Universität Mannheim
Fakultät für Sozialwissenschaften

INTERVIEWER EFFECTS IN PIAAC GERMANY 2012

Inauguraldissertation zur Erlangung des akademischen Grades einer Doktorin der Sozialwissenschaften der Universität Mannheim

vorgelegt von Daniela Ackermann-Piek
Mannheim, 08.11.2017

Dekan:	Prof. Dr. Michael Diehl
Erstbetreuerin & Erstgutachterin:	Prof. Annelies G. Blom, Ph.D.
Zweitbetreuer & Zweitgutachter:	Prof. Dr. Oliver Arránz-Becker
Drittgutachterin:	Prof. Dr. Beatrice Rammstedt
Tag der Disputation:	08. Februar 2018

Table of Contents

Acknowledgments

A large number of people have guided and supported me throughout my time as a PhD student. First of all, I feel privileged that I have a supervisor who provided not only guidance for topic related ideas, critical questions, constructive discussions and feedback, but also believed in me, gave me time to write and kept me motivated. Many thanks to my first supervisor, Annelies Blom! I also thank my second supervisor, Oliver Arránz-Becker for providing helpful feedback and sharing expertise on methodological issues. Lastly, I want to thank Beatrice Rammstedt for reviewing my thesis at short notice.

Writing a dissertation is not possible without a good working environment. Thus, I want to thank all current and former colleagues at the University of Mannheim and GESIS for their fruitful support and discussions. Especially, I`m grateful to my former colleague Natascha Massing who inspired me to study interviewer effects as well as to all other colleagues from the German PIAAC team. I also thank my former colleagues at the SFB 884. Namely, Carina Cornesse and Susanne Helmschrott, for the very intense and motivating dissertation debates as well as Christian Bruch, Barbara Felderer, Franziska Gebhard, Jessica Herzing, Marina Jesse, Ulrich Krieger, Margarita Maklakova and Dayana Sieger. I will miss our office discussions.

Finally, and maybe most important, I thank my family and friends for their love and support throughout success and failure. I thank my parents, who taught me to keep trying even when it sometimes doesn't seem worth it. I thank my brother and his family, who always remind me of my roots and helped me to keep my feed on the ground. I especially thank my husband, who was most encouraging by patiently listening to my constantly recurring reports of successful and less successful research days, giving me sufficient time to be without him but at my computer, and support with all his unconditional love. Thank you, Christian for being the most beautiful person I know. Last, but not least, I thank my daughter Carla Lieselotte for reminding me every day what is actually important in life.

List of Tables

List of Figures

List of Appendices

1 Introduction

In this dissertation, I study interviewer effects on different aspects of survey quality in a face-to-face survey in Germany. Studying interviewer effects is relevant, because human interviewers have various tasks during the data collection process of surveys and can thus have an effect on multiple aspects of a survey. In my dissertation, I use the data from the German implementation of the Programme of the International Assessment of Adults Competencies 2012 (PIAAC) as well as from various auxiliary data sources.

Surveys are an important instrument to study society (Couper, 2013, p. 111). Typically, in surveys, data on a subset of the population is collected to make inferences about the target population of interest. This is only possible, when the data derived from surveys are of high quality, which has multiple definitions. Data quality is commonly assessed by the total quality framework, including the following dimensions: Accuracy, credibility, comparability, usability/interpretability, relevance, accessibility, timeliness/punctuality, completeness and coherence (Biemer, 2010).

When studying interviewer effects on survey quality, *survey accuracy* is of special interest. Survey accuracy is a concept described in the Total Survey Error (TSE) framework, a conceptual framework of the survey life cycle describing survey errors at all stages of a survey (see Groves & Lyberg, 2010). In terms of the TSE framework, survey accuracy is defined as the "deviation of an estimate of substantive survey variables from its underlying true parameter value" (Biemer, 2010, p. 817). In other words, an accurate survey should produce data that is error free to answer the intended research questions. As interviewers are often involved in implementing surveys – for example interviewers build sampling frames, attempt contact with sample persons and persuade them to participate, and administer the survey instrument – the data quality of a survey also depends on the work of interviewers.

Interviewer effects in surveys have been studied across many countries, for many surveys and with a focus on every possible aspect of the TSE framework (for an overview see West & Blom, 2017). To date, there have been various studies describing and some studies explaining interviewer effects on the different aspects of the TSE, including the construction of the sampling frame and sampling (Eckman, 2013; Eckman & Kreuter, 2011; Tourangeau, Kreuter, & Eckman, 2012), nonresponse (e.g., Blom, de Leeuw, & Hox, 2011; Durrant, D'Addio, & Steele, 2013; Jäckle, Lynn, Sinibaldi, & Tipping, 2013), and measurement (e.g. Durrant, Groves, Staetsky, & Steele, 2010; Rice, 1929). Furthermore, there are a few studies that looked into interviewer effects on multiple aspects of the TSE simultaneously (Brunton-Smith, Sturgis, & Williams, 2012; Loosveldt & Beullens, 2014; Olson, Kirchner, & Smyth, 2016; West, Kreuter, & Ursula, 2013; West & Olson, 2010).

This dissertation is structured as follows: I start with introducing the TSE framework in detail, followed by a description of the role and tasks of interviewers in administering a survey (Chapter 2). Next, I introduce the survey and auxiliary data I use in detail (Chapter 3). Subsequently, interviewer effects for all relevant stages of PIAAC Germany where interviewers were involved are examined including a detailed description of the literature on interviewer effects for these error sources. In doing so, I follow the structure of how surveys are implemented: I start with interviewer effects on unit nonresponse (Chapter 4), followed by a chapter on interviewers' deviations from standardized interviewing techniques (Chapter 5)[1] and a chapter on interviewer effects on the estimates of substantive survey variables (Chapter 6). After studying interviewer effects on unit nonresponse and measurement individually, I conduct combined analyses on interviewer effects on those two different survey errors simultaneously (Chapter 7). The last and final chapter summarizes the key findings and gives first insights into how interviewer effects may be countered (Chapter 8).

[1] Chapter 5 has been already published in a peer reviewed journal: Ackermann-Piek and Massing (2014).

Overall, the results of my dissertation show that there are interviewer effects on survey unit nonresponse as well as on the estimates of the substantive survey variables under research. Nevertheless, the magnitude of the interviewer variance differs between the different aspects of the survey data collection process: The interviewer variance of a sample persons' propensity to be successfully contacted is 20% and interviewer variance of a sample persons' propensity to participate in PIAAC it is 2%. The interviewer variance in the estimates of the substantive survey variables is between 0 and 7%. Furthermore, I show that interviewers deviate from standardized interviewing techniques. Although, I can draw on a very rich data source of interviewer characteristics – using data from PIAAC Germany – I find no clear pattern of variables that explain interviewer effects on survey unit nonresponse and on the estimates of the substantive survey variables. However, I find interviewer characteristics that explain interviewer variance in the estimates of some substantive variables under research. In addition, my results of the combined analyses of interviewer effects on survey unit nonresponse and on estimates of survey variables show a significant relationship between the predicted probabilities of successfully making contact with sample persons and the predicted means of the majority of the substantive survey variables at the interviewer level. However, I find no significant relationship between the predicted probabilities of successfully gaining sample persons' cooperation and the predicted means of the majority of the substantive survey variables under research at the interviewer level.

2 The Total Survey Error Framework and Interviewers' Role in Surveys

Within the present chapter, I briefly describe the Total Survey Error (TSE) and interviewers' role in surveys. A detailed description of interviewer effects for the individual error sources described in the TSE framework, including a detailed literature review, can be found in Chapters 4, 5, 6 and 7.

The TSE is a conceptual framework describing the life cycle of a survey from a quality perspective including statistical error properties for each part of a survey (Groves et al., 2009b; Groves & Lyberg, 2010). The TSE is divided into two major segments: Measurement and representation (see also Figure 2.1). The measurement part of the TSE starts with *constructs*, the elements of information that are operationalized to be *measured* within the next step. Subsequently, *responses* are obtained from respondents and, in another step, those responses are *edited*. The representation part of the TSE considers the population described by the survey. The description of the representation part starts with the definition of the *target population*, which is the set of sample units to be studied in the survey. In the next step, a *sample* is selected form the *sampling frame* of the target population. Since not all sample units are typically observed due to nonresponse, survey data is typically collected only from a subset of the sample, the *respondents*. Finally, before survey statistics are derived from the collected responses, the data may undergo *post-survey adjustments*.

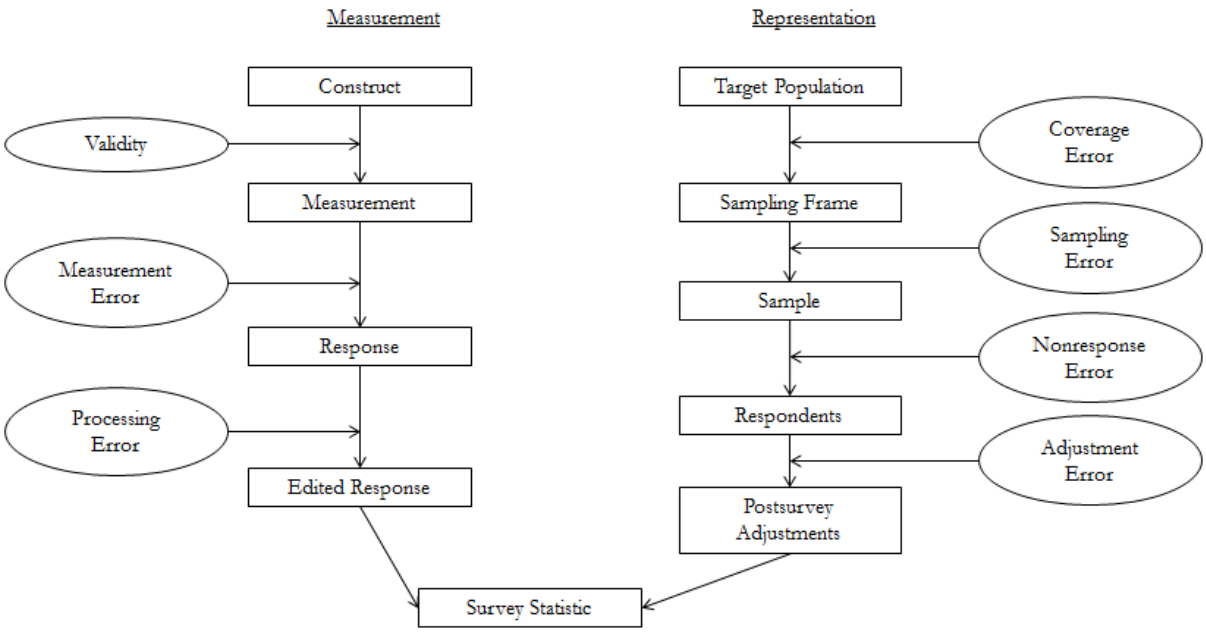


Figure 2.1 Survey life cycle from a quality perspective, Groves et al. (2009b, p. 48)

Groves et al. (2009b, p. 49-62) show, that at each step of the survey life cycle, errors can occur. For the measurement part of the survey life cycle they start with describing problems of *validity*, which can occur when the construct is not related to the measure in an optimal way. For example, in measuring education, the question might miss one important aspect and thus education is invalid measured. In addition, *measurement error* is possible when an obtained response is not equal to the value a respondent intends to provide and *processing error* can occur when responses are not correctly transferred into data, e.g. by incorrect coding or recoding. For the representation part of the survey, Groves et al. (2009b) describe four potential error sources: First, *coverage error* might occur when the sampling frame does not perfectly cover the target population, i.e. either sample persons are incorrectly excluded from or included in the sampling frame. Second, regarding *sampling error* they describe two types of error: Sampling bias and sampling variance. Sampling bias can occur, when some members of the sampling frame are systematically excluded from selection and sampling variance arises, because sampling – by definition – is a selection process ending up with different sets of frame elements, each set having different values in the survey statistic. Third, *nonresponse error* can occur when members of the sample do not respond to the survey request. Finally, post-survey adjustments are used to correct

for possible representation errors by means of statistical techniques. However, when the adjustments fail to correct for any misrepresentation, *adjustment error* can occur.

All errors described in the TSE framework may cause incorrect estimates of substantive survey variables. Thus, it is important to closley look at the quality of survey data with regard to the error sources described in the TSE famework. This also applies when interviewer effects are the main research interest: Interviewers are involved with almost all steps of the survey process and thus have an impact on almost all aspects of the TSE (for an overview of interviewers' impact on various aspects of the TSE see West & Blom, 2017). More specifically, interviewers are crucial in the data collection process and their tasks in surveys to complete are manifold: They build sampling frames by listing addresses, they detect eligible sampling units, they attempt contact with sample persons and persuade them to participate, they clarify the goal of the survey and explain to respondents their role in the interview, they ask questions and record answers, and they transmit data (Groves et al., 1992; Loosveldt, 2008; Schaeffer, Dykema, & Maynard, 2010). Interviewers may introduce errors during the production of sampling frames, during the sampling process, in the set of respondents, in the actual measures collected, and during the coding and editing processes (see Figure 2.2). In fact, through the years survey methodologists have found interviewer effects for all of these error sources, including coverage and sampling error (Eckman, 2013; Eckman & Kreuter, 2011; Tourangeau et al., 2012), nonresponse error (e.g., Blom et al., 2011; Durrant et al., 2013; Jäckle et al., 2013), and measurement and processing error (Durrant et al., 2010; Rice, 1929).

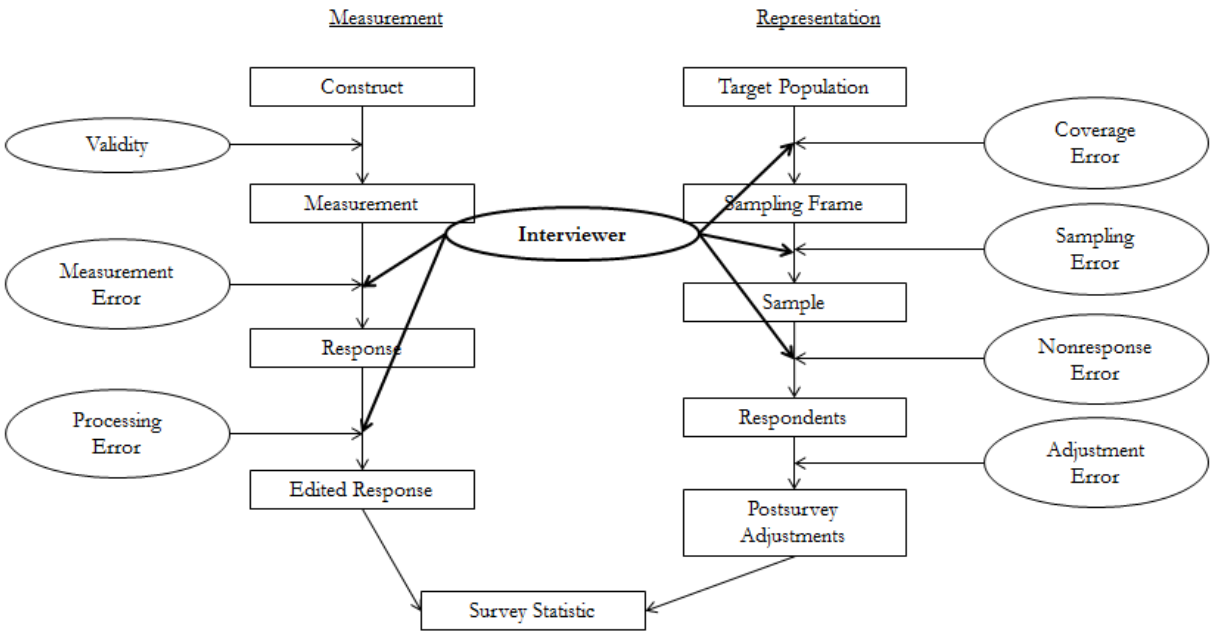


Figure 2.2 Interviewer effects in terms of the Total Survey Error framework, adapted from Groves et al. (2009b, p. 48)

The different errors described in the TSE framework can lead to either *bias* (difference between the expected and the true value) or *variance* (variability of values; Biemer & Lyberg, 2003). The literature on interviewer effects for the different errors described in the TSE framework often focuses on studying interviewer variance instead interviewer bias (see West & Blom, 2017). For example, for analyzing interviewer bias in estimates of substantive survey variables, the true value of an answer given by a respondent is required as a reference, whereas interviewer variance analyses can be studied with the collected survey data.

In my dissertation, I focus on describing and explaining *interviewer variance* components for both parts of the TSE framework: measurement and representation. More specifically, I examine interviewer effects on survey unit nonresponse and estimates of substantive survey variables, the two types of survey error described in the TSE in which the German PIAAC interviewers were involved.

3 PIAAC Germany 2012 and the German PIAAC Interviewers

To date, interviewer effects have been studied in many countries, surveys and with a focus on every possible aspect of the TSE framework (for an overview see West & Blom, 2017). However, there are only very few studies that simultaneously look into interviewer effects on multiple aspects of the survey process described in the TSE framework. This is partially due to the fact that sufficient data on multiple error sources as well as on interviewer characteristics are seldom available in a single survey (for a notable exception see Brunton-Smith et al., 2012). In addition, analyzing just a single error source is usually already sufficiently relevant to warrant publication.

In my dissertation, I study interviewer effects on multiple aspects of the TSE using data from the German implementation of PIAAC. PIAAC is a large-scale face-to-face survey, which aims to compare adults´ competencies across countries in three major domains: *Literacy*, *numeracy*, and *problem solving in technology-rich environment*s. PIAAC was initiated by the Organization for Economic Co-operation and Development (OECD). To date, the survey has been conducted in more than 40 countries around the world, and a second survey cycle is planned to be fielded in 2021.

For my analyses of interviewer effects, I use the data from the PIAAC survey interview as well as from various auxiliary data sources (for an overview see Table 3.1): Data from the PIAAC sampling frame, data from the PIAAC contact protocols, data on interviewers' behavior with regard to deviations from standardized interviewing techniques derived from audio recordings of the PIAAC survey interviews, data on characteristics of PIAAC interviewers collected via an interviewer survey, and data from official area statistics[2]. Having access to such rich data sources

[2] The official area statistics stem from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (2015). This data source offers a large amount of information, which is free of charge and available online. They are mainly based on official statistics from the Federal and State authorities. For my analyses I used the following indicators available for each sample person: Share of Germans, share of single-person households and unemployment rate.

is rare and enables me, on the one hand, to analyze interviewer effects on individual aspects of survey quality, as well as to conduct combined analyses of interviewer effects on survey unit nonresponse and estimates of substantive survey variables. On the other hand, the data enables me to assess whether interviewer characteristics explain any identified interviewer effect.

Table 3.1 Interviewer effects in PIAAC Germany: dependent, explanatory, and control variables

Variables	Data	Data source
Dependent variables	Contact and cooperation outcomes	Contact protocols
	Every day skill use (literacy reading, literacy writing, numeracy, ICT), about yourself (learning strategies, cultural engagement, political efficacy, social trust), number of books, proficiency scores (literacy, numeracy)	Survey interview
Explanatory variables	Interviewers' age, gender, education, occupational status, experience, working hours, expectations, personality, attitudes, behavior	Interviewer survey
	Coded data on interviewers' behavior regarding collecting administrative data, incorrectly skipped questions, incorrectly read questions	Audio recordings
Control variables	Respondents' age, gender, nationality, PSU size, region	Sampling frame
	Respondents' age, gender, nationality, education, work status, health, social background	Survey interview
	Share of Germans, share of single person households, unemployment rate	Official area statistics

Notes. ICT = information and communication technology. PSU = primary sampling unit.

In this chapter, I describe aspects of PIAAC Germany with a special focus on interviewers' role in PIAAC Germany and introduce all data sources that I use. A more detailed description of the international PIAAC study can be found in the international technical report (OECD, 2016), while the implementation of PIAAC in Germany is described in the German national technical report (Zabal et al., 2014).

3.1 PIAAC Germany 2012: A General Overview and Data Description

PIAAC Germany is based on a random sample of persons who are between 16 and 65 years old and live in private households. The German sample was drawn from local community registers via a two-stage sampling process (with a sample of municipalities at the first stage and a

sample of persons clustered within municipalities at the second stage). Overall, 10,240 persons were sampled from the registers within 320 primary sample units (PSUs) and 277 municipalities for PIAAC Germany. The sample persons were processed by 129 interviewers from the German survey organization TNS Infratest Sozialforschung during an 8-months data collection period between August 2011 and March 2012. Most interviewers worked in two or three PSUs and received 32 addresses per PSU. Due to organizational arrangements, a few interviewers worked in only one PSU, while others in up to five PSUs. The overall interview duration was 1 hour and 40 minutes on average and respondents received 50 Euro conditional cash incentive for a completed interview. According to the response rate definition in PIAAC, the overall response rate for Germany is 55% with 5,465 completed cases according to the survey definition interviews (Rammstedt et al., 2013).

3.1.1 Sampling frame

The German PIAAC sample was drawn from German municipal registers. In addition to the names and addresses, municipal registers typically give access to the following information for all the sampled persons: Age, gender and nationality. Furthermore, the sampling frame data contains information on the PSU: PSU size and region. For analyses of interviewer effects, data from population registers is attractive, because the data is available for all sample persons, i.e. respondents and nonrespondents, and thus allows unit nonresponse analyses. Furthermore, the data is usually complete. Indeed, in the case of the PIAAC Germany data, only 64 cases had missing age information (0.6% of the full sample of 10,240 cases), while the sampling frame data was complete regarding gender, nationally, PSU size and region.

3.1.2 Contact protocols

The German PIAAC interviewers were instructed to fill out an extensive contact protocol – the so-called case folder – for each sample person. The contact protocols contained a long list of questions, such as information on sample persons' dwelling units, each contact attempt, disposition codes, and information on completed interviews. One of the key objectives of this

dissertation is analyzing interviewers' success during the recruiting phase, operationalized by two dependent variables: *Contact* and *cooperation*. The data on whether contact or cooperation is derived from the disposition codes collected via the PIAAC contact protocols. A detailed description of these two key variables is presented in Chapter 4.

3.1.3 Survey interview

The PIAAC survey interview consists of two parts: A core questionnaire and a test of key information-processing cognitive competencies. Both survey interview parts are fully computer-based, which is one innovative aspect of the implementation of PIAAC. However, respondents without sufficient computer experience could work on the test items using a paper-and-pencil version.

The core questionnaire was administered by interviewers by means of computer-assisted personal interviewing (CAPI). This part of the interview contains questions on socio-demographic characteristics such as education or work history, participation in education and training, economic and social outcomes, as well as information on the use of skills at work and in private life. In contrast, the cognitive test was completed by the respondents themselves by means of computer-assisted self-interviewing (CASI) under the supervision of the interviewers. Interviewers were instructed to not help respondents and to only intervene if, for instance, technical problems occurred (Zabal et al., 2014). Interviewers had to adapt their behavior during the cognitive test, because they had to learn to be more passive in their role as test administrators.

The cognitive test contained items measuring every-day competencies in the domains *literacy*, *numeracy*, and *problem solving in technology-rich environments* (PS-TRE). The domain PS-TRE was measured computer-based only: Around 20% of respondents did not work on the PS-TRE as they had no experience in operating computers or were not able to work properly with a computer mouse (Rammstedt et al., 2013). Since complete cognitive test data is only available for numeracy and literacy, I focus on these two domains as dependent variables when examining the cognitive test data. (Sample items of the two domains literacy and numeracy can be found in

Appendix A.) The values of respondents' literacy and numeracy competency scores range from 0 to 500. The competency scores represent a continuum of the respective competency domain, with higher values indicating higher competencies.

For more information on the survey design see the international technical report (OECD, 2016). A detailed description of the substantive survey variables from the CAPI and the CASI parts of the PIAAC survey interview that I use to assess interviewer effects on unit nonresponse and substantive survey variables is presented in the respective Chapters 4, 6, and 7.

3.2 The German PIAAC Interviewers: Tasks, Quality Control, Training, and Data

3.2.1 Interviewers' tasks, quality control, and interviewer training in PIAAC Germany

As discussed in the previous chapter, interviewers have many tasks when implementing a survey. This also applies to PIAAC Germany. More specifically, in PIAAC Germany, interviewers had the first in-person contact with sample persons. To increase the chance of a successful contact attempt with a sample person, they were asked to make at least 4 in-person contact attempts per case before they could close a case as a noncontact. In addition, interviewers had to attempt contacting the sample persons during different days of the week and times of day. After attempting contact, they tried to gain the sample persons' cooperation to participate in PIAAC. They were instructed to make use of several pieces of information, which were developed to help gaining the sample persons' cooperation. Once a sample person agreed to participate in PIAAC, interviewers had to administer the complex survey instrument, including an interviewer-administered core questionnaire and a self-completion cognitive test for respondents. In addition, they were responsible for transferring the data and hand in extensive documentation materials on the recruitment process and the survey interview for each case they worked on.

To ensure that the PIAAC data achieved high quality, specific and comprehensive technical standards and guidelines were defined by an international consortium (OECD, 2014). Each country participating in PIAAC had to comply with these standards when carrying out PIAAC. The implementation of the standards was monitored very closely by the consortium for each

country participating in PIAAC, and every single deviation or variation had to be approved in detail. Next to many other aspects, quality control of interviewers' work was specified in the technical standards and guidelines (see also Massing, Ackermann, Martin, Zabal, & Rammstedt, 2013). These standards and guidelines included the following instructions for quality control of interviewers' work: First, interviewers had to be made aware that their work was monitored on a regular basis. Second, at least 10% of each interviewer's work had to be validated (including nonrespondents). Third, for every interviewer, two audio recordings of the PIAAC interviews had to be collected and checked with respect to interviewers' behavior, regarding deviations from standardized interviewing protocols. In addition to the international requirements, national best practice quality assurance processes were implemented for PIAAC Germany. For example, the national team and the survey institute looked at overall patterns of interviewers' performance and checked the plausibility of the date and time of all interviews.

