), 199–222. <https://doi.org/10.12758/MDA.2014.008>
- American Association for Public Opinion Research. (2016). *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys*. American Association for Public Opinion Research (AAPOR). <https://aapor.org/wp-content/uploads/2022/11/Standard-Definitions20169theditionfinal.pdf>
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V., BVM Berufsverband Deutscher Markt- und Sozialforscher e. V., & Deutsche Gesellschaft für Online-Forschung e. V. (2021). *Richtlinie für telefonische Befragungen*. <https://www.adm-ev.de/wp-content/uploads/2021/01/RL-Telefon-neu-2021.pdf>
- Aust, F., Hess, D., & Prussog-Wagner, A. (2012). *Methodenbericht. NEPS. Startkohorte 4 (Elternbefragung). Haupterhebung Frühjahr 2011 B34*. Bonn. infas Institut für angewandte Sozialwissenschaft. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/Methodenbericht_B34.pdf
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Blossfeld, H.-P., & Roßbach, H. G. (Eds.). (2019). *Education as a Lifelong Process. The German National Educational Panel Study (NEPS)* (Second Edition). Springer VS.
- Böhme, M., & Stöhr, T. (2014). Household Interview Duration Analysis in CAPI Survey Management. *Field Methods*, 26(4), 390–405. <https://doi.org/10.1177/1525822X14528450>
- Fowler, F. J., Jr., & Mangione, T. W. (1990). *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. SAGE.
- Garbarski, D., Dykema, J., Schaeffer, N. C., & Edwards, D. F. (2020). Response Times as an Indicator of Data Quality: Associations with Question, Interviewer, and Respondent Characteristics in a Health Survey of Diverse Respondents. In K. Olson, J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, & B. T. West (Eds.), *Interviewer Effects from a Total Survey Error Perspective* (pp. 253–264). CRC Press.
- Jin, J., Vandenplas, C., & Loosveldt, G. (2019). The Evaluation of Statistical Process Control Methods to Monitor Interview Duration During Survey Data Collection. *SAGE Open*, 9(2), 1-14. <https://doi.org/10.1177/2158244019854652>
- Kirchner, A., & Olson, K. (2017). Examining Changes of Interview Length over the Course of the Field Period. *Journal of Survey Statistics and Methodology*, 5(1), 84–108. <https://doi.org/10.1093/jssam/smw031>
- Kosyakova, Y., Olbrich, L., Sakshaug, J. W., & Schwanhäuser, S. (2021). Positive learning or deviant interviewing? Mechanisms of experience on interviewer behavior. *Journal of Survey Statistics and Methodology*, 1–27. <https://doi.org/10.1093/jssam/smab003>
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236. <https://doi.org/10.1002/acp.2350050305>
- Lepkowski, J., Axinn, W., Kirgis, N., West, B. T., Kruger Ndiaye, S., Mosher, W., & Groves, R. (2010). *Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection* (NSFG Survey Methodology 10-012). Population Studies Center, Institute for Social Research, University of Michigan.