According to the literature, detailed interviewer trainings as well as adequate interviewer payment can prevent interviewer effects in surveys (e.g. Billiet & Loosveldt, 1988; Groves & McGonagle, 2001). Thus, in addition to all post-hoc quality control assurance, it is important to direct effort towards preventing interviewer effects. Thus, for PIAAC Germany various strategies to prevent interviewer effects were applied. First of all, extensive 5-day in-person interviewer trainings were conducted. Second, a special interviewer payment scheme was developed to achieve an adequate level of interviewer payment. Third, interviewers received a take-home reference manual containing comprehensive information on the survey and its implementation. In addition, the survey organization TNS Infratest Sozialforschung was instructed to only select experienced, high-quality interviewers to work for PIAAC. The selection criteria are described in Zabal et al. (2014).

3.2.2 Audio recordings of German PIAAC interviews

Having access to detailed information on interviewers' actual behavior with respect to deviations from standardized interviewing during a survey interview is rare. For PIAAC Germany

such data is available, because the international standards and guidelines of PIAAC demanded two audio recordings to be collected and checked during quality control for each interviewer. Despite of the official instructions, audio recordings are not available for eight interviewers working on PIAAC Germany.

The audio recorded interviews were coded by means of simple coding schemes to detect deviations from standardized survey interviewing for three major topics: administration, completeness, and probing. The coding schemes as well as the descriptive results of the interviewer behavior based on this coding scheme are presented in Chapter 5.

In order to explain interviewer effects on survey unit nonresponse and estimates of substantive survey variables by means of interviewers' actual behavior, I further develop four indicators for interviewers' behavior based on the simple coding scheme described in Chapter 5: (1) collecting formal criteria correctly with regard to the date of the interview, the interviewer ID and the respondent ID, (2) collecting informed consent with regard to the permission to record the interview from the respondent, (3) the number of incorrectly skipped questions, as well as (4) the number of incorrectly read questions. Descriptive statistics for all four indicators can be found in Appendix B. In total, only one third of the interviewers made no mistake in collecting formal information. When collecting informed consent, even fewer interviewers (19%) made no mistakes during interviewing. On average, 160 questions were asked per case and, of those, about 1.5 questions were skipped incorrectly on average by the PIAAC interviewers. With regard to the incorrect reading of questions, 7.5 questions were not read correctly, on average, with a maximum of 47 incorrectly read questions for one interviewer.

3.2.3 The German PIAAC interviewer survey

In PIAAC Germany, an interviewer survey was implemented to collect more information on interviewers' characteristics. The main purpose was to identify interviewer characteristics that explain interviewer effects on survey unit nonresponse and on estimates of substantive survey variables.

The question text of the interviewer questionnaire used in this dissertation can be found in Appendix C. The questionnaire was adapted from an interviewer questionnaire implemented in the Survey of Health, Ageing and Retirement in Europe (SHARE) in 2011 (Blom & Korbmacher, 2013). Interviewers' participation in the interviewer survey was voluntary and interviewers did not receive any incentive. Overall, the interviewer survey yielded a response rate of 89.8% with 115 interviewers sending back the questionnaire via mail.

This chapter provides an overview of the interviewers characteristics collected with the German PIAAC interviewer survey as well as of the indicators derived from the item lists in the questionnaire.

Imputation

To improve the statistical power of the analyses conducted with the interviewer survey data and correct for item nonresponse which is *missing at random* (MAR) or *missing completely at random* (MCAR; Little & Rubin, 1987), I impute missing data. For this purpose, I use single imputations with either linear regression, ordered logistic regression or multinomial logistic regression, depending on the level of measurement of the respective variables. Overall, about 1.8% of the interviewer survey data is missing at the item level from the interviewer survey data set. I use the Little`s MCAR-Test (Little, 1988) to test whether the missing values are MCAR. The MCAR-Test by Little is a chi-square test for missing data. The null hypotheses means that the data is missing completely at random and the alternative hypotheses states that the data is not missing completely at random. The results of the test statistics $\chi^2 = 979.46$, $p = .16$ indicate that the null hypotheses cannot be rejected, which means that there is no clear pattern in the missing data or in other words, this indicates that the data is missing completely at random (MCAR).

Interviewer socio-demographic characteristics, experience, working hours, and expectations

The interviewer questionnaire contains several questions about interviewers´ socio-demographic characteristics as well as about their experience, working hours, and expectations. A

summary of these characteristics and expectations of the German PIAAC interviewers is provided in Table 3.2a and Table 3.2b.

Table 3.2a The interviewer survey: Characteristics of the interviewers

		n	Percent
Gender	Male	62	53.91
	Female	53	46.09
	Total	115	100.00
Age	<= 45 years	10	8.70
	45 - 60 years	44	38.26
	>= 60 years	61	53.04
	Total	115	100.00
Abitur	Yes	52	45.22
	No	63	54.78
	Total	115	100.00
Employed (full-/part-time)	Yes	15	13.04
	No	100	86.96
	Total	115	100.00
Retired	Yes	43	37.39
	No	72	62.61
	Total	115	100.00
Prior PIAAC experience	Yes	35	30.43
	No	80	69.57
	Total	115	100.00
Work experience	<= 5 years	41	35.65
	6 - 10 years	21	18.26
	>= 11 years	53	46.09
	Total	115	100.00
Working hours per week	<= 15 hours	10	8.70
	16 - 30 hours	69	60.00
	>= 31 hours	36	31.30
	Total	115	100.00

Table 3.2b The interviewer survey: Expectations of the interviewers

	Mean	SE	Min	Max
Expectation: Overall cooperation rate in percent	50.31	12.04	20	80
Expectation: Item response rate - income in percent	82.84	17.51	15	100

Notes. Tables based on 115 interviewers. SE = standard error. Min = minimum. Max = maximum.

Interviewers' attitudes, behavior, and personality

Interviewers´ attitudes and behavior were collected with an extensive questionnaire containing several long item lists. To reduce complexity, I construct a limited list of latent variables (indicators) from the interviewer characteristics which were measured via the this interviewer survey for the following sub-dimensions:

- Reasons for working as an interviewer
- How to conduct standardized survey interviews
- How to achieve response
- Trust
- Social desirability
- Data protection concerns

First, I recode all reversed items, such that a high score indicates a high level of the measured attitude and behavior to allow an intuitive interpretation. For easier interpretation and to account for correlations among items after performing factor analyses, the factors are rotated using *promax* rotation. Subsequently, I calculate Cronbach's Alpha for each indicator to gain more information about the reliability of the respective indicator. Finally, I construct indicators by simply calculating the mean of the items included in the indicator. Within the next paragraphs, I describe all indicators in detail.

Reasons for working as an interviewer

The interviewer questionnaire contains seven items on an interviewer's reason for working as an interviewer. No items are inversely coded and, thus, no items have to be recoded. For all items, a high score indicates the particular reason is very important to the interviewer. The factor analysis of the item list statistically shows a four factor solution. However, no item loads on the fourth factor (see Appendix D). Thus, I choose to build three indicators. I construct the three indicators based on the means of all items that load on the respective factor. The resulting indicators are the following:

- Indicator 1: *Science – work for something related to science / research*
- Indicator 2: *People – socialize with other people and interest in life of others*
- Indicator 3: *Formal – formal conditions, like payment*

The first indicator *science* is based on two arguments on working for research or science related projects. The second indicator *people,* describes reasons with regard to socializing with other people and the third indicator *formal* refers to formal arguments to work as an interviewer. Cronbach's Alpha shows an acceptable reliability for all three indicators as they are all based on a small number of items (see Appendix D). For an overview of the resulting indicators and questions see Table 3.3, descriptive statistics can be found in Appendix D.

Table 3.3 The interviewer survey: Indicators on interviewers' attitudes, behavior and personality

Topic	Indicator	Question
Reasons for working as an interviewer	Science	Research Society
	People	Interest Socialize Insight
	Formal	Payment Time
How to conduct standardized survey interviews	TailorContent	Explaining Wording
	TailorTime	Shortening Completing Fast
	TailorNo	Instructions Interrupt Selfadmin
How to achieve response	Motivate	Persuasion Contacting
	Diligent	Effort Righttime Confident
	Voluntary	Privacy Refusal Voluntary
Trust	Trust	Trust1 Trust2
Social desirability	Self	Impression Judgement Preference
	Others	Change Honesty Player

Note. For descriptive statistics for all indicators of interviewers attitutes, behavior and personality see Appendix D.

How to conduct standardized survey interviews

One item list contains ten items about rules of standardized survey interviewing. Out of these ten items, three items have to be recoded as they are inversly coded: *Wording*, *slow* and *instructions*. All items are coded such that a high score indicates a high level of agreeing. Statistically, the factor analysis of the item list shows a five factor solution with only one item loading highest on factor four and one item loading highest on factor five (see Appendix D). Thus, I construct three indicators based on the mean of all items that load on the respective factor. The resulting indicators and single items for analyzing interviewer effects in PIAAC from the item list on how to conduct standardized survey interviews are the following:

- Indicator 1: T*ailorContent – tailoring the content, because interviewer thinks s/he can do it better*
- Indicator 2: T*ailorTime – tailoring by reduce the questionnaire time to "help" the respondent*
- Indicator 3: T*ailorNo – no tailoring but closely following instructions*
- *Slow*
- *Dialect*

Factor one includes four variables and is split into two factors to better represent interviewers´ behavior with regard to *not* following standardized interviewing techniques: The first indicator *tailorContent* refers to tailoring question text because some respondents have problems understanding the question. Instead, the second indicator *tailorTime* refers to tailoring question text to reduce the questionnaire time. Factor three *tailorNo* is computed according to the five factor solution. It includes items describing interviewer behavior following the instructions of standardized interviewing. Cronbach's Alpha shows an acceptable reliability for the three outcome indices (see Appendix D). Lastly, two items (*slow*, *dialect*) are used in their original version from the questionnaire. For an overview of the resulting indicators and corresponding questions see Appendix D.

How to achieve response

The interviewer questionnaire contains nine items on possible strategies on how to achieve response of sample persons. Out of the nine items, three items have to be recoded as they are inversly coded in the original questionnaire version: *persuasion*, *effort,* and *righttime*. All items are then coded so that a high score indicates a high level of agreeing to the statement of the item. The factor analyses of the item list statistically show a five factor solution with only one item loading on factor four and one item loading on factor five (see Appendix D). The latter (*effort*), has a high loading on factor two as well as on factor five. I construct three indicators based on the mean of all items that load on the respective factor. I add the item *effort* to factor two, as Cronbach's Alpha shows an acceptable value for adding the item *effort* to factor two.

Furthermore, this makes most sense considering the content of the item, instead of including this item individually in the models.

Indicator one, *motivate,* refers to strategies of persuading reluctant sample persons by being motivated. Indicator two, *diligent,* describes strategies of achieving response by being diligent. Indicator three *voluntary* refers to items describing respectful behavior regarding the voluntariness of sample persons' willingness to participate. Finally, one item (*answers*) was used in its' original questionnaire version. For an overview of the resulting indicators see Table 3.3.

- Indicator 1: M*otivate – persuade reluctant sample persons*
- Indicator 2: D*iligent – plenty of work pays off*
- Indicator 3: V*oluntary – respect voluntariness*
- *Answers*

Trust, social desirability and data protection concerns

The interviewer questionnaire contains two questions with regard to trust. Cronbach´s Alpha for this indicator is very high (.77). Thus, I use the trust indicator based on the mean of both items.

A short scale of six items measuring social desirability was implemented in the interviewer questionnaire (Winkler, Kroh, & Spiess, 2006). Three out of these six items have to be recoded as they are inverse coded in the questionnaire: *Judgement*, *change* and *player*. All items are then coded so that a high score indicates a high level of agreeing to the respective item. Statistically, the factor analysis of the items measuring social desirability indicates a two factor solution (see Appendix D). The statistical factor solution is in accordance with Winkler et al. (2006). Thus I construct two indicators based on the mean of the three items per indicator.

- Indicator 1: S*oc_desire_self – social desirability with regard to one self*
- Indicator 2: S*oc_desire_others – social desirability with regard to others*

One question of the interviewer questionnaire measures the general data protection concerns of interviewers´ own data. I recode this ordinal scaled question as a dummy variable for my analyses.

4 Study 1: Interviewer Effects on Survey Unit Nonresponse in PIAAC Germany

There are numerous reasons why sample persons may or may not participate in surveys. In order to minimize survey unit nonresponse it is important to understand why some interviewers are more successful at recruiting sample persons than others. Thus, in this chapter, I focus on explaining interviewer effects on the two key aspects of survey unit nonresponse by means of interviewer characteristics: Successfully making contact with sample persons and successfully gaining sample persons' cooperation.

This chapter commences with a literature review of interviewer effects on survey unit nonresponse with a focus on explaining interviewer effects with interviewer characteristics. Subsequently, I describe the data and methods used to analyze interviewer effects on survey unit nonresponse in this chapter. Then I present and discuss two major research questions: First, I aim to answer the question of how the propensity of a sample person to be successfully contacted and to take part in PIAAC Germany differs across interviewers. I answer this question by analyzing the amount of variance introduced by interviewers with regard to these two aspects of the recruitment process for PIAAC Germany. Second, I aim to explain some of the observed differences across interviewers during the recruitment process by means of explanatory factors related to sample composition and interviewer characteristics. In doing so, I investigate to what extent sample composition characteristics (e.g. sample persons` age and region) or interviewer characteristics (e.g. interviewers' age and working experience) affect a sample persons' propensity to be successfully contacted and a sample persons' propensity to participate in PIAAC Germany.

I find considerable interviewer effects on a sample persons' propensity to be successfully contacted for PIAAC Germany, which I can only partly explain by differences in interviewers' characteristics, such as interviewers' socio-demographic characteristics or behavior. Area clustering affects a sample persons' propensity to be successfully contacted to almost the same

amount as interviewer clustering, even after controlling for sample composition effects. In addition, I find very small interviewer effect on successfully gaining sample persons' cooperation for PIAAC Germany. This interviewer effect is so small that it does not warrant further analyses.

4.1 Previous Research

There is a long history of studies evaluating the influence of interviewers on survey unit nonresponse. Overall, more studies look into noncooperation than into noncontact. Theoretically, the skills to successfully make contact with sample persons and those to gain sample persons' cooperation differ considerably. According to Brunton-Smith et al. (2012) "gaining cooperation is essentially a matter of using persuasive verbal and non-verbal interactional techniques (Morton-Williams, 1993) and understanding and reacting appropriately to the sorts of objections householders are likely to raise (Campanelli, Thompson, Moon, & Staples, 1997; Groves & Couper, 1998), successfully making contact is more a case of organizing ones time effectively, following best-practice guidelines and being persistent." (p.270) Nonetheless, many studies report that interviewers who are successful in gaining cooperation are also successful in making contact with sample persons (Blom et al., 2011; O'Muircheartaigh & Campanelli, 1999; Pickery & Loosveldt, 2002; Vassallo, Durrant, Smith, & Goldstein, 2015). These results indicate that good survey interviewers have both sets of skills to successfully make contact and gain sample persons' cooperation.

Following West and Blom (2017) I structure my literature review of interviewer effects on survey unit nonresponse according to the following interviewer characteristics: Socio-demographics, experience, attitudes and personalities, and skills and behavior.

Socio-demographic characteristics

Most studies that look into interviewer effects on survey unit nonresponse and take socio-demographic interviewer characteristics into account focus on gender and age: This is mainly the case, because information on interviewers' age and gender is typically available from interviewer records. For example, Groves and Couper (1998) or Morton-Williams (1993) find a weak

relationship between interviewers' age and gender and sample persons' willingness to cooperate. Blom et al. (2011) find older interviewers to be better at gaining sample persons' cooperation, but they assume this might be confounded with experience. Furthermore, a match of interviewers' and sample persons' socio-demographic characteristics has a positive effect on survey unit nonresponse (Brehm, 1993; Durrant et al., 2010; Lord, Friday, & Brennan, 2005; Moorman, Newman, Millikan, Tse, & Sandler, 1999; Webster, 1996).

Experience

In general, interviewers with more experience have been found to attain higher survey contact and higher survey cooperation rates (Campanelli & O'Muircheartaigh, 1999; M. P. Couper & Groves, 1992; Durbin & Stuart, 1951; Durrant et al., 2010; Jäckle et al., 2013). Interviewers' survey-specific experience seems to be even more important in explaining unit nonresponse than their general job experience (Lipps & Pollien, 2011). However, both effects might be confounded through self-selection (Jäckle et al., 2013): Interviewers who are more successful stay longer in the job and thus are the ones who are more experienced. Furthermore, the effect of interviewer experience might also be confunded with interviewers' interaction skills, meaning that interviewers who are better at tailoring the own behavior to sample persons' needs, stay longer in the job and thus are more experienced (Beerten, 1999; Groves & Couper, 1998; Jäckle et al., 2013; Lemay & Durand, 2002).

Studies attempting to explain the effect of interviewer experience on survey unit nonresponse show diverse results: As already mentioned above, Blom et al. (2011) report that older interviewers are more successful in recruiting sample persons and suggest this effect to be confounded with interviewers' experience. Jäckle et al. (2013) find higher cooperation rates for more experienced interviewers which they can partly explain by differences in personality traits (e.g. extroversion, openness), skills (verbal communication, adaptability, assertiveness and deliberation) and attitudes of interviewers. When controlling for sample person and interviewer characteristics, Blohm, Hox, and Koch (2007) find that the effect of experience on a sample

persons' willingness to cooperate disappears. Further explanations suggest that an interviewer's workload may be related to their success in recruiting sample persons to participate in surveys: Interviewers who have been shown to have higher response rates might be allocated harder cases (Blom et al., 2011) or even more cases to work on (Japec, 2008; Pickery, Loosveldt, & Carton, 2001; Singer, Frankel, & Glassman, 1983) which could reduce their success as a result the other way around. Overall, the effect of interviewer experience is still unclear.

Attitudes and personality

Older research on interviewer effects with regard to survey unit nonresponse has focused on observable interviewer characteristics and not as much was known about the effect of interviewers' attitudes and personality. In recent years, more has been discovered about the effects of interviewer attitudes regarding surveys and interviewers' personality. In general, an interviewer's positive attitude with regard to gain sample persons' cooperation results in higher cooperation rates (Blom et al., 2011; de Leeuw, Hox, Snijkers, & de Heer, 1998; Hox & de Leeuw, 2002; Jäckle et al., 2013; Lehtonen, 1996; Maynard & Schaeffer, 2002). Indeed, a positive attitude could also be the result of better outcomes and greater success in recruiting sample persons during interviewers' previous work. Therefore, a circular effect, where success affects attitudes, which in turn affect success, cannot be ruled out.

Another aspect of interviewers' personality, the role of interviewer confidence, has been studied: More confident interviewers and interviewers with a professional self-image achieve higher cooperation rates (Blom et al., 2011; Durrant et al., 2010). This relationship is even stronger for interviewers with more experience (Jäckle et al., 2013) and during a more complex survey (Durrant et al., 2010). Furthermore, Jäckle et al. (2013) try to explain interviewer effects on survey unit nonresponse by means of the Big 5 personality traits. They find support of their hypotheses regarding the association of personality traits and interviewers success in gaining sample persons cooperation: Greater extroverted interviewers achieve higher cooperation rates. In addition, they find that greater openness and agreeableness is negatively related to

interviewers' success in gaining sample persons' cooperation. Next to that; Vassallo et al. (2015) found no or only weak relationships with interviewer personality traits such as tendency to worry, complacency of survey message, awareness self-portrayal, and conscientiousness and wave participation in a panel study.

Overall, the results of interviewers' personality traits on survey unit nonresponse leave many questions open and should be studied in further research.

Behavior

Jäckle et al. (2013) assume that the interviewer has both, an active and a passive influence on the sample person's decision to participate in a survey: The passive influence is due to the sample person's perception of the interviewer, while the interviewers' behavior actively influences the sample person's decision to participate. Jäckle et al. (2013) suspect that there are three reasons why interviewer behavior is often not found to be predictive of cooperation: 1) statistical power: usually the number of interviewers working on a project is limited; 2) measurement: Interviewers forget "exact components of interaction" when they are asked to report them, and 3) level of measurement: Typically interviewers are asked about their usual behavior with regard to tailoring, even though individualized interaction might be more relevant than a general result of how interviewers behave.

Other studies, that use audio recordings to analyze interviewer behavior on survey unit nonresponse, report that interviewers who introduce themselves and tailor responses to reluctance achieve higher cooperation rates, whereas interviewers who are less successful in gaining sample persons' cooperation give sample persons too much room to escape (Groves & McGonagle, 2001; Morton-Williams, 1993; Schaeffer, Garbarski, Freese, & Maynard, 2013; Snijkers, Hox, & de Leeuw, 1999). In addition, a moderate voice speed is found to have an positive effect on sample persons' willingness to cooperate (Conrad et al., 2013; Groves, O'Hare, Gould-Smith, Benkí, & Maher, 2008; Oksenberg & Cannell, 1988; Oksenberg, Coleman, & Cannell, 1986).

However, de Heer (1999) or Durrant et al. (2010) assume the effects of interviewer behavior on survey unit nonresponse to be selection effects: Interviewers who are successful at recruiting sample persons for a survey might have more skills or more experience in doing so (Campanelli et al., 1997; de Leeuw, 1999; Snijkers et al., 1999; Sturgis & Campanelli, 1998).

Interviewer versus regional effects

In face-to-face surveys, interviewers typically work in a limited number of areas (Primary Sampling Units; PSUs). Thus, effects of the interviewer and the area on survey unit nonresponse are confounded. The best way to disentangle interviewer and area effects on survey unit nonresponse is to implement a *fully interpenetrated* design, meaning that sample persons are assigned to interviewers at random (Groves et al., 2009b; O'Muircheartaigh & Campanelli, 1999). Due to several reasons, such as limited costs or a limited number of available interviewers within an area, this is often not possible for face-to-face-surveys. O'Muircheartaigh and Campanelli (1999) implement a *partially interpenetrated* survey design to disentangle interviewer from area effects by pooling geographic areas and randomly assigning at least two interviewers to sample persons. They report that interviewer effects on unit nonresponse are more important than area effects. In contrast, they find area effects to be more important on successfully making contact with sample persons.

There is another alternative to account for a *non-random assignment* of sample persons to interviewers in surveys: Jäckle et al. (2013) included variables in their models to control for sample composition effects. They show that 1/3 of the variation in unit nonresponse can be explained by the non-random assignment of interviewers to areas and surveys. In addition they find that some variation can also be explained by interviewer characteristics.

4.2 Data Description

The analyses of interviewer effects on survey unit nonresponse in this chapter are based on data from the German implementation of PIAAC 2012 (Rammstedt et al., 2014). For my analyses, I am in the luxurious position of having access not only to the PIAAC gross sample,

including data from the sampling frame, contact protocols, and the survey interview, but also to a variety of characteristics of the German PIAAC interviewers collected via a comprehensive interviewer survey. I described PIAAC, as well as the data on interviewers' characteristics, in detail in the general data section in Chapter 3. In addition, I briefly describe the data used for the analyses in the present chapter in the next paragraphs (for an overview see Table 4.1 and Table 4.2).

Table 4.1 Interviewer effects on survey unit nonresponse: dependent, explanatory, and control variables

Variables	Data	Data source
Dependent variables	Contact and cooperation outcomes	Contact protocols
Explanatory variables	Interviewers' age, gender, education, occupational status, experience, working hours, expectations, personality, attitudes, behavior	Interviewer survey
Control variables	Respondents' age, gender, nationality PSU size, region	Sampling frame
	Share of Germans, share of single person households, unemployment rate	Official area statistics

Note. PSU = primary sampling unit.

Table 4.2 Interviewer effects on survey unit nonresponse: sample description

	Contact	Cooperation
Eligible sample	9,369	9,369
Analyses sample	7,902	7,450
Number of interviewers	115	107
Number of PSUs	251	240

Notes. Analyses sample *contact*: Excluding cases with missing data on explanatory or control variables and cases approached by more than one interviewer. Analyses sample *cooperation*: Excluding cases with missing data on explanatory or control variables, cases approached by more than one interviewer and cases where no contact was successfully made. PSU = primary sample unit.

For PIAAC Germany, 10,240 persons were sampled from the council registers. However, the gross sample of PIAAC consists of 9,369 eligible sample persons, because 871 sample persons are defined as ineligible. After some cases with missing data[3] and cases approached by more than one interviewer[4] are excluded, the analyses sample for the two dependent variables

[3] One possibility to deal with missing data is to assign values to missing items. However, imputation should be based on reliable information about the case for which the data is missing (Lohr, 1999). For the cases mentioned above no or almost no information is available. Thus I do not impute data for these cases.

[4] Survey organizations tend to re-assign difficult cases to their most productive and most experienced interviewers, a practice that may bias results of interviewer-effects analyses, if unaccounted for. It is not possible to disentangle

consists of 7,902 sample persons for *contact* and 7,450 sample persons for *cooperation* (see Table 4.2).