- Loosveldt, G., & Beullens, K. (2013a). 'How long will it take?' An analysis of interview length in the fifth round of the European Social Survey. *Survey Research Methods*, 7(2), 69–78. <https://doi.org/10.18148/srm/2013.v7i2.5086>
- Loosveldt, G., & Beullens, K. (2013b). The impact of respondents and interviewers on interview speed in face-to-face interviews. *Social Science Research*, 42(6), 1422–1430. <https://doi.org/10.1016/j.ssresearch.2013.06.005>
- NEPS Network. (2020). *Scientific Use File of Starting Cohort Grade 9 (Version 11.0.0)* [Computer software]. Leibniz Institute for Educational Trajectories (LifBi). Bamberg, Germany. <https://doi.org/10.5157/NEPS:SC4:11.0.0>
- OECD. (2013). *OECD Framework of statistics on the distribution of household income, consumption and wealth*. OECD Publishing. <https://doi.org/10.1787/9789264194830-en>
- Olson, K., & Peytchev, A. (2007). Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes. *Public Opinion Quarterly*, 71(2), 273–286. <https://doi.org/10.1093/poq/nfm007>
- Olson, K., & Smyth, J. D. (2015). The Effect of CATI Questions, Respondents, and Interviewers on Response Time. *Journal of Survey Statistics and Methodology*, 3(3), 361–396. <https://doi.org/10.1093/jssam/smv021>
- Olson, K., & Smyth, J. D. (2020). What Do Interviewers Learn? Changes in Interview Length and Interviewer Behaviors over the Field Period. In K. Olson, J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, & B. T. West (Eds.), *Interviewer Effects from a Total Survey Error Perspective* (pp. 279–290). CRC Press.
- Olson, K., Smyth, J. D., Dykema, J., Holbrook, A. L., Kreuter, F., & West, B. T. (Eds.). (2020a). *Interviewer Effects from a Total Survey Error Perspective*. CRC Press.
- Olson, K., Smyth, J. D., Dykema, J., Holbrook, A. L., Kreuter, F., & West, B. T. (2020b). The Past, Present, and Future of Research on Interviewer Effects. In K. Olson, J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, & B. T. West (Eds.), *Interviewer Effects from a Total Survey Error Perspective* (pp. 3–15). CRC Press.
- Olson, K., Smyth, J. D., & Kirchner, A. (2020). The Effect of Question Characteristics on Question Reading Behaviors in Telephone Surveys. *Journal of Survey Statistics and Methodology*, 8(4), 636–666. <https://doi.org/10.1093/jssam/smz031>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Statistische Ämter. (2021). *Armutsgefährdung und Einkommensverteilung (MZ-Kern)*. Statistische Ämter des Bundes und des Landes. Gemeinsames Statistikportal. <https://www.statistikportal.de/de/sbe/ergebnisse/einkommensarmut-und-verteilung>
- Stekhoven, D. J. (2022). *missForest: Nonparametric Missing Value Imputation using Random Forest* (Version R package version 1.5) [Computer software].
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Stiegler, A., & Biedinger, N. (2015). *Interviewer Qualifikation und Training* (GESIS Survey Guidelines). GESIS - Leibniz-Institut für Sozialwissenschaften. <https://www.gesis.org/gesis-survey-guidelines/operations/interviewertraining-und-effekte/interviewertraining> https://doi.org/10.15465/GESIS-SG_013
- Sturgis, P., Maslovskaya, O., Durrant, G., & Brunton-Smith, I. (2021). The Interviewer Contribution to Variability in Response Times in Face-to-Face Interview Surveys. *Journal of Survey Statistics and Methodology*, 9(4), 701–721.

- <https://doi.org/10.1093/jssam/smaa009>
- Timbrook, J., Olson, K., & Smyth, J. D. (2018). Why Do Cell Phone Interviews Last Longer? A Behavior Coding Perspective. *Public Opinion Quarterly*, 82(3), 553–582. <https://doi.org/10.1093/poq/nfy022>
- Vandenplas, C., Beullens, K., & Loosveldt, G. (2019). Linking interview speed and interviewer effects on target variables in face-to-face surveys. *Survey Research Methods*, 13(3). <https://doi.org/10.18148/SRM/2019.V13I3.7321>
- Vandenplas, C., Loosveldt, G., Beullens, K., & Denies, K. (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(4), 516–538. <https://doi.org/10.1093/jssam/smx034>
- West, B. T., & Blom, A. G. (2017). Explaining Interviewer Effects: A Research Synthesis. *Journal of Survey Statistics and Methodology*, 5(2), 175–211. <https://doi.org/10.1093/jssam/smw024>