Control variables: Sample composition characteristics

The primary objective of this chapter is to explain interviewer effects during the recruitment phase, in particular, on successfully making contact with the sample person and successfully gaining the sample person's cooperation. In PIAAC, sample persons were not randomly assigned to interviewers. Instead sample persons are clustered within PSUs and PSUs are clustered within interviewers (see method section of this chapter for a detailed description). This non-random assignment of interviews to sample persons may lead to differences in key characteristics in the set of sample persons that an interviewer is assigned to interview. To account for this non-random assignment of sample persons to interviewers, I take sample composition effects into account. More specifically, I control for the following information derived from the sampling frame and from the Federal Institute for Research on Building Urban Affairs and Spatial Development (2015): Sample persons' age, gender and nationality as well as PSU size and region, share of Germans, share of single-person households and unemployment rate. In addition, I include sample persons' age squared in my models, to take a possible (non-linear) U-shaped relationship between age and contact as well as between age and cooperation into account (Jäckle et al., 2013). For an overview of the data I use, as well as the respective data sources see Table 4.1.

success or failure to just one interviewer, because in most cases where the first interviewer was not successful the outcome of the work done by the second interviewer could be affected by the work of the previous interviewer. To achieve a clean picture of interviewer effects in the recruitment process, I exclude cases that were approached by more than one interviewer.

Explanatory variables: Interviewer characteristics

In the present chapter, I aim to explain interviewer effects during the recruitment process by means of interviewer characteristics that were collected via an interviewer survey conducted with the German PIAAC interviewers. The interviewer survey was answered by 115 interviewers, (response rate: 89.8%). The questionnaire contains information on interviewers' socio-demographic characteristics, attitudes, personality, behavior as well as expectations with regard to field work outcomes of PIAAC (for a detailed description of the interviewer survey see Chapter 3).

For the analyses of interviewer effects on survey unit nonresponse, first of all, I include interviewers' socio-demographic characteristics age, gender, education, and occupational status (next to their interviewer job) in my models. Each of these characteristics are included in my analyses coded as dummy variables, which I described in Table 3.2 in Chapter 3 in detail. Next to interviewer socio-demographic characteristics, I include the indicators on interviewers' attitudes about reasons for working as an interviewer, behavior regarding standardized interviewing techniques or strategies on how to achieve response, and interviewer personality traits in my models: S*cience*, *people*, *formal*, *tailorContent*, *tailorTime*, *tailorNo*, *slow*, *dialect*, *motivate*, *diligent*, *voluntary*, *answers*, *trust*, *self*, and *others*. The description of the construction for all of these indicators and descriptive statistics are provided in Chapter 3.

Dependent variable I: Contact

The first key dependent variable *contact* is a dichotomous variable which is derived from the disposition codes in the contact protocols (see Chapter 3). I define a case as successfully contacted if the interviewer had contact to any resident in the household (for alternative definitions see Blom (2009). According to this definition, any interview, break-off or refusal (by the sample person or any other person), is treated as successful contact.

Overall, the analyses of interviewer effects on the dependent variable contact include 7,902 cases approached by 115 interviewers in 251 PSUs (see Table 4.2). In total, no contact was made

for 285 sample persons. The overall contact rate across all 115 interviewers is 96.39%. Figure 4.1 shows the distribution of contact rates achieved by the PIAAC interviewers. The figure depicts a low variation in contact rates between the interviewers: The interviewer with the lowest contact rate achieved contact in 73.81% of her/his sample persons, whereas 37 interviewers (32%) made contact with 100% of their sample persons. In addition, only three interviewers have a contact rate lower than 85% and 13 interviewers have a contact rate below 90%.

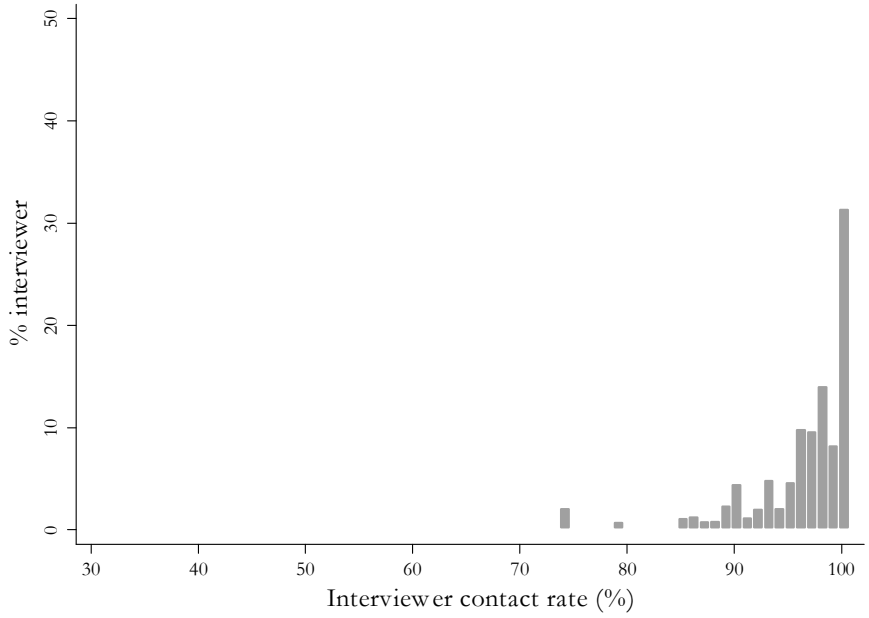


Figure 4.1 Interviewer contact rates by interviewers, PIAAC 2012

Notes. Data based on 115 interviewers. Interviewer contact rate: percentage of cases for which interviewers successfully made contact by total number of sample persons assigned to an interviewer.

Dependent variable II: Cooperation

The second key dependent variable is *cooperation*. Identical to the dependent variable *contact*, the dependent variable *cooperation* is derived using the disposition codes from the PIAAC contact protocols (see Chapter 3). Cooperation is defined according to the "foot-in-the-door" logic, a complience tactic (Freeman & Fraser, 1966): An interviewer was successful at gaining a sample persons' cooperation to participate in PIAAC, if s/he started the interview, independently of whether the interview was completed or broken off later on.

Furthermore, I define cooperation *conditional on* contact, which means that only sample persons for which contact – either to the sample person or to any other person in the household – was successfully made are included in the analyses sample for cooperation. In total, my analyses on interviewer effects on cooperation include 7,450 cases approached by 115 interviewers in 251 PSUs.

In total, 2,890 sample persons were successfully contacted, but no cooperation was achieved with the survey request by the interviewers. The overall cooperation rate across all 115 interviewers is 61.20%. Figure 4.2 shows the distribution of cooperation rates by interviewer. There is considerable variation in cooperation rates across interviewers: The lowest cooperation rate is 33.33%, while the highest cooperation rate is 87.04%.

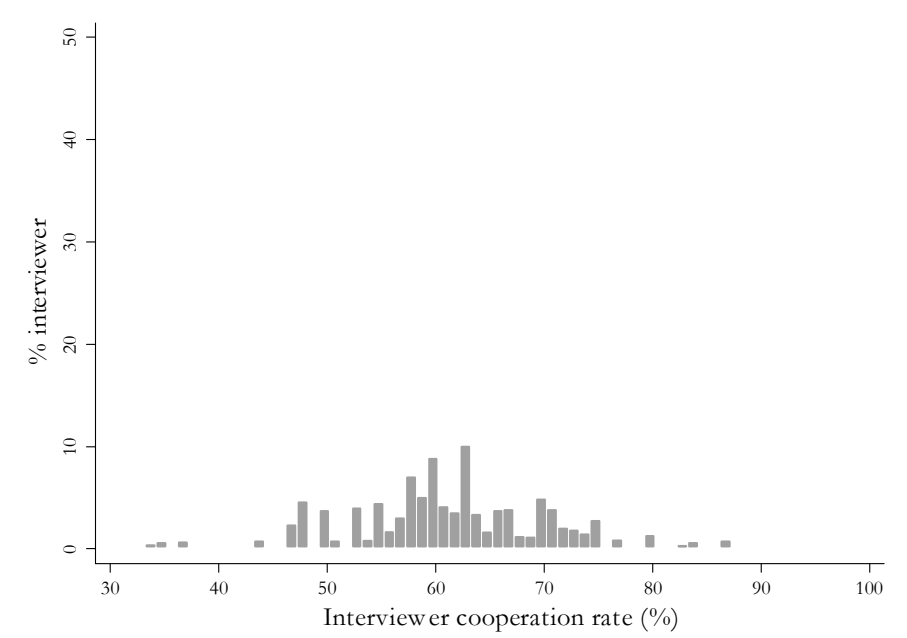


Figure 4.2 Interviewer cooperation rates by interviewers, PIAAC 2012

Notes. Data based on 115 interviewers. Interviewer cooperation rate: percent of cases for which interviewers successfully gained sample persons' cooperation by total number of sample persons.

4.3 Method

I analyze interviewer effects on the two dependent variables *contact* and *cooperation* by means of three-level random effects logistic regression models, thus taking the nested data structure into account: Sample persons (first level) are nested in PSUs (second level) and PSUs are nested in interviewers (third level; see Figure 4.3).

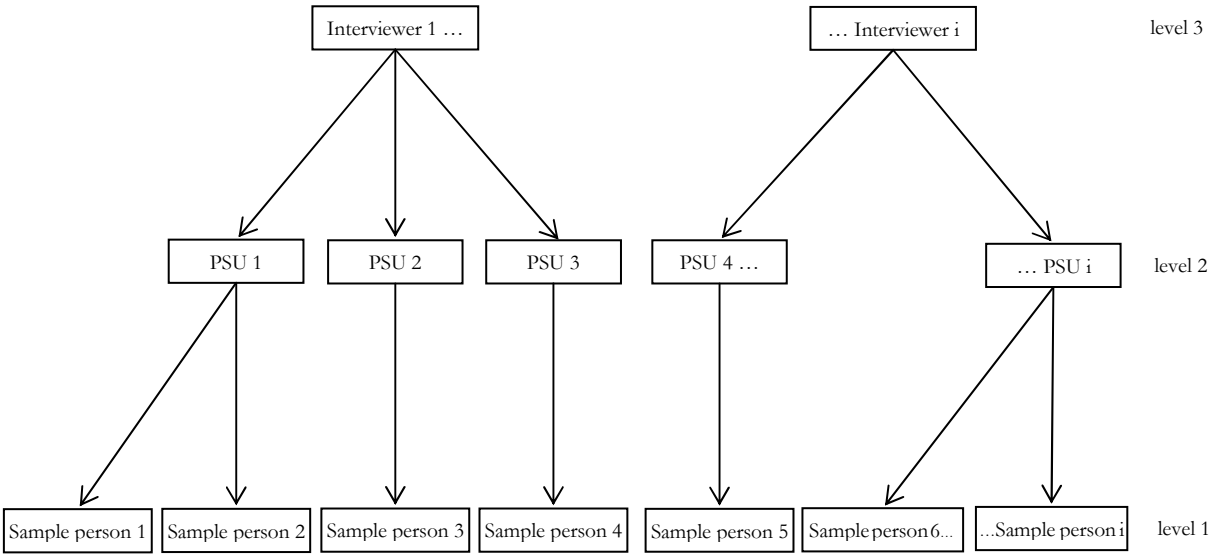


Figure 4.3 Interviewer effects on survey unit nonresponse: Data structure

Note. PSU = primary sample unit.

I tested the necessity to estimate three-level models instead of less complex two-level models. The results of two-level models (see Appendix E) showed a substantial part of the variance of the propensity of a sample person to be successfully contacted can be explained by PSU clustering (in line with findings from Jäckle et al., 2013). Thus, PSU and interviewer clustering have to be taken into account. Results for the second dependent variable cooperation show in the same direction, but are less pronounced than for making contact.

Intercept-only model

Following Hox (2010), I first estimate an intercept only model (null model):

$$\ln\left(\frac{P(Y_{ijk}=1\mid X_i)}{1-P(Y_{ijk}=1\mid X_i)}\right)=\beta_{000}+v_{00k}+u_{0jk}$$

$$\begin{aligned} with \quad & v_{00k}\sim N(0,\sigma^2_{v_{00k}}) \\ & u_{0jk}\sim N(0,\sigma^2_{u_{0jk}}) \end{aligned} \tag{4.1}$$

where the natural logarithm of the propensity P of the outcome Y_{ijk} divided by 1 minus the propensity P of the outcome Y_{ijk} for sample person i nested in PSU j nested in interviewer k is estimated as the intercept of the regression β_{000} and the residuals v_{00k} for the interviewer and PSUs u_{0jk} with the assumption of a normal distribution of the variance over PSUs nested within one interviewer $\sigma^2_{u_{0jk}}$ and the variance over interviewers $\sigma^2_{v_{00k}}$ (see equation 4.1).

In the null model, the Intra-Class Correlation Coefficient (ICC) informs about the proportion of the variance explained by the group variable; i.e. PSUs or interviewers. At each level, ICCs are estimated as shown in equations (4.2), (4.3) and (4.4) (see Hox, 2010). For example, equation (4.2) describes the estimation of the ICC at the interviewer level (level 3)

$$ICC_{\text{int}evriewer_{00k}} = \frac{\sigma^2_{v_{00k}}}{\sigma^2_{v_{00k}}+\sigma^2_{u_{0jk}}+\pi^2/3} \tag{4.2}$$

$$ICC_{PSU_{0jk}} = \frac{\sigma^2_{u_{0jk}}}{\sigma^2_{u_{0jk}}+\sigma^2_{v_{00k}}+\pi^2/3} \tag{4.3}$$

$$ICC_{PSUwithinInterviewer_{0jk}} = \frac{\sigma^2_{u_{0jk}}+\sigma^2_{v_{00k}}}{\sigma^2_{u_{0jk}}+\sigma^2_{v_{00k}}+\pi^2/3} \tag{4.4}$$

where $\sigma^2_{v_{00k}}$ describes the variance between interviewers, $\sigma^2_{u_{0jk}}$ describes the variance over PSUs nested within interviewers, and $\pi^2/3$ is the variance in the logistic regression model at level 1.

Equation (4.3) describes the estimation of the ICC for the proportion of the variance between PSUs and equation (4.4) describes the estimation of the ICC for the proportion of the variance between PSUs clustered within one interviewer.

Full models

To account for sample composition effects, I include the sample composition variables in the full models. In addition, I estimate lots of models where I include one interviewer characteristic per model to the models including the control variables to test whether the respective interviewer characteristics can explain parts of the interviewer variance of the propensity of a sample person to be successfully contacted and to participate in PIAAC. An example of an equation for a full model including two variables V at each level is provided in equation (4.5):

$$\begin{aligned} &\ln\left(\frac{P(Y_{ijk}=1\mid X_i)}{1-P(Y_{ijk}=1\mid X_i)}\right)=\beta_{000} \\ &+\beta_1 \text{sp\_V1}_{ijk}+\beta_2 \text{sp\_V2}_{ijk} \\ &+\beta_3 \text{psu\_V1}_{ij}+\beta_4 \text{psu\_V2}_{ij} \\ &+\beta_5 \text{i\_V1}_{k}+\beta_6 \text{i\_V2}_{k} \\ &+v_{00k}+u_{0jk} \end{aligned} \qquad (4.5)$$

$$\begin{aligned} \textit{with} \quad & v_{00k} \sim N(0,\sigma^2_{v_{00k}}) \\ & u_{0jk} \sim N(0,\sigma^2_{u_{0jk}}) \end{aligned}$$

where the natural logarithm of the propensity *P* of the outcome Y_{ijk} divided by 1 minus the propensity *P* of the outcome Y_{ijk} for sample person *i* nested in PSU *j* nested in interviewer *k* is estimated as the intercept of the regression β_{000} and variables at the sample person level $\beta_1 \text{sp\_V1}_{ijk}$ and $\beta_2 \text{sp\_V2}_{ijk}$, PSU level $\beta_3 \text{psu\_V1}_{ij}$ and $\beta_4 \text{psu\_V2}_{ij}$ as well as interviewer level $\beta_6 \text{i\_V1}_{k}$ and $\beta_7 \text{i\_V2}_{k}$ and the residuals for the interviewers v_{00k}, and PSUs u_{0jk} for *with* the assumption that the variance across interviewers $\sigma^2_{v_{00k}}$ and the variance across PSUs nested within one interviewer $\sigma^2_{u_{0jk}}$ are normally distributed.

Random slopes

By including random slopes in the models, I test whether the propensity of a sample person to be successfully contacted differs across interviewers for each group of an explanatory variable, e.g. male and female sample persons. Equation (4.6) shows an example for estimating a random slope including one sample person variable:

$$\begin{aligned} &Y_{ijk} = \beta_{000} + \beta_x x_k + u_{1jk}\,\text{sp\_V1}_{ijk} + v_{00k} + u_{0jk} \\ &with \quad v_{00k} \sim N(0, \sigma^2_{v_{00k}}) \\ &\qquad u_{0jk} \sim N(0, \sigma^2_{u_{0jk}}) \end{aligned} \tag{4.6}$$

where the natural logarithm of the propensity *P* of the outcome Y_{ijk} divided by 1 minus the propensity *P* of the outcome Y_{ijk} for sample person *i* nested in PSU *j* nested in interviewer *k* is estimated as the intercept of the regression β_{000} and sample composition control variables at the PSU and sample person level and all interviewer characteristics from the previous models $\beta_x x_k$, the random slope $u_{1jk}\,\text{sp\_V1}_{ijk}$ and the residuals v_{00k} for the interviewer and PSUs u_{0jk} with the assumption that the variance across interviewers $\sigma^2_{v_{00k}}$ and the variance across PSUs nested within one interviewer $\sigma^2_{u_{0jk}}$ are normally distributed.

4.4 Results

To identify interviewer effects on a sample persons' propensity to be successfully contacted and participate in PIAAC, I started with estimating null models of three-level logistic regressions to identify the amount of variance located at the interviewer level during the recruitment phase (see Appendix E). Around 20% of the variance of a sample persons' propensity to be successfully contacted can be explained by interviewer clustering. And, for a sample persons' propensity to participate in PIAAC, around 2% of the variance can be explained due to interviewer clustering. The very small interviewer effect on a sample persons' willingness to participate in PIAAC does

not warrant further analyses to explain this low interviewer effect by means of interviewer characteristics. However, a relevant part of the variance of a sample persons' propensity to be successfully contacted is located at the interviewer level and is worth explaining using interviewer characteristics while controlling for sample composition effects.

Similar as done by Jäckle et al. (2013), I continue my analyses of interviewer effects on a sample persons' propensity to be successfully contacted by including variables in my models to account for an *a non-random assignment* of sample persons to interviewers in PIAAC Germany, in other words, to account for sample composition effects. This is relevant, because it is not possible to fully disentangle interviewer effects and area effects due to the data structure in PIAAC: Sample persons (level 1) are clustered within PSUs (level 2) and PSUs are clustered within interviewers (level 3).

Results of the three-level logistic regression on a sample persons' propensity to be successfully contacted that include sample composition control variables are shown in the coefficient plot in Figure 4.4 (also see the regression table in Appendix E, model 2). In the coefficient plot, the markers are coefficients and the horizontal lines are confidence intervals. The results of the sample person characteristics are as follows: I find that younger persons have a significantly lower propensity to be successfully contacted (-.15 logODDS), men have a significantly lower propensity to be successfully contacted (-.38 logODDS), and Germans have a significantly higher propensity to be successfully contacted (.41 logODDS). In addition, my findings show that persons living in larger cities have a significantly lower propensity to be successfully contacted (-1.65; -1.55; -2.49 logODDS).

By including sample composition control variables in the model, almost all variance of the propensity of a sample person to be successfully contacted because sample persons are clustered within PSUs and 44% of the variance of a sample person to be successfully contacted because sample persons are clustered within PSUs within one interviewer is significantly explained (likelihood-ratio test; see Rabe-Hesketh & Skrondal, 2012). After I include the sample

composition control variables, the variance of the propensity of a sample person to be successfully contacted at the PSU level is 2% and the variance of the propensity of a sample person to be successfully contacted that can be explained because sample persons are clustered within PSUs within one interviewer is 24%. In addition, the interviewer variance *is not reduced* by including sample composition control variables. Instead, the amount of variance at the interviewer level is slightly higher when sample composition control variables are included in the model (22%; see Appendix E). These results indicate an effect on the propensity of a sample person to be successfully contacted of the sample composition which is not related to the interviewers.

The results for the sample composition control variables described above will not be interpreted anymore for the models including interviewer characteristics, as they are included as control variables in all further models.

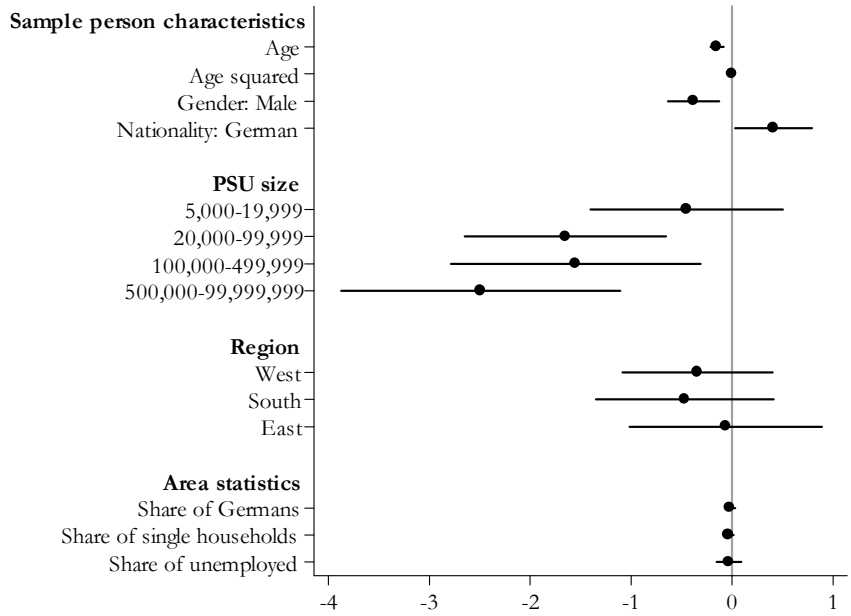


Figure 4.4 Coefficient plot for the odds rations of successful contact, model 2, three-level logistic regression, PIAAC 2012

Notes. Parameter estimates with 95% confidence intervals from three-level logistic regression for the dependent variable of successful contact. Significant = $p < 0.05$. Reference PSU size: 0-4,999. Reference region: North. Number of interviewers = 115. Number of PSUs = 251. Number of sample persons = 7,902. The full model can be found in Appendix E, model 2. PSU = primary sample unit.

Next, I investigate, how interviewer effects on a sample persons' propensity to be successfully contacted can be explained by means of interviewer characteristics – which is my major research interest. To answer this question I include one interviewer characteristics per model. Again, in all models, I control for possible sample composition effects by including sample composition variables. Overall, the results show that almost none of the interviewer characteristics significantly explain the differences in a sample persons' propensity to be successfully contacted between interviewers. However, three of the interviewer characteristics are found to have a significant effect on a sample persons' propensity to be successfully contacted (see Figure 4.5, Figure 4.6 and Appendix E).

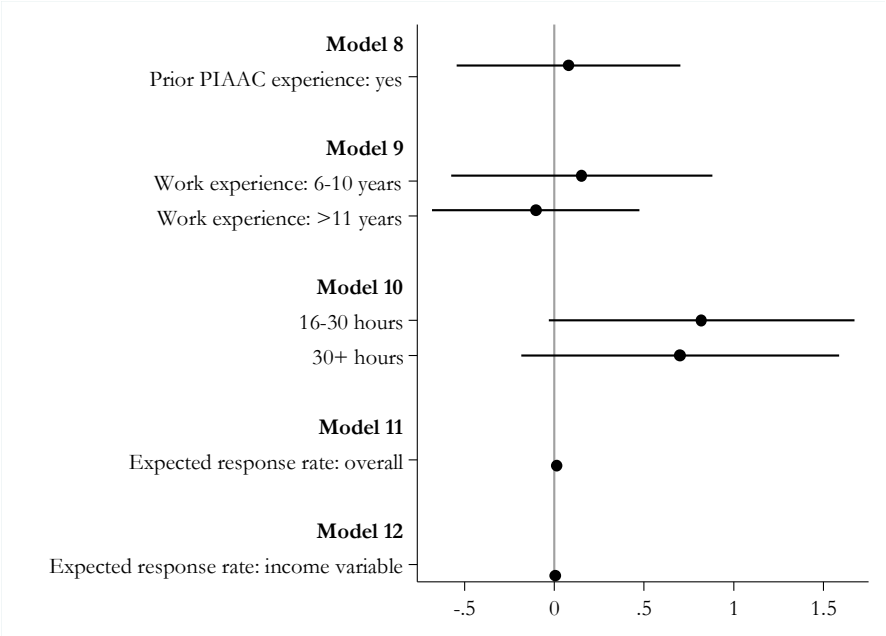


Figure 4.5 Coefficient plot for the odds rations of successful contact, model 8 – model 12, three-level logistic regression, PIAAC 2012

Notes. Parameter estimates with 95% confidence intervals from three-level logistic regression for the dependent variable of successful contact. Significant = $p < 0.05$. All models control for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). Reference model 9: <=5 years. Reference model 10: <=15 hours. Number of interviewers = 115. Number of PSUs = 251. Number of sample persons = 7,902. The full models can be found in Appendix E, models 8-12.

Sample persons have a higher propensity to be successfully contacted (see Model 10) when interviewers with a medium amount of working hours per week (16-30 hours) attempt contact with them compared to sample persons with whom interviewers attempt contact who work a low amount of working hours per week (less than 15 hours). However, this effect is only significant at

the 10%-level and I find no significant effect for interviewers with the highest amount of working hours per week (30+ hours). Thus there is only a weak indication that interviewers should be allocated a moderate amount of cases to ensure they are able to have a balanced time management to attempt contact with sample persons to achieve high contact rates. In contrast, interviewers who work less than 15 hours a week might have other duties (e.g. another job[5]) and have not enough time to attempt contact with sample persons often enough to successfully make contact. In addition, no interviewer variance can be explained significantly by interviewers' amount of working hours per week (likelihood-ratio test; see Rabe-Hesketh & Skrondal, 2012).