Appendix

Table A.1 Descriptive statistics

Variable	N	Mean (SD) / Proportion %	Min	Max
<i>Dependent variable</i>				
Interview duration	8,622	31.27 (8.41)	16.5	60.8
<i>Main explanatory variables</i>				
Within-survey interviewer experience	8,622	43.15 (39.74)	1	253
Overall interviewer experience	180			
< 2 years	49	27.2%		
2–3 years	50	27.8%		
3–4 years	53	29.4%		
> 5 years	28	15.6%		
<i>Interviewer-level controls</i>				
Gender	180			
Male	92	51.1%		
Female	88	48.8%		
Age	180			
< 30	62	34.4%		
30–49	62	34.4%		
50–65	45	25%		
> 65	11	6.1%		
Education	175			
Lower	38	21.1%		
Intermediate	32	17.8%		
Higher	105	58.3%		
<i>Respondent-level controls</i>				
Gender	8,622			
Male	1,451	16.8%		
Female	7,171	83.2%		
Age	8,622	45.68 (5.16)	25	92
Education	8,608			
Lower	762	8.9%		
Intermediate	4,936	57.3%		
Higher	2,910	33.8%		
Employment status	8,615			
Employed	7,265	84.3%		
Unemployed	1,350	15.7%		

Table A.1 (continued)

Variable	N	Mean (SD) / Proportion %	Min	Max
Net equivalent income	7,128			
Risk of poverty	2,153	30.2%		
Average income	4,316	64.8%		
High income	359	5.0%		
Household size	8,620			
1–3 persons	2,593	30.1%		
4 persons	3,772	43.8%		
> 4 persons	2,255	26.2%		
Type of child's school	8,622			
Other school	5,196	60.3%		
Gymnasium	3,426	39.7%		
<i>Interview-level controls</i>				
Number of contact attempts	8,622	5.84 (8.01)	1	100
Interview at the first call	8,622			
First call	1,453	16.9%		
Not first call	7,169	83.1%		
Days since advance letter	8,622	47.89 (24.32)	7	165
Telephone connection	8,622			
Landline	5,813	67.4%		
Mobile phone	435	5.0%		
Undefined	2,374	27.5%		
Time of day	8,622			
Morning	1,956	22.7%		
Afternoon	4,151	48.1%		
Evening	2,515	29.2%		
Day of the week	8,622			
Weekday	6,885	79.9%		
Weekend	1,737	20.1%		
Item nonresponse	8,622	0.77 (1.02)	0	11.5
Number of questions	8,622	254.17 (30.72)	165	327

Table A.2 Hierarchical linear random intercept models: estimated coefficients, standard errors, and variance components (including controls)

	Model 0 Null	Model 1 + Experience	Model 2 + Interviewer	Model 3 + Respondent	Model 4 + Interview	Model 5 + Interaction
<i>Main explanatory variables</i>						
Within-survey interviewer experience (log)		-0.029*** (0.002)	-0.030*** (0.002)	-0.031*** (0.002)	-0.043*** (0.002)	-0.046*** (0.004)
Overall interviewer experience [ref. < 2 years]						
2–3 years		-0.029 (0.027)	-0.047 (0.026)	-0.045 (0.026)	-0.030 (0.022)	-0.003 (0.027)
4–5 years		-0.023 (0.027)	-0.053 (0.026)*	-0.052* (0.026)	-0.032 (0.022)	-0.049 (0.027)
5 years		-0.031 (0.032)	-0.083*** (0.032)	-0.081*** (0.032)	-0.070** (0.027)	-0.122*** (0.032)
<i>Interviewer-level controls</i>						
Gender [female]			-0.009 (0.019)	-0.014 (0.019)	-0.015 (0.016)	-0.017 (0.016)
Age [ref. < 30 years old]						
30–49			0.087*** (0.024)	0.080*** (0.023)	0.064*** (0.020)	0.067*** (0.020)
50–65			0.120*** (0.026)	0.120*** (0.026)	0.106*** (0.022)	0.105*** (0.022)
> 65			0.157*** (0.045)	0.153*** (0.044)	0.147*** (0.038)	0.149*** (0.038)
Education [ref. intermediate]						
Lower education			-0.0214 (0.031)	-0.014 (0.030)	0.006 (0.026)	0.002 (0.026)
Higher education			-0.018 (0.026)	-0.012 (0.026)	-0.009 (0.022)	-0.010 (0.022)