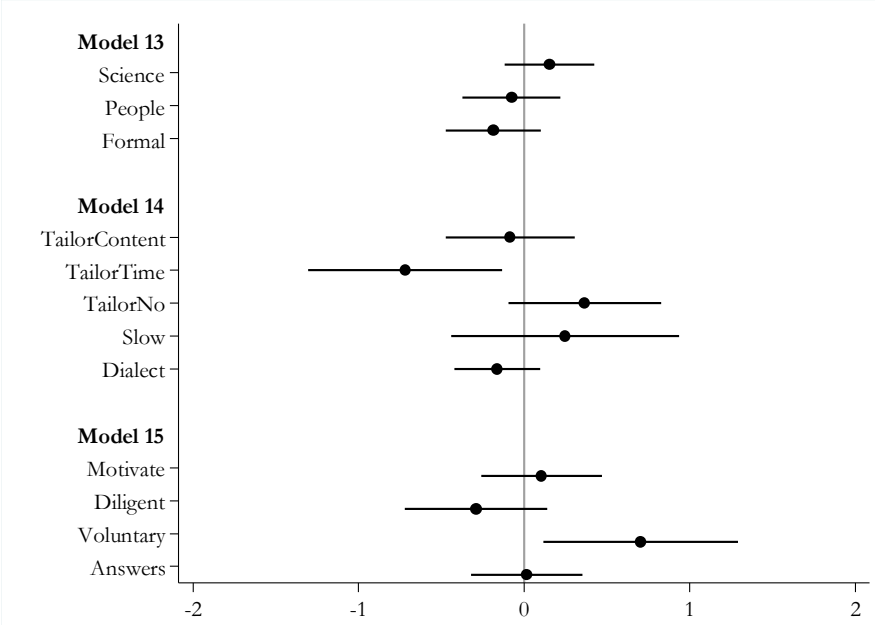


Figure 4.6 Coefficient plot for the odds rations of successful contact, model 13 – model 15, three-level logistic regression, PIAAC 2012

Notes. Parameter estimates with 95% confidence intervals from three-level logistic regression for the dependent variable of successful contact. Significant = $p < 0.05$. All models control for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). Number of interviewers = 115. Number of PSUs = 251. Number of sample persons = 7,902. The full models can be found in Appendix E, models 13-15.

Sample persons have a significantly lower propensity to be successfully contacted when interviewers attempt contact with them who say they tailor question text to shorten the interview time to meet sample persons requirements (see Model 14). Thus, I assume that interviewers who do not follow the standardized survey interview techniques by tailoring the question text to meet

[5] The effect remains significant when controlling for all other interviewer characteristics including occupational status.

sample persons' requirements also tend to tailor their contact strategies to meet sample persons requirements. However, no interviewer variance can be explained significantly by including this indicator in the models (likelihood-ratio test; see Rabe-Hesketh & Skrondal, 2012).

The last interviewer characteristic which is found to have a significant effect on a sample persons' propensity to be successfully contacted refers to one indicator on interviewers' attitude regarding statements on how to achieve response. Explicitly, sample persons have a significant higher propensity to be successfully contacted (see Model 15) when interviewers attempt contact with them who report a higher importance with respect to the statement that a sample persons' voluntariness to cooperate in a survey should be respected. However, the interviewer variance is reduced by 2 percentage points and not found to be significantly reduced when this interviewer characteristics is included in the model (likelihood-ratio test; see Rabe-Hesketh & Skrondal, 2012). de Leeuw et al. (1998) or Lehtonen (1996) for example found that the effect of interviewers' rating of the importance to respect a sample persons' voluntariness to participate in a survey is negatively related to cooperation. Thus the unexpected result of a positive relationship of the statement regarding a sample persons' voluntariness to participate in a survey with the propensity to be successfully contacted suggests that contact and cooperation strategies do not require the same competencies of interviewers.

In addition, I tested random slope effects of sample composition variables for the following variables: Sample person age, gender, nationality, and PSU size. My findings indicate that the propensity of a sample person to be successfully contacted does not differ significantly across interviewers by the size of the PSU as well as age and nationality of the sample person. However, the propensity of a sample person to be successfully contacted is significantly different across interviewers for female and male sample persons. However, results of performing a likelihood-ratio test comparing the random-intercept and random-slope model show that the random slope model is rejected (Rabe-Hesketh & Skrondal, 2012). Thus, analyses including random slopes are not presented for further investigation. In addition, this implies that testing cross-level

interactions between interviewer and sample person characteristics does not warrant further analyses.

It is especially important to have a closer look at the amount of interviewer variance for the models which include interviewers' characteristics compared to the amount of interviewer variance in the model containing only sample composition control variables. My results show that, for all models, the variance introduced through interviewer clustering remains around 20%. The interviewer variance explained by including all interviewer characteristics at the same time in one model is 10.3% (see Appendix E, model 19). However, the reduction of the interviewer variance is not found to be significant (likelihood-ratio test; see Rabe-Hesketh & Skrondal, 2012). Therefore, interviewer characteristics included in the models seem to not efficiently explain the variance introduced through interviewer clustering with regard to sample persons propensity to be successfully contacted. When all interviewer characteristics are included in one model at the same time, the identified effects remain significant, although the significance levels are reduced.

4.5 Discussion and Conclusion

The main goal of this chapter is to understand why some interviewers are more successful than others with respect to two aspects of survey unit nonresponse: Successfully making contact with sample persons and gaining sample persons' cooperation to participate in the survey. This question is studied using data from the German implementation of PIAAC, supplemented by auxiliary data on the sample composition as well as interviewer characteristics.

The findings show three main results: First, I find a very small interviewer effect on the willingness of sample persons to cooperate with 2% variance located at the interviewer level. This can be interpreted as an ignorable interviewer effect on gaining sample persons' cooperation and is not investigated any further in the present chapter. I assume this very low interviewer effect on cooperation, to be a positive outcome of the very intense interviewer trainings conducted with PIAAC interviewers prior to the field work. During the trainings, the high standards for gaining

sample persons' cooperation were especially emphasized and might explain why interviewers were more homogenous in gaining sample persons' cooperation. However, achieving high contact rates was also an important aspect during the PIAAC interviewer trainings. Thus the interviewer effect of around 20% variance located at the interviewer level for a sample persons' propensity to be successfully contacted is surprising. An alternative explanation of the different amount of interviewer variance identified for the two aspects of survey unit nonresponse can be found by looking at the distribution of contact and cooperation rates by interviewers (see Figure 4.1 and Figure 4.2). It becomes obvious that the variation in contact rates between interviewers is much lower than the variation in cooperation rates between the interviewers. In other words: Interviewers seem more diverse with respect to gaining sample persons' cooperation, whereas interviewers seem to be more similar in making successful contact with sample persons. However, the differences between interviewers with different contact rates seem to be much higher than the differences between interviewers with different cooperation rates. In addition, only three interviewers who have a contact rate below 85%. Thus, I assume that the high interviewer effects on a sample persons' propensity to be successfully contacted to be the effect of some outliers.

Second, I investigate whether there are sample composition effects that explain part of the interviewer effect on a sample persons' propensity to be successfully contacted. My results show that the variance explained by PSU clustering, as well as the variance that can be explained because sample persons are clustered within PSUs within an interviewer, can be significantly reduced by including sample composition control variables in the models. In addition, I find that the interviewer variance is slightly higher when sample composition control variables are taken into account. This indicates an effect of the sample composition on a sample persons' propensity to be successfully contacted which is not related to interviewers. However, as in PIAAC, the sample persons were not randomly assigned to interviewers it is not possible to fully disentangle sample composition and interviewer effects. Instead, I handle this limitation of the data by

following the best practice to account for the nested data structure by controlling for sample composition effects by means of including as many available sample composition variables as possible in my models.

Third, I tried to explain some of the identifyed interviewer effect on successfully making contact with sample persons by means of interviewer characteristics meassured with an extensive interviewer questionnaire. Here, I find significant effects for three interviewer characteristics on a sample persons' propensity to be successfully contacted: A medium amount of working hours per week, tailoring question text to shorten the interview to meet sample persons' requirements, and of voluntariness of sample persons' decision to cooperate in a survey. Although my analyses are based on a very rich source of interviewer characteristics, none of the interviewer characteristics can significantly explain interviewer variance of a sample persons' propensity to be successfully contacted. These results allow two conclusions: First, the data needed to explain interviewer effects on a sample persons' propensity to be successfully contacted is not available, which is in accordance to conclusions that are drawn in many other studies (e.g. see West & Blom, 2017). And secondly, I already explained that I assume outliers with very low contact rates to be the major reason for the high interviewer effect on a sample persons' propensity to be successfully contacted. Thus, as it likely that these outliers do not differ much with respect to the characteristics collected via the interviewer survey and that the identified interviewer effect cannot be explained by means of the available interviewer characteristics.

Based on my findings of the present chapter, I highly recommend reporting interviewer effects during the recruitment process and, if interviewer characteristics are available, controlling for any available interviewer characteristics. Furthermore, I suggest adapting interviewer trainings according to any identified interviewer effect. More specifically, in PIAAC, the focus of the interviewer trainings could be intensified with regard to successfully making contact with sample persons during interviewer trainings for further cycles of PIAAC.

Furthermore, further research could correct for the difficulty of cases in the analyses, like for example done by Brunton-Smith et al. (2012). They find that it is worthwhile implementing an adjustment for difficulty of cases by means of variables that have been found to be predictive of survey unit nonresponse when interviewer effects are analyzed.

In addition, it is often criticized in the literature (e.g. Wang, Kott, & Moore, 2013), that many studies that assess interviewer effects on survey unit nonresponse do not take measurement error in account. Interviewers good at contacting or persuading sample persons are not necessarily good at interviewing persons using standardized interviewing techniques. In other words, the variance introduced by interviewer clustering during the recruitment process could be one reason for variance reported in estimates of substantive survey variables (Stokes & Yeh, 1988). To contribute to the literature, I analyze interviewer effects on estimates of substantive survey variables in my dissertataion. In addition, I conduct combined analyses of interviewer effects on two aspects of the TSE: unit nonresponse error and meassurement error.

5 Study 2: Interviewer Behavior and Interviewer Characteristics in PIAAC Germany[6]

5.1 Introduction

Face-to-face surveys rely on interviewers for data collection. However, behavior regarding standardized interviewing techniques may differ across interviewers. As a result, interviewers can influence – intentionally or unintentionally – various aspects of the data collection process. Concerns about interviewer effects in interviewer-mediated surveys have accompanied generations of survey researchers. According to Groves et al. (2009a), Loosveldt (2008), Schaeffer et al. (2010), and Blom and Korbmacher (2013), interviewers have many different roles in administering a survey: They contact sample persons and persuade them to participate, they clarify the goal of the survey and explain to respondents what is expected of them, as well as ask questions and record answers. Thus, the behavior of interviewers affects nearly all aspects of survey errors, including sampling (Eckman, 2013; Eckman & Kreuter, 2011; Tourangeau et al., 2012), nonresponse (e.g., Blom et al., 2011; Durrant et al., 2013; Jäckle et al., 2013), measurement (Durrant et al., 2010; Rice, 1929), and coding or editing of survey responses (e.g., Campanelli et al., 1997). The focus of the present paper is on the measurement perspective of interviewer behavior: interviewers´ behavior with regard to deviations from standardized interviewing techniques during interviews.[7]

In terms of the total survey error framework, as many error sources as possible should be taken into account when designing a survey (for a survey see Groves & Lyberg, 2010). When it comes to errors during face-to-face interviews, standardized interviewing techniques are commonly used as a strategy to reduce errors introduced by interviewers (e.g., Fowler & Mangione, 1990; Mangione, Fowler, & Louis, 1992). In a standardized interview, interviewers are

[6] This chapter was published as Ackermann-Piek, D., & Massing, N. (2014). Interviewer behavior and interviewer characteristics in PIAAC Germany. Methods, Data, Analyses, 8(2), 199-222. doi: 10.12758/mda.2014.008.
Chapter 5, uses a slightly different coding of the variables and indicators available from the interviewer questionnaire, since this chapter has already been published in a peer-reviewed journal.

[7] For more information regarding nonresponse in PIAAC Germany see Helmschrott and Martin (in this volume).

expected a) to read aloud questions, as well as instructions, as they are scripted, b) provide adequate nondirective probing, if necessary, and c) be unbiased towards respondents and record answers accurately (Fowler & Mangione, 1990, p. 14). All steps should be conducted in exactly the same way by each interviewer and therefore no differences between them should occur. Accordingly, all respondents are provided with identical stimuli and the "general assumption is that when all interviewers do their job in a standardized way and adhere to the interview rules, and when they interview a comparable group of respondents, they will get comparable answers." (Loosveldt, 2008, p. 216).

However, several studies have shown that interviewers deviate from standardized techniques. Hyman and Cobb (1954) were among the first to present results of errors introduced by interviewers who did not follow standardized interviewing techniques. Several other studies followed and revealed, for example, effects introduced by autonomously reworded text (e.g. Billiet & Loosveldt, 1988; Brenner, 1982; Haan, Ongena, & Huiskes, 2013; Ongena, 2005). Maynard and Schaeffer (2002) summarized the debate on standardization and concluded that understanding why interviewers deviate from the expected behavior helps to improve data quality.

Two approaches are commonly used to explain why interviewers deviate from standardized interviewing techniques. The first approach focuses on the survey instrument and the second on the interaction in the question-answer process. With respect to the survey instrument, many guidelines have been written on how survey questions should be scripted (e.g. Porst, 2008). Firstly, formulating survey questions of good quality reduces the bias introduced by interviewers, because they do not feel the need to deviate from the question text (Schaeffer, 1991; Schaeffer et al., 2010; Schaeffer & Maynard, 1996). Secondly, Schober and Conrad (2002) concluded that, due to the nature of communication, interviewers collaborate with respondents when trying to improve question understanding, which might affect responses. Additionally, interviewers might not want to appear ignorant or impolite and therefore tailor the question text (Ongena &

Dijkstra, 2006). Further studies suggest that conversationally structured interviews reduce interviewers´ burden and therefore minimize the chance of mistakes, because there are no rules for standardization (e.g. Cannell, Miller, & Oksenberg, 1981; Houtkoop-Steenstra, 2000; Schober & Conrad, 1997). Although these authors state that a flexible interviewing technique has many advantages – especially for interviewers – they admit that it is very time consuming and more challenging when controlling interviewers´ work.

However, these two approaches do not fully explore the reasons for interviewers' deviations from standardized techniques. The literature suggests a third approach: using interviewer characteristics, such as attitudes or behavior, as predictors for nonresponse and measurement error (Blom & Korbmacher, 2013; Durrant et al., 2010). However, research into the effects of interviewers' background characteristics, such as gender, age or education, has yielded inconsistent findings (for an overview see Schaeffer et al., 2010). Groves (2004) concluded that interviewers´ characteristics are mostly associated with measured constructs when both are related (e.g., questions on respondents´ weight might be affected by interviewers´ gender). For example, interviewers´ experience is often used to explain differences in the success of reaching contact or gaining cooperation.[8] Gfroerer, Eyerman, and Chromy (2002) related interviewers´ experience to standardized interviewing techniques and found that less experienced interviewers tend to be more accurate in reading questions. Furthermore, Groves et al. (2009a) and Groves and Lyberg (2010) reported that interviewers with more experience introduce greater measurement error to the data. However, other studies did not find an effect of experience and conclude that any effects might be overcome with training (e.g. Collins, 1980).

Nevertheless, detailed data on interviewers' actual behavior during the interview and interviewers´ characteristics are often not available. Because these data is available for the Programme for the International Assessment of Adult Competencies (PIAAC) Germany, we used the third approach. The combination of detailed background information about interviewer

[8] This relationship is usually linear (e.g. Jäckle et al., 2013) or, rarely, U-shaped (Singer et al., 1983).

characteristics with actual interview behavior is special and enables us to fill a gap in the literature and explain deviations of interviewers´ behavior from standardized interviewing techniques. We first describe the behavior of the interviewers in the standardized structured background questionnaire of PIAAC Germany. Subsequently, we present findings from analyses of the association between interviewer behavior during the PIAAC interview and interviewer characteristics.

5.2 Data Description

In comparison to many other studies that use auxiliary data to evaluate interviewers´ behavior, we could rely on factual data from the German PIAAC survey. We used data about interviewers that were either on the interviewer level or on the respondent level. Data on interviewers´ background characteristics came from an interviewer questionnaire that was implemented in order to collect more data on interviewers, their attitudes, and reported behavior. Data on interviewers´ actual behavior regarding standardized interviewing techniques were derived from audio recordings of interviews collected during the fieldwork in PIAAC Germany. In the following section, we first briefly explain the interviewers´ role in PIAAC Germany[9] and then describe both data sources in more detail.

5.2.1 PIAAC Germany and the role of interviewers

PIAAC is an international survey, initiated by the OECD (2014) and implemented by an international Consortium. Its aim is to investigate how adults´ competencies are distributed across and within countries. All participating countries collected data via face-to-face interviews with randomly sampled persons. In Germany – like in almost all other participating countries – the data collection took about eight months, between August 2011 and March 2012.[10] In total, 129 interviewers from the German survey organization TNS Infratest worked for PIAAC in Germany. The cases were organized in sample points based on a random sample of the adult

[9] The description of PIAAC is based on our own experience during the implementation of PIAAC in Germany, as well as on the international technical report (OECD, 2016) and the German PIAAC technical report (Zabal et al., 2014).

[10] This included two main fieldwork phases as well as several re-issue phases.

population in Germany (16-65 years of age). Most interviewers worked in two or three sample points with 32 addresses per point. However, due to organizational arrangements, a few interviewers worked in only one or in up to five sample points. In total, the target size of approximately 5,000 respondents was exceeded, with a final number of 5,465 completed interviews.[11]

In PIAAC, the role of the interviewers differed somewhat from their normal tasks. The design of PIAAC included not only a computer-based background questionnaire, which interviewers are accustomed to administer, but also a computer-based assessment of every-day skills in the domains *literacy*, *numeracy* and *problem solving in technology-rich environments*. The background questionnaire was administered as a computer-assisted personal interview and contained questions about the respondent, such as education or the use of skills at work and in every-day life. The assessment was in a self-completion format administered under the supervision of the interviewer. Although we did not use the data collected in the skills assessment for the analysis in this paper, it is important to note that the interviewers had to adapt their behavior for the assessment, because they had to learn to be more passive in their role as test administrators.

To ensure that the PIAAC data was of high quality, specific and comprehensive technical standards and guidelines were defined by the international Consortium (OECD, 2014) and each participating country had to comply with these standards when carrying out PIAAC. The implementation of the standards was monitored very closely by the Consortium and every single deviation from the standards had to be approved. One important aspect of the international requirements referred to quality control of the fieldwork: interviewers' work, as well as the data quality, had to be closely monitored.[12] The analyses in this paper that deal with deviations from standardized interviewing techniques were based on the information retrieved from audio recordings of interviews from the PIAAC background questionnaire that was collected and

[11] For a definition of a *completed case* in PIAAC see OECD (2014).
[12] All standards and guidelines related to interviewers are described in detail in Massing et al. (2013).

reviewed in this context. The international requirements for quality control stipulated that each interviewer had to produce two audio recordings (for more details see below).

Another important aspect in the PIAAC standards and guidelines was that interviewers received intensive in-person trainings, to provide them with adequate information and practice for carrying out their various tasks. The training included a special focus on standardization for the data collection in the background questionnaire. Conducting such extensive interviewer trainings is relatively uncommon in Germany. In other countries, however, this is best practice and several studies have demonstrated a positive effect of interviewer trainings on response rates and on the overall data quality (e.g. Billiet & Loosveldt, 1988; M. P. Couper & Groves, 1992; Fowler & Mangione, 1990; Japec, 2008). Furthermore, German PIAAC interviewers were carefully selected.[13]

In addition to their training, interviewers were provided with substantial information material. For instance, they received an extensive manual that included detailed descriptions of each relevant aspect of PIAAC in Germany, as well as a small interviewer booklet. Providing interviewers with such extensive material is also uncommon in German surveys.

5.2.2 Interviewer questionnaire

To date, interviewer behavior, or even interviewer effects, has often only been described but not explained, because data to explain those effects are lacking (Blom & Korbmacher, 2013; Brunton-Smith et al., 2012). In Germany, detailed data on interviewer characteristics are normally not provided by survey agencies. To overcome this gap, additional data on the PIAAC interviewers were collected by the authors, using a questionnaire that was adapted from the questionnaire implemented in the Survey of Health, Ageing and Retirement in Europe (SHARE) 2011 (Blom & Korbmacher, 2013). Interviewers' participation was voluntary and the interviewers did not receive any kind of incentive. Data from the interviewer survey were not intended to be used for quality control measures during PIAAC but rather to gain more information about the

[13] The selection criteria are described in detail in Zabal et al. (2014).

interviewers, in order to analyze differences in interviewers´ behavior and success, related to their characteristics. It contained questions about the interviewers' background, their attitudes, and their expectations, related to their fieldwork in PIAAC.[14] The questionnaire was sent to 128 interviewers and 115 interviewers completed and returned the questionnaire, resulting in a response rate of almost 90%. However, 15 questionnaires were received without an interviewer ID (see Table 5.1). These cases could not be matched with interviewer behavior retrieved from the audio data. Therefore, they were excluded for joint analysis of interviewer characteristics and interviewer behavior. Their exclusion did not alter the results.

Table 5.1 Overview of the interviewer questionnaire

	n	**Percent**
Interviewer received questionnaire	128	100.00
Interviewer returned questionnaire	115	89.8
Questionnaire contained interviewer ID	100	78.1

Notes. One interviewer was excluded after a short time. Therefore, the questionnaire was sent to 128 interviewers.

A summary of the interviewers´ background characteristics, collected through the interviewer survey, is provided in Table 5.2. The results for gender and age were equivalent to the information provided by the survey agency TNS Infratest in their technical report (Zabal et al., 2014, p. 54). TNS Infratest provided additional information on how long interviewers had been working for their survey institute: 71% of the interviewers had worked for TNS Infratest for ten years or less. However, our results show that over 45% stated that they had worked as interviewers for more than ten years. Another interesting issue is related to the experience of PIAAC interviewers: compared to interviewers from other German surveys, PIAAC interviewers were very experienced (Blom, Ackermann, Korbmacher, Krieger, & Massing, 2013). This is not surprising, because one criterion for selection as a PIAAC interviewer required candidates to be a senior interviewer.

[14] The source questionnaire is presented in Blom and Korbmacher (2013).

Table 5.2 Characteristics of the German PIAAC interviewers

		n	Percent
Gender	Male	62	53.91
	Female	53	46.09
	Total	115	100.00
Age	<= 45 years	10	8.70
	45 – 55 years	21	18.26
	56 – 65 years	51	44.35
	>= 66 years	33	28.70
	Total	115	100.00
Work experience	< 2 years	10	8.77
	2 – 5 years	31	27.19
	6 – 10 years	21	18.42
	11 – 15 years	10	8.77
	> 15 years	42	36.84
	Total	114	100.00
Education	Lower-level or medium-level school and no vocational or university qualification	1	0.93
	Medium-level school qualification and vocational education	36	32.73
	Advanced technical college entrance qualification or university entrance qualification	42	38.18
	Tertiary education	31	28.18
	Total	110	100.00
Working hours per week	<= 10 hours	6	5.66
	11 – 20 hours	31	29.25
	21 – 30 hours	36	33.96
	31 – 40 hours	18	16.98
	>= 31 hours	15	14.15
	Total	106	100.00

Notes. Data from the PIAAC interviewer survey. 115 interviewers included in analysis. Number of cases varies because of item nonresponse.

5.2.3 Audio recordings and coding scheme

As mentioned above, the PIAAC standards stated that each country had to evaluate at least two audio recordings, per interviewer, of interviews made during administration of the background questionnaire (OECD, 2014). Analyzing recordings is considered to be a good way of monitoring interviewers' behavior and interviewing techniques, without affecting respondents' behavior (Fowler & Mangione, 1990; Sykes & Collins, 1992). In addition, such recordings provide insights into the complex interaction process between interviewers and respondents (Ongena, 2005). The audio recordings were taken early in the field period. The interview was

recorded via an external digital voice recorder and the interviewer had to manually start and stop the recording. Table 5.3 shows an overview of the expected as well as the recordings actually delivered by the interviewers. In total, 258 recordings were expected. Recordings were not available for some interviewers, whilst others delivered more than two recordings. In total, 245 recordings were received, coded, and reviewed during quality control of the fieldwork in PIAAC Germany.

To use the information from the audio recordings for quality control, information first had to be coded. In the literature, several coding schemes are available, indicating that the choice of coding scheme depends on the purpose of the analysis (Ongena & Dijkstra, 2006).

The main reason for evaluating interviewer behavior using audio recordings in PIAAC was quality control. The aim was to obtain information about the interviewers' interviewing techniques and their actual behavior during the interview as early as possible during the data collection in order to intervene, if necessary. Because coding and reviewing audio recordings is very time consuming[15] and information was needed as early as possible, we developed a simple coding scheme that focused on crucial deviant interviewer behavior in the background questionnaire.[16] A major problem was defined as a deviation from the standardized script that potentially affects the level of accuracy of the response (Ongena, 2005).