Table A.2 (continued)

	Model 0 Null	Model 1 + Experience	Model 2 + Interviewer	Model 3 + Respondent	Model 4 + Interview	Model 5 + Interaction
<i>Respondent-level controls</i>						
Gender [female]				-0.045*** (0.006)	-0.016** (0.005)	-0.016** (0.005)
Age (centered)				0.001** (0.000)	0.002*** (0.000)	0.002*** (0.000)
Education [ref. intermediate]						
Lower education				0.141*** (0.008)	0.118*** (0.007)	0.118*** (0.007)
Higher education				0.010* (0.005)	0.126** (0.004)	0.126** (0.004)
Employment status [employed]				0.005 (0.006)	0.000 (0.005)	0.000 (0.005)
Net equivalent income [ref. average]						
Risk of poverty				0.014* (0.005)	0.049*** (0.005)	0.049*** (0.005)
High income				0.017 (0.011)	-0.006 (0.010)	-0.006 (0.010)
Household size [ref. 4 persons]						
1–3 persons				-0.028*** (0.005)	-0.010* (0.004)	-0.010* (0.004)
> 4 persons				0.022*** (0.005)	-0.010* (0.005)	-0.010* (0.005)
Type of child's school ["Gymnasium"]				-0.051*** (0.005)	-0.024*** (0.004)	-0.024*** (0.004)

Table A.2 (continued)

	Model 0 Null	Model 1 + Experience	Model 2 + Interviewer	Model 3 + Respondent	Model 4 + Interview	Model 5 + Interaction
<i>Interview-level controls</i>						
Number of contact attempts (centered)					0.008*** (0.000)	0.008*** (0.000)
Interview at the first call [not first call]					0.066*** (0.005)	0.066*** (0.005)
Days since advance letter (centered)					-0.000*** (0.000)	-0.000*** (0.000)
Telephone connection [ref. landline] Mobile phone					0.045*** (0.009)	0.044*** (0.009)
Undefined					0.023*** (0.004)	0.023*** (0.004)
Time of day [ref. morning] Afternoon					-0.026*** (0.005)	-0.025*** (0.005)
Evening					-0.006 (0.006)	-0.005 (0.006)
Day of the week [weekday]					-0.001 (0.005)	-0.002 (0.005)
Item nonresponse (%)					0.029*** (0.002)	0.029*** (0.002)
Number of questions (centered)					0.027*** (0.000)	0.027*** (0.000)

Table A.2 (continued)

	Model 0 Null	Model 1 + Experience	Model 2 + Interviewer	Model 3 + Respondent	Model 4 + Interview	Model 5 + Interaction
Within-survey * overall interviewer experience						
Within-survey * 2–3 years						-0.011 (0.006)
Within-survey * 4–5 years						0.006 (0.005)
Within-survey * > 5 years						0.018** (0.006)
Intercept	3.466*** (0.011)	3.564*** (0.021)	3.539*** (0.031)	3.575*** (0.032)	3.467*** (0.029)	3.477*** (0.030)
Residual variance (interviewer)	0.020	0.017	0.015	0.014	0.010	0.010
Residual variance (respondent)	0.047	0.046	0.046	0.043	0.032	0.032
ICC	0.294	0.264	0.238	0.245	0.240	0.241
Marginal R ²	-	0.020	0.050	0.103	0.319	0.318
Conditional R	0.294	0.276	0.275	0.323	0.482	0.482
LogLikelihood	703.09	770.51	783.57	1,091.53	2,377.67	2,388.44
Pr (>Chisq)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
Sample size (respondents/interviewers)	8,622/180					

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.