To avoid coder effects, coding was conducted by six different coders. It was ensured that two persons coded the recordings of one interviewer. Any inconsistencies or difficulties in the codes were resolved by two lead coders. After a review of the coding, a summary of the behavior of each individual interviewer was written by the lead coders and feedback was provided to the survey agency. All codes were derived directly from the audio recordings.

[15] Coding the background questionnaire took about one hour per recording and was conducted directly from the recordings, using the software *Audacity* (Mazzoni & Dannenberg, 2012).
[16] The PIAAC technical standards and guidelines only required this part of the interview to be reviewed via recordings.

Table 5.3 Overview of the audio recorded interviews

	n	Percent
Interviewer	129	100.0
Interviewer with no recordings	8	6.6
Audio tapes to be recorded	258	100.0
Received audio taped interviews	245	95.0
Interviewer with 1 recording	1	0.8
Interviewer with 2 recordings	116	95.9
Interviewer with 3 recordings	4	3.3

Note. Reference: Zabal et al. (2014).

For the present analysis, we reorganized the original coding scheme used for quality control in PIAAC, based on the coding scheme of Beullens, Loosveldt, and Dries (2013). Each single code represents one aspect of standardization. The resulting seven codes were grouped into three categories: administration, completeness, and probing (see Figure 5.1).

The first category contained administrative information that interviewers were asked to record at the beginning of the interview. The first code *admin I* consisted of a combination of the following information: the date of the interview, the interviewer ID and the respondent ID. Only if the interviewer ID or the respondent ID was recorded incorrectly (missing or incomplete) was this coded as incorrect interviewer behavior. *Admin II* covered whether interviewers announced the recoding to the respondent and whether they explicitly asked for permission to record the interview. This was especially crucial because data protection regulations are strict in Germany. Only if the announcement of the recording was completely absent on the recording was this coded as incorrect interviewer behavior. However, because a digital voice recorder, and not the laptop, was used to record the interview, it was obvious for all recordings that all respondents were aware that the interview was being recorded. This was further confirmed by the audio recordings, which contained no indication of any secret recording of interviews. Nevertheless, because this was a formal requirement, this code provided information on how accurately interviewers worked.

For the second category, *completeness*, the two codes referred to question text.[17] We will explain these codes by using the example of a question wording, provided in Figure 5.2 , to illustrate deviations from standardized interviewing techniques. We coded each single incidence of an incorrectly skipped question as incorrect interviewer behavior.

	Administration	Completeness	Probing
Admin I: collected date of interview, interviewer ID, respondent ID	x		
Admin II: collected permission to record interview from respondent	x		
Question is read out (not incorrect skipped)		x	
Question is completely read out		x	
Probing (if applicable)			x
Probing overall correct			x
3-point scale for probing quality			x

Notes. ID = Identification number. Admin = Administration.

Figure 5.1 Coding scheme for audio recordings of the background questionnaire of PIAAC in Germany

With respect to the question wording provided in Figure 5.2 , we found that interviewers often deviated from the script, using information from the previous part of the interview. For example, in one interview, the interviewer assumed that the respondent was a student instead of part-time employed, because both talked about forthcoming holidays. Because the question was not asked, the interviewer collected incorrect information. As a consequence, various filters of the following questionnaire did not fit the respondent's situation and data was incorrect. Although incorrectly skipped questions do not necessarily result in incorrect data, this example shows that each piece of information obtained from the previous conversation has to be verified by asking each single question (Ongena, 2005). Luckily, in our example, the respondent realized

[17] During quality control, two additional codes were used, referring to answer categories and showcards. However, coding could not be derived from the audio recordings for all cases and we thus excluded these codes from our analysis.

the error introduced by the interviewer and asked to go back, to change the information that applied to her situation.

Question

Please look at this card and tell me which ONE of the statements best describes your current situation. If more than one statement applies to you, please indicate the statement that best describes how you see yourself.

Instruction

1. Hand show card 9.
2. Mark only one answer.

Answer Categories

01 Full-time employed (self-employed, employee)
02 Part-time employed (self-employed, employee)
03 Unemployed
04 Pupil, student
05 Apprentice, internship
06 In retirement or early retirement
07 Permanently disabled
08 In compulsory military or community service
09 Fulfilling domestic tasks or looking after children/family
10 Other
DK
RF

Notes. DK = don´t know. RF = refused. DK and RF were not printed on showcards in general.

Figure 5.2 Example of a question from the PIAAC background questionnaire

With respect to the second aspect of completeness, we assume that rewording or shortening a question has either no, a minor, or a major impact on the respondents' answers, and use the example provided in Figure 5.2 to explain the differences. For the wording presented in Figure 5.2 , the interviewer might simply leave out the first word "Please". We assume that this has no effect on question understanding. However, this rewording could also have a minor effect, if respondents think that the question is not worded very politely or that the interviewer is impolite. We assume that minor rewordings have no major effect on the accuracy of responses. On the other hand, we assume that complete rewordings of the original question text (e.g., changing the question wording presented in Figure 5.2 to: "Are you employed?") will have major effects on the accuracy of responses, if further information is not provided by the interviewer about how respondents should answer the question and, thus, respondents do not have the opportunity to assign themselves to the correct answer category. In comparison, a minor effect of

this completely reworded question could be that respondents ask for clarification and interviewers probe to provide respondents with the missing information. As mentioned above, we decided to focus on major problems and did not code minor rewordings as incorrect interviewer behavior during quality control. We only coded major deviations from the original question text which, we assumed, would have major effects on the responses, as incorrect interviewer behavior.

Finally, three codes referred to probing (Figure 5.1), an interviewing technique in which additional information is provided on request. This is usually triggered by respondents, when, for example, they ask for clarification of the question or give an inaccurate answer (e.g., one that does not fit the answer scheme). Each time interviewers had to probe, the quality of the probing was coded. The first code included information on whether probing was necessary or not. We subsequently constructed a dichotomy code that included information about whether probing was correct or not. Because there is a wide range of probing quality, we decided to additionally build a three-point scale to differentiate between a) excellent probing, b) probing that was not good, but for which it was assumed that it would not have a major negative effect on the respondent´s answer and, c) poor probing. The scale was constructed by combining the number of good and poor probes, based on the overall distribution: More than three correct probes were considered to be excellent probing on the scale; if an interviewer conducted only bad probing, without any good probing, we considered this to be poor probing, and everything in between was assigned to the middle category. A good probe is nondirective and neutral, which means that it does not influence the content of the answer. In contrast, a poor probe influences the answer of the respondent (Fowler & Mangione, 1990). Due to limited details in the original coding schema, this scale could be applied to approximately only half of the recordings.**5.3 Results**In this section, we present results of the descriptive analysis of the interviewer behavior retrieved from the audio recordings. We start by describing how many interviews we identified in which interviewers collected administrative information incorrectly and then proceed to provide

information on interviewers' behavior using standardized interviewing techniques such as reading questions without incorrect skipping or rewording. Finally, we provide information on interviewers' probing behavior. In the second part of this section, we show whether interviewers´ behavior in the interviews is associated with interviewers´ background characteristics. For this purpose, we crossed the information from the audio recordings with interviewers´ characteristics from the PIAAC interviewer survey and estimated several regressions. All results in the following section are based on those cases for which the interviewer ID was available from the interviewer questionnaire. Nevertheless, results including all cases do not differ substantively.

Table 5.4 Interviewer behavior for collecting administrative information

Admin I **collected date of interview, interviewer ID, respondent ID**		
	n	Percent
Incorrect	94	43.32
Correct	123	56.68
Total	217	100.00
Admin II **collected permission to record interview from respondent**		
	n	Percent
Incorrect	52	23.96
Correct	165	76.04
Total	217	100.00

Note. Data based on 107 interviewers and 217 recordings.

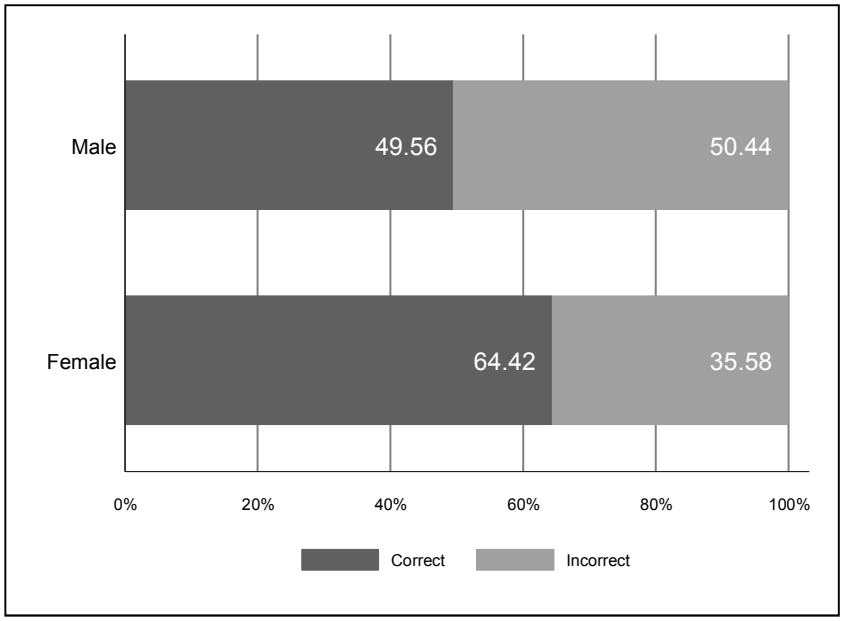
5.3.1 Administration

The interviewers were asked to record some administrative information, such as the date of the interview or the interviewer ID. The results presented in

Table 5.4 show that, in 43% of the recordings, either the date of the interview, the respondent ID or the interviewer ID were missing on the recording (admin I). Furthermore, it was a formal requirement for interviewers to record the permission of the respondent for recording the interview (admin II). In almost 25% of the cases, the recording was not announced in the standardized way; i.e., according to the instructions the interviewers had received. As

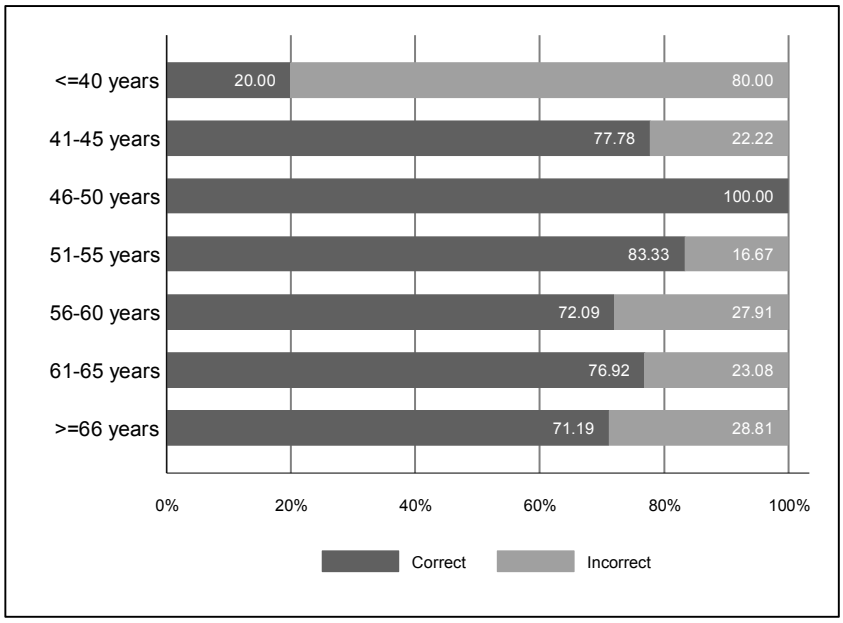
already mentioned, we did not find any case in which recordings were not announced at all to respondents.

Crossing admin I with interviewers´ characteristics revealed that there were significantly fewer mistakes in recording the date of the interview, the interviewer ID, as well as the respondent ID in interviews conducted by female interviewers, compared to interviews conducted by their male colleagues (Figure 5.3). In terms of age, working experience, education, and working hours, a clear pattern was not evident. Results of a logistic regression that included all five interviewer characteristics in one model did only reveal a positive significant association with gender (Odds Ratio = 0.1853, $p = 0.048$).



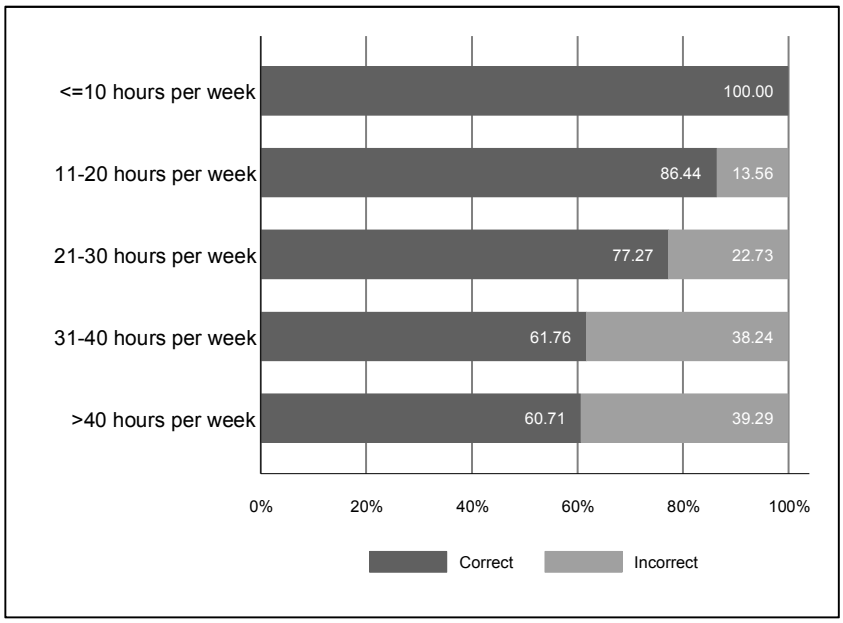
Notes. $\chi^2 = 4.874$. $p = 0.027$. Data based on 107 interviewers and 217 recordings.

Figure 5.3 Interviewer behavior for collecting administrative information I and interviewer´s gender



Notes. $\chi^2 = 17.2574$. $p = 0.008$. Data based on 107 interviewers and 217 recordings.

Figure 5.4 Interviewer behavior for collecting administrative information II and interviewer´s age



Notes. $\chi^2 = 14.8856$. $p = 0.005$. Data based on 98 interviewers and 199 recordings.

Figure 5.5 Interviewer behavior for collecting administrative information II and interviewer´s working hours per week

For collecting permission to record the interview (admin II), our analyses yielded a significant association with age and working hours per week: For age, no clear pattern was found (Figure 5.4). However, we found significantly more mistakes in interviews conducted by interviewers with longer working hours per week (Figure 5.5). For interviewers´ gender, experience, and education, a significant association was not evident. Results of a logistic

regression including all five interviewer characteristics in one model supported these results: a significant negative association was present only for working hours per week (Odds Ratio = 0.588, $p = 0.001$).

5.3.2 Completeness

We investigated several aspects of completeness, including the correct use of filters (questions not incorrectly skipped) and the accuracy of reading a question as scripted. Starting with the number of incorrectly skipped questions, our results showed that, in 55% of the recordings, every question was read out (Table 5.5). In 27% of the cases, up to two questions were incorrectly skipped and, in 10%, five or more questions were incorrectly skipped. No significant differences were identified for any of the tested interviewer characteristics, neither through cross tabulation nor with a linear regression[18].

Table 5.5 Interviewer behavior regarding incorrect skipping of questions

Number of incorrect skipped questions	n	Percent
0	120	55.30
1	43	19.82
2	16	7.37
3	8	3.69
4	8	3.69
>= 5	22	10.14
Total	217	100.00

Notes. Data based on 107 interviewers and 217 recordings. On average, around 160 questions were asked per case.

With regard to reading questions as they are scripted (e.g. shortening or rewording), our results showed that, in 58% of all recordings, up to ten questions were read incorrectly. Additionally, more than ten questions were not read correctly in 26% of the recordings (see Table 5.6). Examples of how interviewers reworded questions are provided in section 2.2. No significant differences were identified for any of the tested five interviewer characteristics, using cross tabulation or a linear regression model[19].

[18] Results available from corresponding author upon request.
[19] Results available from corresponding author upon request.

Table 5.6 Interviewer behavior regarding incorrect reading of questions

Number of incorrect read questions	n	Percent	Cummul. percent
0	35	16.13	16.13
1	32	14.75	30.88
2	17	7.83	38.71
3	16	7.37	46.08
4	12	5.53	51.61
5	13	5.99	57.60
6	10	4.61	62.21
7	8	3.69	65.90
8	5	2.30	68.20
9	9	4.15	72.35
10	3	1.38	73.73
11 – 20	36	16.59	90.32
21 – 30	12	5.53	95.85
> 30	9	4.15	100.00
Total	217	100.00	100.00

Notes. Cummul. = cumulative. Data based on 107 interviewers and 217 recordings. On average, around 160 questions were asked per case.

Table 5.7 Interviewer behavior regarding probing quality

	n	Percent
Excellent probing	35	29.41
Satisfying probing	62	52.10
Inaccurate probing	22	18.49
Total	119	100.00

Note. Data based on 84 interviewers and 119 recordings.

5.3.3 Probing

In almost all recorded interviews, respondents triggered interviewers to probe for at least one question (96%). In these cases, 29% of the interviewers performed excellently, probing was satisfactory in 52%, and probing was inadequate in almost 19% (Table 5.7). No significant association was found for any of the five tested interviewer characteristics[20].

5.4 Discussion

Using data from PIAAC Germany, we provide detailed information on interviewers' behavior regarding several aspects of standardized interviewing techniques, such as using correct filters without skipping questions incorrectly, reading questions as scripted, and neutral

[20] Results available from corresponding author upon request.

communication. Furthermore, we investigated how interviewers´ background characteristics were associated with deviations from the expected behavior with regard to these standardized interviewing techniques. During field work, some problems – such as incorrect reading of questions or incorrect probing[21] – were detected; analyses of interviewer behavior therefore seemed worthwhile. The overall results showed that the majority of the interviewers fulfilled the requirements and predominantly used standardized interviewing techniques. Some further analyses focused on the following aspects: Do the interviewers capture administrative information correctly? Do interviewers read each single question correctly (including answer categories)? Do interviewers probe accurately?

Capturing administrative information is one part of interviewers´ daily work. Nonetheless, over 40% of interviewers did not correctly capture information, such as their own interviewer ID, on the recordings, and, in almost 25% of the cases, the interviewers did not announce the recording in the mandatory way. We consider the source of this error to be the way interviewer trainings are typically conducted. Usually, interviewer trainings in Germany have focused on providing study-specific information, such as how specific questions need to be administered. We assume that aspects of interviewers´ daily work, especially accuracy of simple tasks, are covered in more general trainings that are often only conducted at the beginning of an interviewer's career. According to our analyses, there is a need to improve interviewers´ understanding on how important it is to accurately capture administrative data, for example, for monitoring and controlling the fieldwork.

Another aspect of a standardized survey interview is that each single question is read completely as it is scripted. On average, around 160 questions were asked per case in the PIAAC background questionnaire. Results showed that, in almost half of the recorded interviews, interviewers incorrectly skipped at least one question and, in one fourth of the interviews, they even skipped more than two questions incorrectly. Additionally, in approximately one third of the

[21] In total, 14 out of 129 interviewers were identified who had major problems with their interviewing technique and, consequently, received re-training.

recorded interviews, more than ten questions were not read out as scripted. Instead of reading out the question, interviewers, for example, used information from the previous part of the interview to answer the question by themselves. Yet, by not reading a question at all, interviewers "may overlook specific terms of questions or specific situations that the respondent did not report" (Ongena, 2005, p. 25). There is a real chance that the resulting data is incorrect and results drawn from this data contain errors. The same applies for reworded questions: While slightly rewording a question might have no, or even a positive effect, e.g., Haan et al. (2013), major deviations are more likely to affect the accuracy of responses (see also Ongena & Dijkstra, 2006; Smit, Dijkstra, & Van der Zouven, 1997). Differences across respondents may thus be artifacts of the effect interviewers had during the response process (Fowler & Mangione, 1990).

Furthermore, we examined the probing quality: for about one third of the interviews, the probes were excellent. However, we identified inaccurate probing in one fifth of our recordings (e.g., directive probing or providing incorrect information). According to Smit et al. (1997), suggestive probing has an impact on respondents´ answers and can be considered to be a serious problem. Again, interviewers have to be made aware of the importance of correct probing and should be continuously trained and re-trained to make proper use of interviewer instructions and supportive material.

In most cases, we did not find significant differences in deviant behavior with regard to standardized interviewing techniques that were related to interviewers´ characteristics (gender, age, education, experience, and working hours). With respect to education it is not surprising that significant differences are not found, because the level of education among the interviewers is relatively homogeneous. On the other hand, some associations were identified. For example, our data showed that, for interviews conducted by female interviewers, fewer mistakes were made in capturing administrative data such as interviewer or respondent ID. This might be mediated through other factors, because, for example, women tend to be more conscientious (Weisberg,

DeYoung, & Hirsh, 2011). Training and monitoring activities could be adapted accordingly to intensify the attention on the way men perform their work.

Our results showed that, for interviews conducted by interviewers who reported having longer working hours per week, permission to record the interview was significantly less frequently collected. The interviewers´ workload is likely to have an effect on the accuracy of interviewers' daily work. The amount of time interviewers can spend per case is lower when they have many cases to work on. Survey administration should ensure that interviewers´ workload is manageable, as, for example, already stated early in the fifties by Collins (1980) and recently confirmed by Japec (2008), since this is one way of reducing interviewers´ burden. However, it is not always possible to reduce interviewers´ workload, for example, due to the availability of interviewers. Additionally, we are aware that some of the interviewers work for more than one survey agency, which we, unfortunately, cannot account for in this analysis.

Although interviewers were aware of the recordings, because they started the recording themselves manually, our results showed that interviewers did not always follow standardized interviewing techniques. In this study, some interviewers received feedback on their interviewing techniques after we had reviewed their audio recordings. Accordingly, they might have adapted their behavior. However, we have not checked their behavior again and we only provided feedback to those interviewers for whom we detected serious deviant interviewer behavior. According to Biemer (2010), interviewers tend to divert from standardized procedures in the same way over repeated interviews (e.g., they always read out a particular question incorrectly). In summary, we consider that recordings are a good way to gain information on interviewers´ overall behavior, and we assume that our results can be generalized across interviews.

5.5 Conclusion and Outlook

In PIAAC Germany, extensive interviewers trainings were conducted, which is relatively uncommon in Germany (Zabal et al., 2014, p. 54f). An emphasis was placed on the importance of standardized interviewing techniques. However, even with this more intense training, it was

not possible to completely avoid deviant interviewer behavior with regard to standardized interviewing. This suggests that, in many surveys, the problem of deviant behavior is underestimated. Of course, as interviewers are human beings, some degree of deviation from the standardized script has to be expected. Nonetheless, deviations may affect data quality and thus results in quantitative studies conducted by interviewers.

Our analyses did not show many associations between interviewers´ behavior, with regard to standardized interviewing techniques, and interviewers´ background characteristics. Thus, the trainings might have been effective in reducing the variability between interviewers (see also Collins, 1980). This is consistent with our preliminary analyses with regard to interviewer effects on cooperation, using the same database. Here, we find that only 1.7% of the variability in cooperation rate can be attributed to interviewers (Blom et al., 2013; Massing & Ackermann, 2013). In comparison to similar surveys, which report interviewer effects of approximately 7% (Blom et al., 2013), this is particularly low. Another explanation for the lack of associations between interviewers´ background characteristics and deviant interviewing might be that interviewer characteristics other than socio-economic ones are more important in this respect (for an overview see Schaeffer et al., 2010).

Deviations from standardized interviewing techniques result in inhomogeneous answers and hence may reduce the quality of the data or introduce measurement error, and should therefore be minimized. Several studies have already concluded that formulating good survey questions, intensive, tailored interviewer training and supervision as well as several monitoring strategies are a good way to minimize such effects. Based on a joint analysis of interviewers´ success in gaining contact or cooperation and measurement, Brunton-Smith et al. (2012) suggest monitoring measures of interviewers´ success, such as the contact or cooperation rate, which are indicators of key aspects of interviewer performance. This can lead to significant improvements in overall survey quality. We suggest, additionally, checking measures related to data quality by using recordings and giving feedback to interviewers on a regular basis during fieldwork. Simply

training interviewers before they start to work might not be enough to keep them motivated and to ensure that they work consistently in the best possible way throughout the entire field period.

In this paper, our intention was not to explain interviewer effects but rather to demonstrate how interviewers deviated from expected behavior with regards to standardized interviewing techniques and to examine first associations between deviations and interviewers´ background characteristics. Further analyses that make use of the rich data PIAAC Germany offers are necessary to explain the results. For example, other interviewer characteristics, such as interviewers´ attitudes and expectations, respondents´ characteristics or question characteristics can be used to explain deviation from standardized interviewing techniques. Based on analyses by Brunton-Smith et al. (2012), a combination of the relationship of interviewers´ contact behavior and their workload is also worth analyzing. It would also be worthwhile to address the important issue of question quality, in order to reduce interviewers´ burden.

6 Study 3: Interviewer Effects on Estimates of Substantive Survey Variables in PIAAC Germany

When it comes to errors during face-to-face interviews, standardized interviewing techniques are commonly used as a strategy to reduce errors introduced by interviewers (e.g., Fowler & Mangione, 1990; Loosveldt, 2008; Lyberg, 2012; Mangione et al., 1992; OECD, 2013b). However, many studies describe that interviewers deviate from standardized survey interviewing techniques (for an overview see Maynard & Schaeffer, 2002). In Chapter 5, I described these deviations for PIAAC Germany. It is well known, that interviewers' deviations from standardized interviewing techniques, introduce measurement error into survey data (e.g. Sinibaldi, Durrant, & Kreuter, 2013). Measurement error is the observable gap between the ideal measurement and the response obtained (Groves et al., 2009b). With regard to interviewer effects on estimates of substantive survey variables, this means that answers respondents provide to substantive survey questions are biased by interviewers, because interviewers differ in recording these answers.

In the present chapter, I evaluate interviewer effects on the estimates of a large number of substantive survey variables for PIAAC Germany. More specifically, in my analyses I compare the amount of variance introduced by interviewers in two types of estimates of substantive survey variables: Means of survey interview variables collected interviewer-administered by means of computer-assisted personal interviewing (CAPI) and mean scores of direct measures of respondents' competencies from a self-completion part of the interview by means of computer-assisted self-interviewing (CASI). I aim to not only to *identify* interviewer effects on the estimates of substantive survey variables, but also to *explain* some of the observed differences across interviewers by means of interviewer characteristics. Overall, I find interviewer effects for all estimates of substantive survey variables I look at. I can explain these interviewer effects for some of the surveyed estimates of substantive survey variables by means of available interviewer characteristics. However, I find no interviewer characteristics which explain interviewer effects across all estimates of the substantive survey variables. In addition, I find differences between the

interviewer-administered part and the self-completion part of the survey interview regarding which interviewer characteristics explain parts of the identified interviewer variance in the estimates of substantive survey variables.

In the following paragraphs, I review the literature on interviewer effects on estimates of substantive survey variables with a focus on explaining these effects as well as a focus on differences in interviewer effects between interviewer-administered parts and self-completion parts of survey interviews. Then, I describe the data and the methods I use to analyze the data. Finally, I present and discuss the results.

6.1 Previous Research

There are numerous studies explaining interviewer effects on estimates of substantive survey variables by means of interviewer characteristics such as interviewer socio-demographics, expectations, attitudes or behavior. I briefly summarize these findings in the next paragraphs. A detailed overview of the results in the literature can also be found in West and Blom (2017).

Studies that examine interviewer effects on estimates of substantive survey variables which focus on interviewers' gender to explain the observed differences between male and female interviewers generally find that the effect of interviewers' gender is mostly moderated by respondents' gender and the sensitivity of the questions to gender-related topics (Flores-Macias & Lawson, 2008; Landis, Sullivan, & Sheley, 1973; Lui & Stainback, 2013; Nealon, 1983; O'Muircheartaigh & Campanelli, 1998; and many others). This is often explained by the concept of gender stereotypes (e.g. Tourangeau, Couper, & Steiger, 2003). Although this topic has been studied quite extensively, the literature does not provide researchers with any suggestions on how the effect of interviewers' gender could be avoided.

Many studies also assess the relationship of interviewers' age and survey data quality. These studies report either no relationship (e.g. Collins, 1980; Olson & Bilgen, 2011), some studies report that older interviewers collect less accurate data (Brüderl, Huyer-May, & Schmiedeberg, 2013; Ehrlich & Riesman, 1961; Sudman & Bradburn, 1974), and some studies show opposite

results (Cleary, Mechanic, & Weiss, 1981; Sudman & Bradburn, 1974). The latter ones suggest the positive relationship between data quality and interviewers' age to be related to interviewer experience. This points in the same direction as findings from Freeman and Butler (1976) or Moum (1998), who show that interviewers' age moderates effects caused by other interviewer characteristics.

Results of studies that explain any interviewer effect on estimates of substantive survey variables by means of interviewer experience vary a lot. This is mainly caused by differences in the definition of interviewer experience: Some studies use the overall experience of interviewers as survey interviewers; some studies use the within-survey experience and some studies use the current-survey experience instead. For instance, there are many studies that reveal that interviewers with more overall interviewer experience produce data of poor quality compared to interviewers with less overall interviewer experience. Studies trying to explain this result, point in several directions: Interviewers with more overall interviewer experience tend to read questions less accurately or follow protocols less carefully (Gfroerer et al., 2002), they achieve fewer positive answers to screening questions which also results in shorter questionnaires (Matschinger, Bernert, & Angermayer, 2005), and they receive fewer answers for count questions (Cannell, Marquis, & Laurent, 1977). Contrary to that, there are also studies that report a positive relationship of the overall interviewer working experience and answers to open-ended questions (Feldman, Hyman, & Hart, 1951-52), coding of questions (Stember & Hyman, 1949-50), interviewer observations and paradata (Feldman et al., 1951-52; Sinibaldi et al., 2013).

Furthermore, interviewers' behavior has been studied in many ways to explain interviewer effects on estimates of substantive survey variables. In general, most studies report that deviations from standardized survey interviewing – that may occur intentionally or unintentionally – have an impact on the data quality (for an overview see Maynard & Schaeffer, 2002). More precisely, interviewers have been found to affect data quality by skipping questions (Fowler & Mangione, 1990; Ongena, 2005), rewording questions (Billiet & Loosveldt, 1988;

Brenner, 1982; Haan et al., 2013; Ongena, 2005) or probing with different intensity (Ackermann-Piek & Massing, 2014; Smit et al., 1997). Reasons for interviewers' deviant behavior with regard to standardized survey interviewing techniques are explained by means of the poor quality of survey questions (Schaeffer, 1991; Schaeffer et al., 2010) or due to conversational reasons (Schober & Conrad, 2002). Some studies also suggest interviewers' deviations from standardized interviewing techniques could be avoided by conversationally structured interview styles to enlarge response quality (e.g. Cannell et al., 1981; Houtkoop-Steenstra, 2000; Y. P. Ongena & W. Dijkstra, 2006). However, it is an open question whether this leads to smaller interviewer effects during the survey process. A more detailed review of the literature on interviewers' behavior with regard to standardized interviewing techniques is provided in Chapter 5.

Interviewer-administered versus self-completion parts of survey interviews

Interviewer effects on estimates of substantive survey variables are usually expected when interviewers have an active role in asking survey questions and capturing responses. Instead, interviewer effects on estimates of survey variables from self-completion parts of survey interviews are often found to be lower compared to interviewer-administered parts (Aquilino, Wright, & Supple, 2000; Gfroerer et al., 2002; Tourangeau, Rasinski, Jobe, Smith, & Pratt, 1997; Tourangeau & Smith, 1996; Turner et al., 1998).

Indeed, self-completion survey interview parts reduce the differences in data quality across interviewers, but do not eliminate them (Couper & Rowe, 1996; Kroh, 2004; O'Muircheartaigh & Campanelli, 1999; West & Peytcheva, 2014). Interviewers' physical appearance, the so-called *bystander* effect, is one possible explanation that interviewer effects are also found on estimates of survey variables collected via self-completion survey parts (Aquilino, 1993; Aquilino et al., 2000; Kroh, 2004; Weisband & Kiesler, 1996; West & Peytcheva, 2014). In addition, Mneimneh (2012) or Aquilino (1993) assume that interviewers are not well trained on how to ensure privacy during a self-completion part of a survey interview. Moreover, respondents' behavior is suspected to motivate interviewers to take action instead of following the instructions to let respondents work

independently on the survey instrument during the self-completion part of the survey interview: Respondents may ask for help when reading questions, need clarification of words and question meanings (Mneimneh, Heeringa, Tourangeau, & Elliott, 2014) or they might not be motivated or even able to work completely autonomously on the survey instrument (Couper & Rowe, 1996).

However, interviewer-administered modes have advantages compared to self-completion interview parts, too: Tourangeau and Smith (1996) and Tourangeau et al. (1997) report interviewers increase the motivation of respondents to answer questions and thus reduce item nonresponse, as well as reduce ordering or formatting effects.

6.2 Data Description

The analyses in this chapter are based on the scientific use file (SUF) of the German implementation of PIAAC 2012 (Rammstedt et al., 2016). I am in the fortunate position to not only have access to the survey interview data, but also to a variety of interviewers' characteristics which can be used to explain interviewer effects on estimates of substantive survey variables: Interviewers' age, gender, education, occupational status, experience, working hours, expectations, personality, attitudes, and behavior were collected via an comprehensive interviewer survey. In addition, information on interviewers' actual behavior with regard to deviations from standardized interviewing techniques, derived from audio recordings of survey interviews, is available. Both, the interviewer survey as well as the data on interviewers' behavior retrieved from the audio recordings of the interviews were described in detail in Chapter 3. However, I will briefly describe the data used for the analyses in the present chapter within the next paragraphs (for an overview see Table 6.1).

Table 6.1 Interviewer effects on estimates of substantive survey variables: dependent, explanatory, and control variables

Variables	Data	Data source
Dependent variables	Every day skill use (literacy reading, literacy writing, numeracy, ICT), about yourself (learning strategies, cultural engagement, political efficacy, social trust), number of books, proficiency scores (literacy, numeracy)	Survey interview
Explanatory variables	Interviewers' age, gender, education, occupational status, experience, working hours, expectations, personality, attitudes, behavior	Interviewer survey
	Coded data on interviewers' behavior regarding collecting administrative data, incorrectly skipped questions, incorrectly read questions	Audio recordings
Control variables	PSU size, region	Sampling frame
	Respondents' age, gender, nationality, education, work status, health, social background	Survey interview

Notes. ICT = information and communication technology. PSU = primary sampling unit.

The PIAAC SUF includes 5,465 interviews. However, I exclude 86 cases which were defined as *literacy related nonrespondents* according to the PIAAC technical standards and guidelines and thus were counted as completed case, but the respondents did not receive the interview (see Zabal et al., 2014). The eligible sample thus contains 5,379 cases. Furthermore, the analyses sample of 4,132 cases does not include cases with missing data[22], either on one of the dependent, explanatory or control variables (see Table 6.2).

Table 6.2 Interviewer effects on estimates of substantive survey variables: sample description

	Number of cases
Eligible sample	5,379
Analyses sample	4,132
Number of interviewers	107
Number of PSUs	240

Notes. The section on ICT skill use was asked to persons who use a computer outside their working hours only. Thus models on ICT skill use run on a reduced sample of 3,613 cases.

[22] One possibility to deal with missing data is to assign values to missing items. However, imputation should be based on reliable information about the case for which the data is missing (Lohr, 1999). For the cases mentioned above no or almost no information is available. Thus I do not impute data for these cases.

Control variables: Sample composition characteristics

My main research interest in the present chapter is to explain interviewer effects on estimates of substantive survey variables. In PIAAC, sample persons were not assigned to interviewers randomly. Instead, sample persons are clustered within PSUs and PSUs are clustered within interviewers (see method section of this chapter for a detailed description). I account for this non-random assignment of sample persons to interviewers by controlling for sample composition effects. More specifically, I include sample composition variables derived from the sampling frame: PSU size, region. In addition, I take respondents' characteristics derived from the survey interview into account that are related to the dependent variables (for an overview see also Rammstedt et al., 2013): Age, gender, nationality, education, work status, health status, and social background. For an overview of the variables included in the models, as well as data sources see Table 6.1.

Explanatory variables: Interviewer characteristics

In this chapter, I aim to explain interviewer variance in estimates of substantive survey variables by means of interviewer characteristics collected via a comprehensive interviewer survey conducted with the German interviewers. I described the interviewer survey and the collected data in Chapter 3 in detail. Overall, the interviewer survey yielded a response rate of 89.8% with 115 interviewers taking part in the survey. The questionnaire contained information on interviewers' socio-demographic characteristics, attitudes, personality, behavior as well as expectations with regard to field work outcomes of PIAAC.

For the analyses of interviewer effects on meassurement, I first of all, include interviewers' socio-demographic characteristics age, gender, education, and occupational status (next to their interviewer job) in my models. Each of these characteristics are included in my analyses coded as dummy variables, which I described in Table 3.2 in Chapter 3 in detail. Next to interviewer socio-demographic characteristics, I include the indicators on interviewers' attitudes about reasons for working as an interviewer, behavior regarding standardized interviewing techniques or strategies

on how to achieve response, and interviewer personality traits in my models: *Science*, *people*, *formal*, *tailorContent*, *tailorTime*, *tailorNo*, *slow*, *dialect*, *motivate*, *diligent*, *voluntary*, *answers*, *trust*, *self*, and *others*. The description of the construction for all of these indicators and descriptive statistics are provided in Chapter 3.

Furthermore, I use four indicators on interviewers' actual behavior during the interview with regard to deviations from standardized interviewing techniques which were derived via behavior coding from auto recorded interviews: (1) collecting formal criteria correctly with regard to the date of the interview, the interviewer ID and the respondent ID, (2) collecting informed consent with regard to the permission to record the interview from the respondent, (3) the number of incorrectly skipped questions, as well as (4) the number of incorrectly read questions. A detailed description on how these indicators were set up can be found in Chapter 3 and Chapter 5. Descriptive statistics for all of the four indicators are provided in Appendix B.

Dependent variables I: Means of core questionnaire variables

Previous studies have shown that interviewer effects on estimates of substantive survey variables differ by question type (e.g. Billiet & Loosveldt, 1988; Hox, de Leeuw, & Kreft, 1991; Japec, 2005, 2008; O'Muircheartaigh, 1976; Pannekoek, 1991). To cut through the complexity of the very extensive PIAAC core questionnaire, I select key dependent variables that are more likely to be affected by interviewers from the core questionnaire according to the findings from the literature: Schaeffer et al. (2010) summarize the current state of the literature suggesting that attitudinal, sensitive, ambiguous, complex, and open-ended questions are more likely to be affected by interviewers. Thus, I select variables including instructions for the interviewers because these questions are more likely to cause interviewer effects on the estimates of these variables as deviations from this instruction might occur during the administration of the questions. In addition, I choose variables which are at least ordinally scaled. Next to that, I focus on variables that have a contextual link to the second data set (data from the cognitive test). However, variables *about yourself* and *background* do not have this link with regard to the content.

Instead, they are used for sensitivity analyses to research whether interviewer effects on the estimates of variables of other topic are also identified. Furthermore, I also conduct my analyses for variables from a part that was filtered as this part might be more prone to interviewer effects because interviewers might want to reduce the overall interview time by shortening the interview as much as possible. Accordingly, from the CAPI part of the PIAAC interview, I select the following 36 substantive survey variables for my analyses on interviewer effects on estimates of substantive survey variables [23]:

- Skill use in everyday life
    - 12 variables about literacy skill use (8 about reading and 4 items about writing)
    - 6 variables about numeracy skill use
    - 7 variables about ICT skill use
- About yourself
    - 6 variables on learning strategies
    - 1 variable on cultural engagement
    - 1 variable on political efficacy
    - 2 variables on social trust
- Background
    - 1 variable on the number of books

The descriptive statistics for all dependent variables derived from the CAPI part of the PIAAC interview can be found in Table 6.3. With one exception, all variables are ordinal scaled and have a minimum of 1 and a maximum of 5 with a high mean indicating a high level of skill use or agreement. Variable J_Q08 (number of books) has a minimum of 1 and a maximum of 6, also with a high value indicating a high number of books. The distributions of the variables show various patterns with means ranging from 1.13 (H_Q02b) to 4.32 (H_Q01c).

[23] The Original German version and the English translation of the question text for all variables can be found in Appendix F.

Table 6.3 Interviewer effects on estimates of substantive survey variables: Statistics for the dependent variables of the CAPI part

Topic		Variable	Mean	SE	Min	Max
Everyday skill use	Literacy - reading	H_Q01a	2.76	1.18	1	5
		H_Q01b	4.27	1.10	1	5
		H_Q01c	4.32	1.07	1	5
		H_Q01d	2.80	1.25	1	5
		H_Q01e	3.07	1.43	1	5
		H_Q01f	2.63	1.15	1	5
		H_Q01g	3.51	0.96	1	5
		H_Q01h	2.19	1.17	1	5
	Literacy - writing	H_Q02a	3.88	1.22	1	5
		H_Q02b	1.13	0.48	1	5
		H_Q02c	1.45	0.92	1	5
		H_Q02d	2.32	0.87	1	5
	Numeracy	H_Q03b	2.92	1.34	1	5
		H_Q03c	2.48	1.33	1	5
		H_Q03d	2.86	1.23	1	5
		H_Q03f	1.57	0.92	1	5
		H_Q03g	2.12	1.27	1	5
		H_Q03h	1.32	0.82	1	5
	ICT	H_Q05a	4.15	1.13	1	5
		H_Q05c	3.90	1.04	1	5
		H_Q05d	2.75	1.23	1	5
		H_Q05e	1.96	1.08	1	5
		H_Q05f	2.97	1.19	1	5
		H_Q05g	1.20	0.64	1	5
		H_Q05h	1.99	1.48	1	5
About yourself	Learning strategies	I_Q04b	3.17	0.83	1	5
		I_Q04d	3.92	0.86	1	5
		I_Q04h	3.59	0.86	1	5
		I_Q04j	3.47	0.92	1	5
		I_Q04l	3.44	0.93	1	5
		I_Q04m	4.05	0.81	1	5
	Cultural engagement	I_Q05f	1.77	1.15	1	5
	Political efficacy	I_Q06a	2.71	1.18	1	5
	Social trust	I_Q07a	2.28	1.06	1	5
		I_Q07b	2.24	0.96	1	5
Background	Number of books	J_Q08	3.55	1.39	1	6

Notes. Table based on 4,132 cases for literacy and numeracy skill use, learning strategies, cultural engagement, political efficacy, social trust, and number of books. Table based on a reduced sample of 3,613 cases for ICT skill use. CAPI = computer-assisted personal interviewing. SE = standard error. Min = minimum. Max = maximum.

Dependent variables II: Proficiency scores

Next to interviewer effects in interviewer-administered interview parts, also for self-completion interview parts, interviewer effects have been identified (Couper & Rowe, 1996; Kroh, 2004; O'Muircheartaigh & Campanelli, 1999; West & Peytcheva, 2014). Thus, I also examine interviewer effects on literacy and numeracy proficiency scores collected in the self-completion part of the PIAAC interview by means of CASI, administered under the supervision of the interviewers.

There are ten proficiency scores per competency domain ranging from 0 to 500, where a higher score indicates higher competencies. Typically, when working with the PIAAC proficiency scores all ten scores have to be taken into account using replicate weights to make an inference about the proficiency of the general population (OECD, 2016; Zabal et al., 2014). However, I make no inference about the proficiency of the general population. Instead, I aim to identify interviewer effects in this part of the interview and try to explain parts of the interviewer variance with interviewer characteristics. Thus, I use the mean of the ten proficiency scores for literacy and the mean of the ten proficiency scores numeracy for each respondent. Sample items for both domains can be found in Appendix A.

The descriptive statistics for the two dependent variables *literacy* and *numeracy* derived from the CASI part of the PIAAC interview can be found in Table 6.4. Both variables are continuous and have a minimum of 0 and a theoretical maximum of 500[24] with a high score indicating a high level of the respective competency. Both variables are normally distributed: The mean score of literacy is 275.9 with a minimum of 103 and a maximum of 393, the mean score of numeracy is 277.8 with a minimum of 75 and a maximum of 418.

[24] In theory, the scale ranges from -∞ to +∞.

Table 6.4 Interviewer effects on estimates of substantive survey variables: Statistics for the dependent variables of the CASI part

		Mean	SE	Min	Max
Proficiency score	Literacy	275.9	42.72	103	393
	Numeracy	277.8	47.10	75	418

Notes. Table based on 4,132 cases. CASI = computer-assisted self-interviewing. SE = standard error. Min = minimum. Max = maximum.

6.3 Method

I analyze interviewer effects on the estimates of 38 substantive survey variables by means of three-level random effects linear regression models. More specifically, I estimate linear regressions for the means of 36 CAPI variables and the mean scores of the two CASI variables. By using multilevel-modelling I take the nested data structure of PIAAC Germany into account: Respondents (first level) are nested in PSUs (second level) and PSUs are nested in interviewers (third level; see Figure 6.1).

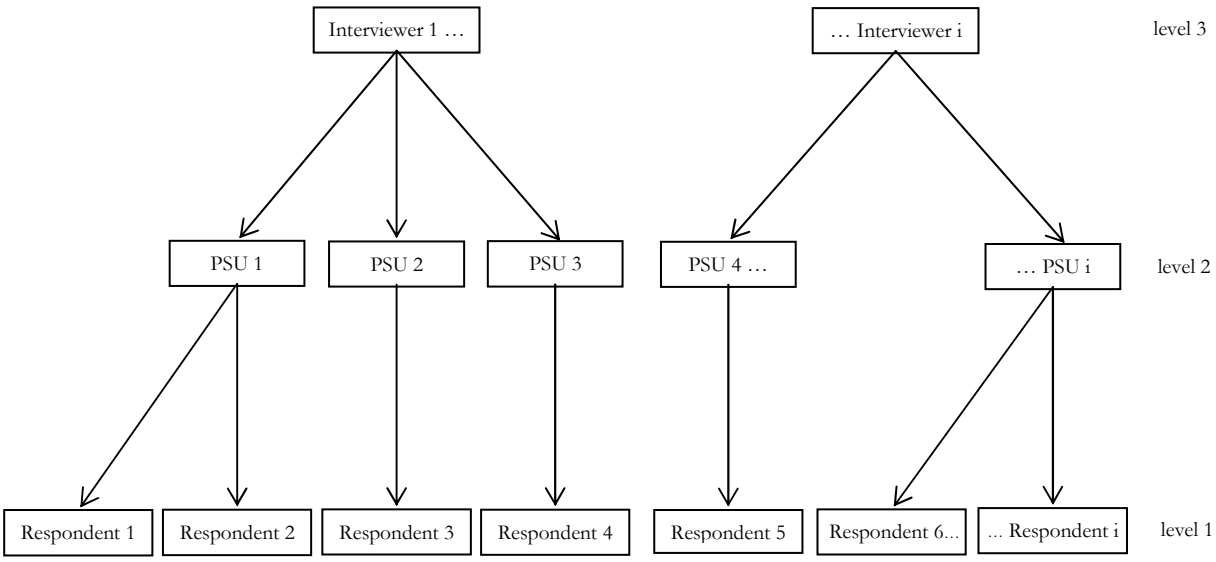


Figure 6.1 Interviewer effects on estimates of substantive survey variables: Data structure

Note. PSU = primary sample unit.

I tested the necessity to estimate three-level models instead of less complex two-level models. Results of the two-level models (see Appendix G) show a substantial part of the variance in the estimates of substantive survey variables under research can be explained by PSU clustering. Thus, I take the PSU and interviewer clustering into account in my analyses of interviewer effects on estimates of substantive survey variables in PIAAC Germany.

Intercept-only model

Following Hox (2010), I first estimate an intercept only model (null model; see equation 6.1):

$$Y_{ijk} = \beta_{000} + v_{00k} + u_{0jk} + e_{ijk}$$

$$\begin{aligned} with \quad & v_{00k} \sim N(0, \sigma^2_{v_{00k}}) \\ & u_{0jk} \sim N(0, \sigma^2_{u_{0jk}}) \\ & e_{ijk} \sim N(0, \sigma^2_{e_{ijk}}) \end{aligned} \tag{6.1}$$

where the outcome Y_{ijk} for respondent *i* nested in PSU *j* nested in interviewer *k* is estimated as the intercept of the regression β_{000} and the residuals v_{00k} for the interviewers, PSUs u_{0jk} and respondents e_{ijk}. All residuals are assumed to follow a normal distribution with variances regarding interviewers $\sigma^2_{v_{00k}}$, PSUs nested within one interviewer $\sigma^2_{u_{0jk}}$ and respondents nested within one PSU and one interviewer $\sigma^2_{e_{ijk}}$.

In the null model, the Intra-Class Correlation Coefficient (ICC) informs about the proportion of the variance explained by the group variable; i.e. PSUs or interviewers. At each level the ICC is estimated as shown in equations (6.2), (6.3) and (6.4; see Hox, 2010). For example, equation (6.2) reflects the estimation of the ICC at the interviewer level (level 3)

$$ICC_{\text{interviewer}_{00k}} = \frac{\sigma^2_{v_{00k}}}{\sigma^2_{u_{0jk}} + \sigma^2_{v_{00k}} + \sigma^2_{e_{ijk}}} \tag{6.2}$$

$$ICC_{PSU_{0jk}} = \frac{\sigma^2_{u_{0jk}}}{\sigma^2_{u_{0jk}} + \sigma^2_{v_{00k}} + \sigma^2_{e_{ijk}}} \tag{6.3}$$

$$ICC_{PSUwithinInterviewer_{0jk}} = \frac{\sigma^2_{u_{0jk}} + \sigma^2_{v_{00k}}}{\sigma^2_{u_{0jk}} + \sigma^2_{v_{00k}} + \sigma^2_{e_{ijk}}} \tag{6.4}$$

where $\sigma^2_{v_{00k}}$ describes the variance over interviewers, $\sigma^2_{u_{0jk}}$ reflects the variance over PSUs nested within interviewers, and $\sigma^2_{e_{ijk}}$ describes the variance over respondents nested within PSUs and interviewers. Equation (6.3) describes the ICC estimation for the proportion of the variance

between PSUs and equation (6.4) describes the ICC estimation for the proportion of the variance between PSUs clustered within one interviewer. I focus on interviewer effects on substantive survey variables, thus ICCs at level 2 – ICC at PSU level and at PSU-within-interviewer-level – are reported, but not interpreted separately.

Full models

Working with data from PIAAC Germany does not allow to fully disentangling PSU and interviewer effects (because sample persons are not randomly assigned to interviewers). To control for sample composition effects, I include various sample composition variables in the full models (see data description of control variables in section 6.2 of the present chapter). An example for a full model including two variables V at each level is provided in equation (6.5):

$$
\begin{aligned}
&\mathrm{Y}_{ijk} = \beta_{000} \\
&+ \beta_1 \mathrm{r\_V1}_{ijk} + \beta_2 \mathrm{r\_V2}_{ijk} \\
&+ \beta_3 \mathrm{psu\_V1}_{ij} + \beta_4 \mathrm{psu\_V2}_{ij} \\
&+ \beta_5 \mathrm{i\_V2}_{k} + \beta_6 \mathrm{i\_V2}_{k} \\
&+ v_{00k} + u_{0jk} + e_{ijk} \qquad (6.5)
\end{aligned}
$$

$$
\begin{aligned}
\textit{with} \quad & v_{00k} \sim N(0, \sigma^2_{v_{00k}}) \\
& u_{0jk} \sim N(0, \sigma^2_{u_{0jk}}) \\
& e_{ijk} \sim N(0, \sigma^2_{e_{ijk}})
\end{aligned}
$$

where the outcome for respondent *i* nested in PSU *j* nested in interviewer *k* is estimated as the intercept of the regression β_{000} and variables at the respondent level $\beta_1 \mathrm{r\_V1}_{ijk}$ and $\beta_2 \mathrm{r\_V2}_{ijk}$,PSU level $\beta_3 \mathrm{psu\_V1}_{ij}$ and $\beta_3 \mathrm{psu\_V2}_{ij}$, as well as interviewer level $\beta_5 \mathrm{i\_V1}_{k}$ and $\beta_6 \mathrm{i\_V2}_{k}$ and the residuals v_{00k} for the interviewers, PSUs u_{0jk} and respondents e_{ijk} with the assumption of a normal distribution of the variance over interviewers $\sigma^2_{v_{00k}}$, variance over PSUs nested within one interviewer $\sigma^2_{u_{0jk}}$, and variance over respondents nested within one PSU and within one interviewer $\sigma^2_{e_{ijk}}$.

6.4 Results

Both, for means of interviewer-administered core questionnaire variables and for proficiency mean scores collected via the self-completion cognitive test, I estimated several three-level linear models in order to answer the question whether there are interviewer effects on estimates of substantive survey variables in PIAAC Germany. I started with estimating null models of the three-level linear regression to identify the percentage of variance introduced by interviewer clustering without including any independent variables at any level.

For both parts of the survey interview, ICCs from the null models at the interviewer level are shown in Figure 6.2 (see also model 1 in Appendix G). Between 0 and 7.3% of the variance in the means of the CAPI variables can be explained because sample persons are clustered within interviewers. For the respondent-completed CASI part, 4.9% of the variance in the literacy mean score and 4.3% of the variance in the numeracy mean score is located at the interviewer level.

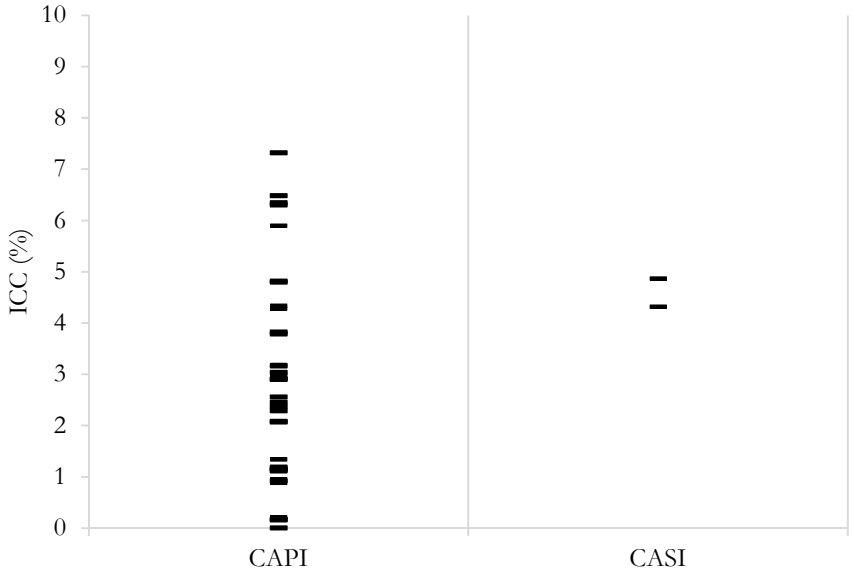


Figure 6.2 ICCs at the interviewer level for the estimates of the CAPI and CASI variables, models 1, three-level linear regression, PIAAC 2012

Notes. CAPI: means of the following core questionnaire variables: H_Q01a, H_Q01b, H_Q01c, H_Q01d, H_Q01e, H_Q01f, H_Q01g, H_Q01h, H_Q02a, H_Q02b, H_Q02c, H_Q02d, H_Q03b, H_Q03c, H_Q03d, H_Q03f, H_Q03g, H_Q03h, H_Q05a, H_Q05c, H_Q05d, H_Q05e, H_Q05f, H_Q05g, H_Q05h, I_Q04b, I_Q04d, I_Q04h, I_Q04j, I_Q04l, I_Q04m, I_Q05f, I_Q06a, I_Q07a, I_Q07b, J_Q08. Question text for CAPI variables can be found in Appendix F. CASI: literacy and numeracy mean scores. Sample items for literacy and numeracy can be found in Appendix A. Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132 (3,613 for H_Q05c). ICC = intra-class correlation coefficient. CAPI = computer-assisted personal interviewing. CASI = computer-assisted self-interviewing.

As mentioned above, sample persons were not assigned to interviewers randomly. Thus, it is not possible to fully disentangle interviewer and area effects. I account for this non-random assignment of sample persons to interviewers by controlling for sample composition effects. More specifically, I control for sample composition effects by including the following sample composition variables in all models (see also section 6.2 of this chapter)[25]: Respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region.

The ICCs at the interviewer level of the models controlled for sample composition effects do not differ considerably from the results of the null models (see Figure 6.3 and model 2 in Appendix G): Between 0 and 6.9% of the variance in the means of the CAPI variables can be explained due to interviewer clustering. For the CASI part of the survey interview the variance in the proficiency mean scores that can be explained due to interviewer clustering is slightly lower in the models that include sample composition control variables, than in the intercept only models (3.7% for numeracy, 3.9% for literacy).

[25] Respondent as well as PSU variance components including standard errors and ICCs at the PSU level and PSU-within-interviewer-level are reported in Appendix G.

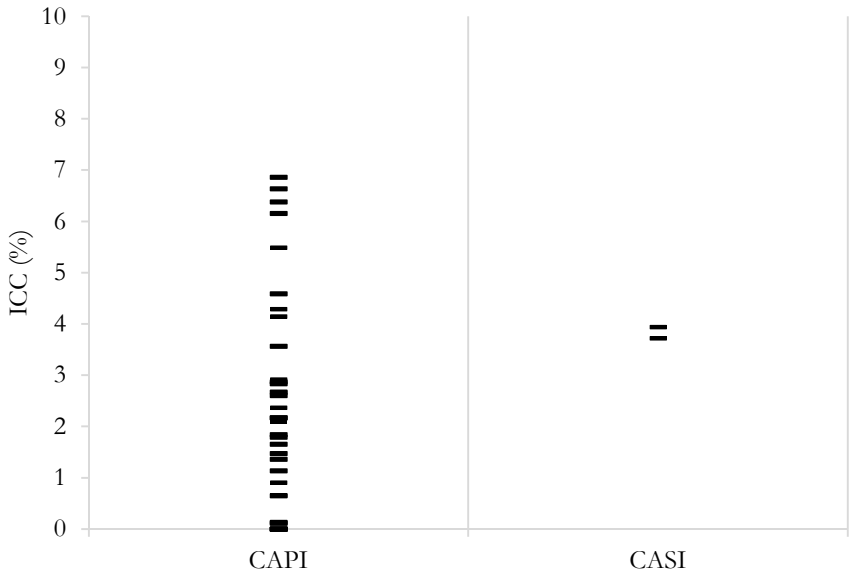


Figure 6.3 ICCs at the interviewer level for the estimates of the CAPI and CASI variables, models 2, three-level linear regression, PIAAC 2012

Notes. CAPI: means of the following core questionnaire variables: H_Q01a, H_Q01b, H_Q01c, H_Q01d, H_Q01e, H_Q01f, H_Q01g, H_Q01h, H_Q02a, H_Q02b, H_Q02c, H_Q02d, H_Q03b, H_Q03c, H_Q03d, H_Q03f, H_Q03g, H_Q03h, H_Q05a, H_Q05c, H_Q05d, H_Q05e, H_Q05f, H_Q05g, H_Q05h, I_Q04b, I_Q04d, I_Q04h, I_Q04j, I_Q04l, I_Q04m, I_Q05f, I_Q06a, I_Q07a, I_Q07b, J_Q08. Question text for CAPI variables can be found in Appendix F. CASI: literacy and numeracy mean scores. Sample items for literacy and numeracy can be found in Appendix A. All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132 (3,613 for H_Q05c). ICC = intra-class correlation coefficient. CAPI = computer-assisted personal interviewing. CASI = computer-assisted self-interviewing.

Next, I compare the amount of the identified interviewer effect between the two survey parts, CAPI and CASI. I find the highest ICCs for the interviewer-administered CAPI part with ICCs over 6% for the means of variables on numeracy everyday skill use (H_Q03b, H_Q03g, H_Q03c) and literacy-writing everyday skill use (H_Q02d). Overall, ICCs are slightly higher for the means of the variables from the CAPI part with an average of 5.3% compared to an average of 3.8% for the proficiency mean scores of the CASI part. For the majority of the estimates of the substantive survey variables (N=17), I find an ICC under 2%, and for even more estimates of the substantive survey variables (N=27) the ICC is under 3.9%, which is the lowest ICC I find for a proficiency mean score of the CASI part (see Appendix G). Nevertheless, the difference in the amount of identified interviewer effect between the two survey interview parts is not very high, although, in the CAPI part the interviewer had an active role in administering the survey questionnaire, whereas the CASI part was self-completed by the respondents.

Explaining interviewer effects

The major research interest in the present dissertation is to investigate, how interviewer effects on estimates of substantive survey variables, collected via the interviewer-administered CAPI part and the self-completion CASI part of the German PIAAC interview, can be explained by means of interviewer characteristics. To answer this question, I estimate three-level linear models for the means of 38 dependent survey variables and add one interviewer characteristic measured through the interviewer survey or collected via the audio recordings per model. In the next step, I compare the amount of interviewer variance between the model including the explanatory interviewer characteristics and the model without the interviewer characteristic (likelihood-ratio test; see Rabe-Hesketh and Skrondal, 2012). Again, all models control for sample composition effects.

A very low ICC indicates a very low interviewer effect on the respective estimate. Such low interviewer effects do not warrant further analyses to explain these low interviewer effects by means of interviewer characteristics. As a consequence, I estimate models including explanatory interviewer characteristics only for those variables where I find an ICC over 3.5% for the means of the variables in the models including sample composition control variables (nine variables of the CAPI part: H_Q01a, H_Q01f, H_Q01h, H_Q02d, H_Q03b, H_Q03c; H_Q03g, H_Q05c, I_Q04l, and the two variables of the CASI part: Literacy and numeracy; see Appendix G).

In Table 6.5, I show for which interviewer characteristics I find a significant effect on the means of at least one of the dependent substantive survey variables at the 5%- and 1%-significance level. Overall, I find significant effects at the 5%-significance-level for the interviewer characteristics in 25% of all models (22 models for 11 estimates of the substantive variables, 61 significant results; see Appendix G). However, for the proficiency mean scores of two dependent CAPI variables, I find no significant effect for any of the available interviewer characteristics (H_Q03b, H_Q03c).

Table 6.5 Significant effects of interviewer characteristics on the estimates of CAPI and CASI variables at the 1% and 5%-significance-level, one interviewer characteristic per model, three-level linear regression, PIAAC 2012

	CAPI							CASI	
	H_Q01a	**H_Q01f**	**H_Q01h**	**H_Q02d**	**H_Q03g**	**H_Q05c**	**I_Q04l**	**Literacy**	**Numeracy**
Experience									
In-study experience			**			***		**	***
Overall work experience			-***		-***		**		
Working hours	-***								
Expectations (overall response rate)								-***	-***
Reasons for working as an interviewer							-**		
How to conduct standardized survey interviews						***		**	**
How to achieve response	-**	-***	-***					-**	-**
Trust							-***		
Data protection concerns	**								
Audio indicators									
Collected formal criteria			-**	-**			**	-**	
Collected permission to record by respondent					-**				
Number of incorrect skipped questions	-***		-**		-**	**			
Number of incorrect read questions	-***		-**		-***	***			

Notes. CAPI variables: Means. Question text for CAPI variables can be found in Appendix F. CASI: Literacy and numeracy mean scores. Sample items for literacy and numeracy can be found in Appendix A. Interviewer characteristics which significantly explain interviewer variance are presented in bigger font. No significant effects at the 1%- or 5%-significance level for the following interviewer characteristics: Age, gender, education, occupation status, expectations (response rate income variable) and social desirability. No significant effects at the 1%- or 5%-significance level for the following core questionnaire variables: H_Q03b, H_Q03c. All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Only one interviewer characteristic included per model. Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132 (3,613 for H_Q05c). CAPI = computer-assisted personal interviewing. CASI = computer-assisted self-interviewing.

** $p < 0.05$, *** $p < 0.01$, - negative coefficient

Interviewer characteristics, which first, are found to have a significant effect on any of the means of the substantive survey variables and second, significantly explain interviewer variance in the means of the respective survey variable, are presented in bold letters in Table 6.5 (likelihood-ratio test; see Rabe-Hesketh & Skrondal, 2012). Overall, the results show that there is no interviewer characteristic that systematically explains interviewer variance across the nine dependent substantive survey variables. However, for most of the means of these survey variables at least one interviewer characteristic is found to significantly explain parts of the interviewer variance. The amount of explained interviewer variance across all estimates of the substantive survey variables ranges from 3.3% till 25.7% with an average of 11.6% (see Table 6.6 and Figure 6.4).

Table 6.6 Percentage explained interviewer variance by interviewer characteristic for which a significant effect on the estimates of CAPI and CASI variables at the 1% and 5%-significance-level was found, by variable and overall variables, one interviewer characteristic per model, three-level linear regression, PIAAC 2012

	Explained interviewer variance by variable (%)									Overall explained interviewer variance (%)		
	CAPI							CASI		CAPI & CASI	CASI	CAPI
	H_Q01a	H_Q01f	H_Q01h	H_Q02d	H_Q03g	H_Q05c	I_Q04l	Literacy	Numeracy			
Experience												
In-study experience			**7.4**			**13.2**		**12.6**	**19.2**	**13.1**	**15.9**	**10.3**
Overall work experience			**12.4**		**11.5**		**13.7**			**12.5**		**12.5**
Working hours	10.3											
Expectations (overall response rate)								**15.9**	**10.2**	**13.0**	**13.0**	
Reasons for working as an interviewer							9.4					
How to conduct standardized survey interviews						**25.7**		10.6	**15.8**	**17.4**	**13.2**	**25.7**
How to achieve response	5.4	3.3	**5.7**					10.3	12.1	**7.4**	**11.2**	**4.8**
Trust							**11.1**					
No/few data protection concerns	**8.1**											
Audio indicators												
Collected formal criteria			8.9	5.1			**10.7**	9.1		**8.5**	**9.1**	**8.2**
Collected permission to record by respondent					8.5							
Number of incorrect skipped questions	**23.6**		**7.8**		**8.4**	**8.1**				**12.0**		**12.0**
Number of incorrect read questions	**20.2**		**7.6**		**9.8**	**10.2**				**11.9**		**11.9**

Notes. CAPI variables: Means. Question text for CAPI variables can be found in Appendix F. CASI: Literacy and numeracy mean scores. Sample items for literacy and numeracy can be found in Appendix A. Interviewer characteristics which significantly explain interviewer variance are presented in bold letters for single variables. No significant effects at the 1%- or 5%-significance level for the following interviewer characteristics: age, gender, education, occupation status, expectations (response rate income variable) and social desirability. No significant effects at the 1%- or 5%-significance level for the following core questionnaire variables: H_Q03b and H_Q03c. Overall explained interviewer variance presented when an interviewer characteristic has a significant effect on more than one substantive variable. All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Only one interviewer characteristic included per model. Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132 (3,613 for H_Q05c). CAPI = computer-assisted personal interviewing. CASI = computer-assisted self-interviewing.

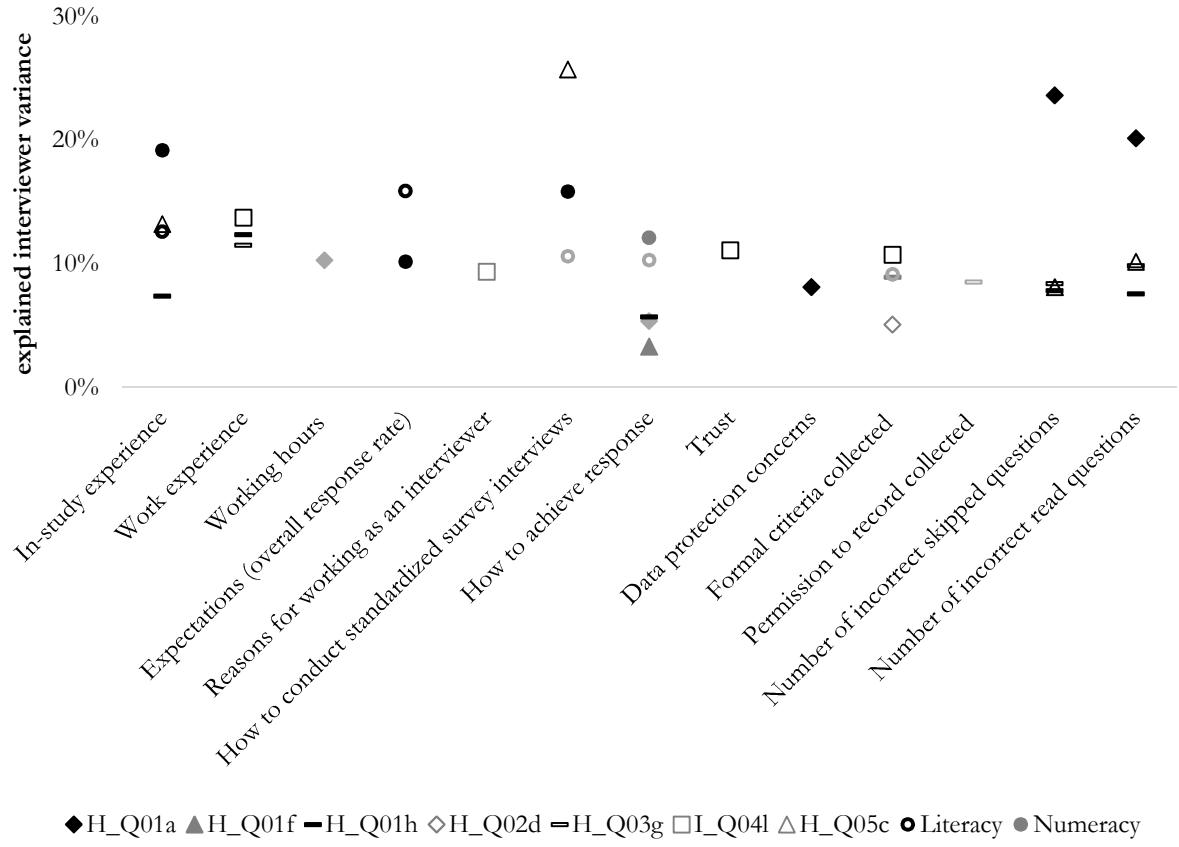


Figure 6.4 Percentage explained interviewer variance by interviewer characteristic for which a significant effect on the estimates of the CAPI and CASI variables at the 1% and 5%-significance-level was found, three-level linear regression, PIAAC 2012

Notes. Core questionnaire variables: Means. Question text for core questionnaire variables can be found in Appendix F. Literacy and numeracy: Mean scores. Sample items for literacy and numeracy can be found in Appendix A. Amount of significant explained interviewer variance is presented in black font. Amount of explained interviewer variance which is not significant presented in grey font. No significant effects at the 1%- or 5%-significance level for the following interviewer characteristics: Age, gender, education, occupation status, expectations (response rate income variable) and social desirability. No significant effects at the 1%- or 5%-significance level for the following core questionnaire variables: H_Q03b and H_Q03c. All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Only one interviewer characteristic included per model. Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132 (3,613 for H_Q05c).

Socio-demographic characteristics

I find no significant effect of the available socio-demographic interviewer characteristics age, gender, education and occupational status on the estimates of substantive survey variables I assessed. Also, interviewers' socio-demographic characteristics do not significantly explain interviewer variance in the means of the survey variables assessed. The latter results indicate that testing random slopes and cross-level interactions between interviewer and respondent

characteristics is not useful to further explain interviewer effects on estimates of substantive survey variables in PIAAC Germany.

Experience

I included two types of interviewer experiences in my models to explain interviewer variance on the means of substantive survey variables: Interviewer in-study experience as well as interviewers' overall working experience. For interviewers' in-study work experience, I find a significant effect on the mean scores of the two CASI variables as well as on the means of the two CAPI variables H_Q01h and H_Q05c (see Table 6.5 and Appendix G) for interviewers, who have in-study work experience compared to interviewers, who have not worked for PIAAC before. Looking at Figure 6.4 and corresponding Table 6.6, it can be seen that the interviewer variance that is significantly explained by interviewers' in-study work experience is 13.1% on average: 15.9% in the mean scores of the CASI variables (literacy: 12.6%, numeracy: 19.2%) and 10.3% in the means of the CAPI variables (H_Q01h: 7.4%, H_Q05c: 13.2%). Instead, for interviewers' overall working experience I find a significant effect on the means of three CAPI variables (H_Q01h, H_Q03g, I_Q04l) for interviewers with more overall work experience compared to interviewers with lower overall work experience. However, no significant effect is found for the proficiency mean scores of the two variables from the CASI part of the interview (see Table 6.5 and Appendix G). The interviewer variance significantly explained by interviewers' overall work experience is 12.5% on average (H_Q01a: 12.4%, H_Q03g: 11.5%, I_Q04l: 13.7%; see Figure 6.4 and Table 6.6). In summary, for interviewers' in-study experience I find significant effects on the means of the substantive survey variables from both parts of the interview, whereas I show that there is no interviewer effect with regard to interviewers' overall experience on the mean scores of the variables of the self-completion part of the interview.

Working hours

Including the amount of hours an interviewer works per week in my models, I find only one significant effect on the mean of a substantive survey variable under research: Interviewers who

work more than 30 hours per week compared to interviewers who work less than 15 hours per week have a significant effect on the mean of the CAPI variable H_Q01a (see Table 6.5 and Appendix G). The interviewer variance explained by interviewers' amount of working hours in the mean of this variable is 10.3%, but not found to be significant (see Figure 6.4 and Table 6.6).

Expectations

With regard to interviewer's expectations, my results presented in Table 6.5 (Appendix G) show a significant effect on the literacy and numeracy mean scores for interviewers who expect a higher overall response rate compared to interviewers who expect a lower overall response rate for their own workload. I find no significant effects for the means of any of the CAPI variables I considered. The interviewer variance significantly explained by interviewers' expectations with regard to their overall response rate in the means of the two CASI variables is 13.0% on average (literacy: 15.9%, numeracy: 10.2%; see Figure 6.4 and Table 6.6).

I have not identified any significant effect of interviewers' expectations with regard to the response rate of the income variable on the means of any of the substantive survey variables I assessed. Furthermore, no interviewer variance is significantly explained in the estimates of substantive survey variables I assessed by including interviewers' expectations with regard to the response rate of the income variable in my models.

Reasons for working as an interviewer

Explaining interviewer effects on the means of substantive survey variables by means of the three indicators on interviewers' reasons for working as an interviewer, show a significant effect for one of the indicators (*science*) on one mean of the CAPI part of the interview only: Interviewers who report higher values on how important it is to work for something related to science or research, have a significant effect on the mean of the CAPI variable I_Q04l (see Table 6.5 and Appendix G). The interviewer variance is reduced by 9.4%, but not found to be significant.

How to conduct standardized survey interviews

In my analyses, I also tested whether interviewers' assessment of their own behavior with regard to following standardized interviewing techniques can explain interviewer effects on estimates of substantive survey variables. I find significant effects on the literacy and numeracy mean scores for interviewers who score higher on the scale for tailoring the speed of speaking according to respondents' needs. Also, for one CAPI variable (H_Q05c), I find that interviewers who score higher on the scale for tailoring their dialect to match respondents' dialect have a significant effect on the mean for this variable (see Table 6.5 and Appendix G). Looking at Figure 6.4 and Table 6.6, it can be seen that the interviewer variance explained by interviewers' assessment of their own behavior with regard to following standardized interviewing techniques is around 17.4% on average: 13.2% in the means of CASI variables on average (10.6% literacy, 15.8% for numeracy) and 25.7% in the mean of the CAPI variable H_Q05c. However, interviewers' assessment of their own behavior with regard to standardized interviewing techniques can only significantly explain interviewer variance in the mean of the CAPI variable H_Q05c.

How to achieve response

With regard to interviewers' behavior, I also tested whether the three indicators on interviewers' assessment of how to achieve sample persons' cooperation can significantly explain parts of the interviewer variance I identified for the means of the substantive survey variables in the PIAAC interview. My results show significant effects on the proficiency mean scores of the two CASI variables literacy and numeracy as well as on the means of three CAPI variables (H_Q01a, H_Q01f, and H_Q01h) for interviewers who reported higher agreeing that plenty of work pays off in achieving respondents' cooperation (indicator: *diligent*; see Table 6.5 and Appendix G). Looking at Figure 6.4 and Table 6.6, it can be seen that the interviewer variance explained by the indicators of interviewers' assessment of how to achieve response in the means of all variables for which significant effects have been identified is 7.4% on average: 11.2% on

average for the CASI part of the interview (literacy: 10.3%, numeracy: 12.1%) and 4.8% on average for the CAPI part of the interview (H_Q01a: 5.4%, H_Q01f: 3.3%, H_Q01h: 5.7%). However, interviewer variance in the means of almost all variables cannot be significantly explained by interviewers' assessment of how to achieve response. Overall, the amount of significantly explained interviewer variance in the means of almost all variables, appear to be lower compared to all other interviewer characteristics.

Trust, social desirability and data protection concerns

I find interviewers who have a higher trust in people to have a significant effect on the mean of the CAPI variable I_Q04l (see Table 6.5 and Appendix G). 11.1% interviewer variance in the mean of this variable is significantly explained by this interviewer characteristic (see Figure 6.4 and Table 6.6). Despite this, no other effects are found for interviewers' trust in people on the means of any of the substantive survey variables under research.

In addition, I have not identified any significant effect of interviewers' social desirability on the mean of any of the substantive survey variables I assessed. Furthermore, no interviewer variance is significantly explained by interviewers' social desirability in the estimates of the substantive survey variables under research.

However, I have identified a significant effect on the mean of one CAPI variable (H_Q01a) for interviewers who are more concerned with regard to the safety of their own personal data (see Table 6.5 and Appendix G). The interviewer variance, significantly explained by adding interviewers' concerns with regard to the safety of their personal data to the model in the mean of this variable is 8.1% (see Figure 6.4 and Table 6.6). Even though, no other effects are found for interviewers' concerns with regard to the safety of their personal data on the estimates of the variables under research.

Audio indicators: Interviewers' actual behavior with regard to standardized survey interviewing

As mentioned above, I´m in the fortunate position to use data on interviewers' actual behavior with regard to following standardized survey techniques during a survey interview. More

specifically, I use four indicators derived from audio recordings on interviewers' behavior with regard to deviations from standardized interviewing techniques to explain parts of the interviewer variance on estimates of substantive survey variables. My results show significant effects of these indicators on the means of seven out of nine substantive variables: H_Q01a, H_Q1h, H_Q02d, H_Q03g, H_Q05c, I_Q04l, and literacy (see Table 6.5 and Appendix G).

For the first indicator, about collecting formal criteria, I find a significant effect on the means of three CAPI variables (H_Q01a, H_Q02d, I_Q04l) for interviewers, who either made mistakes in all interviews or in one interview compared to interviewers where no mistakes with regard to collecting formal criteria was detected on the auto recordings. In addition, I find a significant effect of the audio indicator about collecting formal criteria on the literacy mean score for interviewers who made mistakes in collecting the date of the interview, interviewer identification number (ID) or respondent ID in one interview, but not in the second interview compared to interviewers where no mistakes with regard collecting these formal criteria were detected on the audio recordings. Interviewers' behavior with regard to collecting these formal criteria correctly, can explain 8.2% interviewer variance in the means of the three CAPI variables on average: 8.9% for H_Q01h, 5.1% for H_Q02d and 10.7% for I_Q04l (see Figure 6.4 and Table 6.6). However, no interviewer variance can significantly be explained by this audio indicator in the literacy mean score (see Figure 6.4 and Table 6.6).

For the second audio indicator, which is about collecting the informed consent to record the interview of the respondent, I find only one significant effect on the mean of one CAPI variable (H_Q03g) for interviewers who made no mistakes regarding the collection of these formal criteria compared to interviewers who collected the permission incorrectly (see Table 6.5 and Appendix G). However, no interviewer variance can significantly be explained by this audio indicator in the mean of the CAPI variable H_Q03g (see Figure 6.4 and Table 6.6).

The last two interviewer characteristics I included in my models are the two audio indicators on interviewers' behavior with regard to how many questions were incorrectly skipped or read.

For those two audio indicators, I find a significant effect on the means of four CAPI variables for interviewers who made more mistakes (H_Q01a, H_Q01h, H_Q03g and I_Q05c; see Table 6.5 and Appendix G). I identified no significant effect of these two indicators on the mean scores of the two CASI variables literacy and numeracy. For all variables where I identified significant effects for these two indicators about incorrect skipping and reading of questions, about 12% interviewer variance in the means can be explained significantly. For H_Q01a, I find the highest amount of explained interviewer variance in the mean (number of incorrect skipped questions: 23.6% and number of incorrect read questions: 20.2%). The amount of explained interviewer variance in the means of the other three CAPI variables is between 7.6% and 10.2% (see Figure 6.4 and Table 6.6).

Furthermore, I test whether the independence assumption of linear regressions is violated when I included more than one indicator of interviewers' attitudes on the reasons for working as an interviewer, how to conduct standardized survey interviews, and how to achieve response in my models and find no multi-collinearity between the indicators.

6.5 Discussion and Conclusion

My main research interest of the present chapter is to identify and explain interviewer effects on estimates of substantive survey variables by means of interviewer characteristics for two kinds of survey data: An interviewer administered core questionnaire and a self-completion test of cognitive competencies. There are three main findings, which are discussed within the next paragraphs.

First, I find interviewer effects for all estimates of the substantive survey variables I assessed and second, there is no big difference in the amount of the interviewer effects on the estimates of the substantive survey variables between the interviewer-administered CAPI part and the respondent-administered CASI part of the survey interview. In more detail, I find the variance located at the interviewer level to be between 0% and 7% for the means of the variables from the CAPI part and to be about 4% for the mean scores of the variables from the CASI part of the

interview. Although, the interviewer effects are, overall, lower on the means of the variables of the CASI part, there is no big difference in the amount of the interviewer effects between the two survey parts. This is in line with findings from the literature: Self-completion parts of survey interviews reduce but do not eliminate interviewer effects on estimates of substantive survey variables (Couper & Rowe, 1996; Kroh, 2004; O'Muircheartaigh & Campanelli, 1999; West & Peytcheva, 2014). Instead, due to bystander effects it might not be possible to fully eliminate interviewer effects on estimates of substantive survey variables (Aquilino, 1993; Aquilino et al., 2000; Kroh, 2004; Weisband & Kiesler, 1996; West & Peytcheva, 2014). Emphasizing privacy issues as well as the correct implementation of probing techniques during interviewer trainings, could further reduce interviewer effects on estimates of substantive survey variables in interviewer-administered, but also in self-completion interview parts (Aquilino et al., 2000; Gfroerer et al., 2002; Tourangeau et al., 1997; Tourangeau & Smith, 1996; Turner et al., 1998). Next to that, I find no big differences in the amount of the identified interviewer effects on the estimates of substantive survey variables by question topic or question type. I have theoretically preselected variables which are prone to interviewer effects, thus, this finding implies that interviewer effects are not limited to estimates of a special type of question or interview part. Instead, interviewer effects should be expected and estimated independently of interviewing style and question type.

The third main result of the present chapter is that the available interviewer characteristics explain between 1% and 26% of the interviewer variance in the means of the substantive survey variables under study. I find no interviewer characteristics that explain interviewer effects across the estimates of all substantive variables I assessed. However, I find interviewers' experiences as well as interviewers' actual behavior with regard to the amount of incorrect skipped and incorrect read questions to be most relevant to explain interviewer variance in the estimates of the substantive survey variables. I further discuss these results in more detail within the next paragraphs.

Explaining interviewer effects on estimates of substantive survey variables by means of the two measures of interviewers' experience, which are available for the present study, show the following results: Interviewers' in-study experience significantly explains variance in the means of both parts of the interview, whereas interviewers' overall experience does only explain variance in the means of the interviewer-administered part of the interview. This is in line with results reported by e.g. Gfroerer et al. (2002), who report that less experienced interviewers are mostly found to produce data of poor quality because they work less accurate. Effects of less accurate work on estimates of substantive survey variables are especially relevant, when interviewers have an active role, which is not the case in the respondent-administered interview part. I suggest to make interviewers aware of the effects their experience can have on the data quality (e.g. during interviewer trainings) to increase the change that differences between interviewers with different experience is reduced. Typically, interviewers are instructed to work accurately without showing them consequences of less accurate working styles. I assume this to be the case especially for more experienced interviewers, because interviewer trainings are typically shorter or omitted for more experienced interviewers. Next to that, it is difficult to reduce the effect caused by differences between interviewers' who have and who have no in-study experience. This is mainly the case, because during a survey fieldwork, interviewer stuff has to be refreshed from time to time due to organizational reasons. More research is needed to identify approaches to reduce interviewer effects on estimates of substantive survey variables caused by different in-study interviewer experience.

In explaining interviewer effects during the survey interview process by means of interviewers' actual behavior regarding standardized interviewing techniques I show that, especially the number of incorrect skipped and read questions can significantly explain interviewer variance in the means for most of the CAPI variables under research, but is not relevant to explain interviewer variance in the proficiency mean scores of the CASI part of the PIAAC interview. Results of other studies suggest poor survey questions or conversational

reasons as main source for incorrectly skipped and read survey questions by interviewers (Cannell et al., 1981; Houtkoop-Steenstra, 2000; Schaeffer, 1991; Schaeffer et al., 2010). This is in line with the findings here, because the indicators on the number of incorrect skipped and read questions are only found to explain differences between interviewers for estimates of substantive survey variables of the CAPI part of the PIAAC interview, where the interviewer has an active role in administering the survey questionnaire. However, the identified interviewer effects on the mean scores of substantive survey variables in the CASI part of the interview could be explained by the so-called *bystander* effect: Interviewers might differ in the way they physically appear or how they behave while respondents work on the cognitive test individually (see for example Aquilino et al., 2000; Kroh, 2004; Weisband & Kiesler, 1996; West & Peytcheva, 2014). Again, it might help to reduce interviewer effects by showing consequences of incorrect skipping and incorrect reading of survey questions to interviewers.

It is also of interest to compare the results of the explanatory power of the two interviewer characteristics about interviewers' behavior regarding deviations from standardized interviewer behavior: Interviewers' report of how they follow standardized interviewing techniques – collected via the interviewer survey – and their actual behavior – derived from audio recorded interviews. My results show that the self-report does not significantly explain differences between interviewers in most of the estimates of the survey variables. Instead, the indicators on interviewers' actual behavior regarding standardized interviewing techniques significantly explain parts of the interviewer variance at least in some estimates of the CAPI variables. This leads to the conclusion, that interviewers' behavior with regard to standardized interviewing techniques should be studied in more detail by means of data on their *actual behavior instead using only self-reported information*. Besides, coding interviewers' behavior from audio recorded interviews with regard to other topics than deviations from standardized interviewing techniques might be worth to gain insights in the reliability of interviewers' report of other topics, like for example, how they document the recruitment process gain sample persons' cooperation. More research is needed to

find possibilities of a practice-oriented implementation of quality control of interviewers' behavior during the field work.

Next to interviewers experience and interviewers' actual behavior regarding deviations from standardized interviewing techniques, I have tested a large number of other interviewer characteristics to explain interviewer effects on estimates of substantive survey variables. I find significant effects on some of the means of substantive survey variables as well as significantly reduced interviewer variance in the means of some substantive variables for interviews' expectations of their own response rate as well as for the indicators on how important different aspects on how to achieve response are rated by interviewers. Besides, for some interviewer characteristics, I only find significant effects for one or two of the assessed estimates of the substantive survey variables and almost no interviewer variance significantly reduced: Working hours per week, indicators with regard to reasons for working as an interviewer and indicators regarding how to conduct standardized survey interviews, trust in people and data protection concerns. However, I have not identified any hypotheses to explain these effects, why I expect an unobserved moderate influence for both effects, probably caused by the sample composition. In other words, I expect these interviewer effects on the estimates of the substantive survey variables to be caused by systematic differences in the sample between interviewers which I could not control for.

Moreover, there are some interviewer characteristics for which I find no significant effect for any estimate of the substantive variables I assessed: All socio-demographic characteristics, expectations with regard to the response rate of the income variable, and the social desirability indicator. I assume effects of interviewers' socio-demographic characteristics on estimates of substantive survey variables to be study-specific, because the results of other studies vary a lot. Thus, these interviewer characteristics should be studied on study-specific bases.

Overall, I find no clear pattern in the explanatory power of interviewer characteristics for interviewer effects on estimates of substantive survey variables: I show that different interviewer

characteristics explain interviewer effects on the estimates of the variables of the different parts of the survey interview and I find a large number of interviewer characteristics which are not relevant to explain differences between interviewers in the estimates of substantive survey variables in PIAAC Germany. More research is needed to identify *systematic effects* of interviewer characteristics on estimates of substantive survey variables, to be able to intervene and to ensure high quality of survey data. However, as it is not possible to fully eliminate differences between interviewers, I recommend to control and adjust for available interviewer characteristics for analyses conducted with data that was collected interviewer administered. To prevent interviewer effects on estimates of substantive survey variables, I recommend taking into account the different effects of interviewer characteristics when interviewers are hired and, especially, when interviewers are trained to reduce interviewer effects during the survey interview process. In addition, evaluating interviewer effects during the data collection period is worth to implement appropriate re-trainings according to any identified abnormalities on estimates of substantive survey variables.

I conclude that it is worth studying interviewer effects on estimates of substantive survey variables, instead only studying interviewer effects survey unit nonresponse, which is in accordance to critics from the literature (Lessler, Eyerman, & Wang, 2008; Wang et al., 2013). However, to be able to assess whether successful interviewers with regard to recruiting sample persons are also the ones ending up with lower interviewer effects on estimates of substantive survey variables, the different error sources described in the TSE framework should be studied simultaneously. Thus, I present analyses of multiple survey error sources described in the TSE framework in the next chapter.

7 Study 4: Interviewer Effects for Multiple Aspects of Survey Error in PIAAC Germany

To date, there have been various studies describing interviewer effects on different aspects of the survey process, including the construction of the sampling frame and sampling (e.g. Eckman, 2013; Eckman & Kreuter, 2011; Tourangeau et al., 2012), nonresponse (e.g., Blom et al., 2011; Durrant et al., 2013; Jäckle et al., 2013), and measurement (e.g. Durrant et al., 2010; Rice, 1929). However, very few studies have combined different data sources and analyzed interviewer effects on different aspects of a survey simultaneously (for notable exceptions see Brunton-Smith et al., 2012; Loosveldt & Beullens, 2014; Olson et al., 2016; West et al., 2013; West & Olson, 2010).

From a survey operational viewpoint, it is of great interest to study interviewer effects on multiple aspects of a survey, because if interviewers who are successful in contacting sample persons and gaining sample persons' cooperation are not good at interviewing persons using standardized survey interview techniques in face-to-face interviews (Groves & Lyberg, 2010), this will pose a problem for the selection of good interviewers. Since the tasks interviewers have to fulfill vary enormously between the different parts of a survey, it is likely that the skills needed to be successful in each of these parts vary as well. For example, to gain sample persons' cooperation, interviewers typically have to be flexible, tailor their behavior to sample persons' needs and maintain interaction (Groves & Couper, 1998). In contrast, during the contact phase and the standardized survey interview, interviewers have to adhere on to instructions and follow predefined rules (Brunton-Smith et al., 2012; Fowler & Mangione, 1990).

In the three previous chapters, I looked into interviewer effects on survey unit nonresponse and estimates of substantive survey variables individually. In this chapter, I focus on the relationship between these two error sources. More specifically, I aim to answer the following research questions:

- Is the amount of the identified interviewer effect different for the recruitment process and the survey interview?
- Is there a relationship between interviewers' success in recruiting sample persons and the means of substantive survey variables?
- Are there interviewer characteristics that explain interviewer variance for both survey unit nonresponse and estimates of substantive survey variables?

In the previous chapters, I described interviewer effects on survey unit nonresponse and on estimates of substantive survey variables, which I compare now: The amount of variance introduced due to interviewer clustering is very high for successfully making contact with sample persons, very low for successfully gaining sample persons' cooperation, and moderate for the estimates of substantive survey variables under research. In addition, I have not identified any interviewer characteristics that explain interviewer variance for successfully making contact or successfully gaining sample persons' cooperation. However, some of the interviewer characteristics can significantly explain interviewer variance in some estimates of substantive survey variables. Analyses conducted in the present chapter show that there is a significant relationship between success in contacting sample persons and the proficiency mean scores of the self-administered part of the survey interview as well as most of the estimates of substantive survey variables that were administered by the interviewers at the interviewer level. I discovered almost no significant relationship between interviewers' success in gaining sample persons' cooperation and the estimates of substantive variables from the survey interview by interviewer.

I start this chapter with an overview of the findings in the literature on interviewer effects on multiple aspects of a survey. Subsequently, I briefly describe the data and methods used to analyze interviewer effects in this chapter. Finally, I present and discuss the results following the three research questions I introduced above and finish the chapter with a summary of the results and overall conclusion.

7.1 Previous Research

Interviewer effects have been studied in many countries, surveys and with a focus on every possible aspect of the TSE framework (for an overview see West & Blom, 2017). However, there are only very few studies that have examined interviewer effects for multiple error sources. An important reason for this is that data on multiple error sources are seldom available in a single survey. In the next paragraphs, I will give an overview of the few studies that have presented results of joint analyses of multiple interviewer error sources.

In 2010, West and Olson published a study in which they decompose the total interviewer variance in a telephone survey into nonresponse error variance and measurement error variance. They find that the identified interviewer variance for the reported values for some survey variables may arise from nonresponse variance across interviewers and conclude that interviewer trainings should focus on reducing differences between interviewers instead of decreasing nonresponse rates. For some other estimates of survey variables, however, West and Olson (2010) find measurement variance to be the primary source of interviewer variance. In addition, West et al. (2013) separate out interviewer effects on nonresponse and on estimates of substantive survey variables by applying the same method as West and Olson (2010) for a face-to-face survey. Contrary to the results presented by West and Olson (2010), they find measurement error variance to be the primary source of the total interviewer variance for most of the estimates of the survey variables in their analyses.

In a study about the correlation of interviewer variance in a broad range of estimates of substantive survey variables and interviewers' success in obtaining contact and cooperation with sample persons, Brunton-Smith et al. (2012) identify higher interviewer variance for the group of interviewers that is least successful in making contact with sample persons and gaining their cooperation. Moreover, they find a U-shaped relationship between interviewer cooperation rates and interviewer variance in estimates of substantive survey variables. This implies that the best and worst interviewers with regard to gaining sample persons' cooperation show the highest

interviewer variance in the estimates of substantive survey variables. For contact, however, Brunton-Smith et al. (2012) report a negative relationship between interviewer contact rates and interviewer variance in estimates of substantive survey variables. This indicates that interviewers who are more successful in making contact with sample persons generate lower interviewer variance in estimates of substantive survey variables.

Furthermore, Loosveldt and Beullens (2014) investigate to what extent interviewer effects on nonresponse affect the amount of bias in the achieved sample. They find that interviewers differ in the response propensities for sample persons and thus differ in their risk of introducing nonresponse bias.

Recently, Olson et al. (2016) tried to explain the relationship between interviewer cooperation rates and interviewer variance in estimates of substantive survey variables by means of interviewer behaviors with regard to deviations from standardized interviewing techniques. They find that interviewers with high cooperation rates do not show any signs of being better interviewers with regard to following standardized interviewing techniques.

Overall, the literature so far finds no clear pattern regarding interviewer effects on multiple aspects of surveys. More specifically, some studies report that the differences between interviewers with regard to their ability to successfully make contact with sample persons or gain their cooperation are the major reason for interviewer variance identified in estimates of substantive survey variables, while other studies report measurement error variance to be the primary source of the total interviewer variance in some estimates of the substantive survey variables. However, the studies presenting these results differ in many ways, mainly due to varying data availability.

I contribute to the literature by conducting joint analyses of interviewer effects on survey unit nonresponse and on estimates of substantive survey variables. More specifically, I first compare the amount of the variance introduced due to interviewer clustering during the recruitment phase and the survey interview. Second, I study the relationship between

interviewers' success in recruiting sample persons and the means of substantive survey variables for their set of respondents. And third, I try to shed light on the question whether there are any interviewer characteristics that explain interviewer effects on survey unit nonresponse as well as on estimates of substantive survey variables.

7.2 Data Description and Method

The aim of the present chapter is to find out more about the relationship between interviewers' success during the recruitment process and their influence on estimates of substantive survey variables. I start with comparing the results of the previous chapters regarding interviewer effects on survey unit nonresponse (Chapter 4) and estimates of substantive survey variables (Chapter 6). Thus, the analyses of the present chapter are based on the same data sources as described in the overall data section (Chapter 3) as well in the previous Chapters 4 and 6. The data sources have been described in the respective chapters in detail and I will not repeat that here.

I analyze interviewer effects on survey unit nonresponse and on estimates of substantive survey variables by means of three-level logistic (Chapter 4) and three-level linear models (Chapter 6) and provide a detailed description of the multilevel modelling I used in the respective chapters. In addition to estimating the multilevel-models, I use post estimation tools for multilevel-modeling to estimate the Best Linear Unbiased Predictions (BLUPs) for interviewer effects on survey unit nonresponse and interviewer effects on predicted means of the substantive survey variables. More specifically, for both interview parts, I take the fixed proportion of the model as well as the random effect estimates at the interviewer level into account (Rabe-Hesketh & Skrondal, 2012, p. 111). After I estimate the predicted probabilities of the sample persons og being successfully contacted and participating in PIAAC and the predicted means of the survey variables for each respondent, I calculate the mean of these predicted values per interviewer which were, in the second step, correlated using Spearman's rho.

7.3 Results

7.3.1 Is the amount of the identified interviewer effect different for the recruitment process and the survey interview?

To answer my first research question of the present chapter, I compare intra-interviewer correlation coefficients – which inform about the proportion of the variance explained by interviewer clustering – from my analyses of interviewer effects on survey unit nonresponse and interviewer effects on the means of substantive survey variables (see Figure 7.1). In all models, I control for potential sample composition effects by including all available information on the sample composition. This means, I include sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate in the models on the dependent variables regarding survey unit nonresponse. In the models with the dependent variables coming from the survey interview, I include the following sample composition control variables: Respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region. During the recruitment process, the amount of variance located at the interviewer level of a sample persons' propensity to be successfully contacted is 22%. By contrast, the proportion of the variance due to interviewer clustering is 2% for successfully gaining sample persons' cooperation to participate in PIAAC (see Chapter 4). For the measures collected during the survey interview, the variance located at the interviewer level is between 0 and 7% (see Chapter 6).

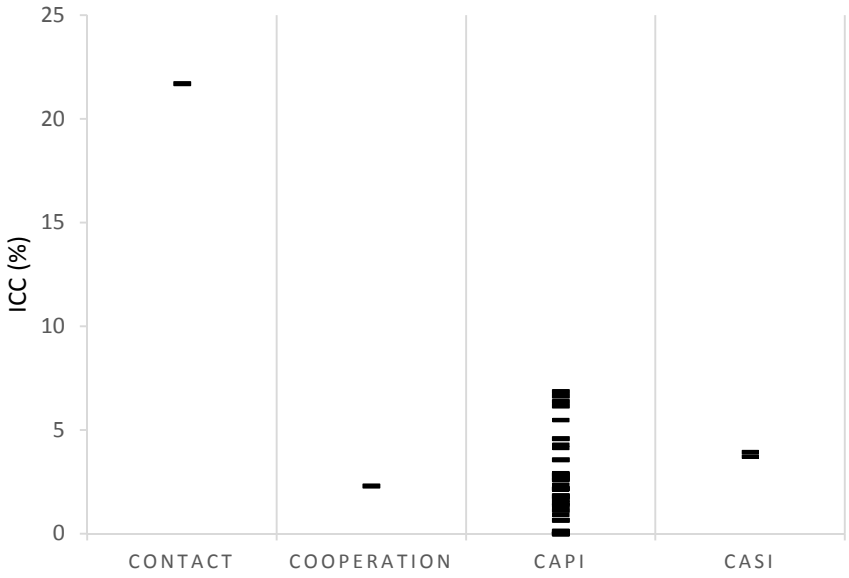


Figure 7.1 ICCs at the interviewer level for successful contact, successful cooperation, and the estimates of the CAPI and CASI variables, models 2, three-level logistic and three-level linear regression, PIAAC 2012

Notes. Contact: successful contact. Cooperation: Successful cooperation. CAPI: means of the following core questionnaire variables: H_Q01a, H_Q01b, H_Q01c, H_Q01d, H_Q01e, H_Q01f, H_Q01g, H_Q01h, H_Q02a, H_Q02b, H_Q02c, H_Q02d, H_Q03b, H_Q03c, H_Q03d, H_Q03f, H_Q03g, H_Q03h, H_Q05a, H_Q05c, H_Q05d, H_Q05e, H_Q05f, H_Q05g, H_Q05h, I_Q04b, I_Q04d, I_Q04h, I_Q04j, I_Q04l, I_Q04m, I_Q05f, I_Q06a, I_Q07a, I_Q07b, I_Q08. Question text for CAPI variables can be found in Appendix F. CASI: literacy and numeracy mean scores. Sample items for literacy and numeracy can be found in Appendix A. Contact and Cooperation: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI and CASI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Contact and cooperation: Number of interviewers = 115. Number of PSUs = 251. Contact: Number of sample persons = 7,902. Cooperation: Number of sample persons = 7,450. CAPI and CASI: Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132 (3,613 for H_Q05c). ICC = intra-class correlation coefficient. CAPI = computer-assisted personal interviewing. CASI = computer-assisted self-interviewing.

Comparing interviewer effects between the recruitment process and the survey interview process shows that for successfully making contact with sample persons, the interviewer effect is very high compared to the interviewer effects identified on the estimates of substantive survey variables. Instead, the interviewer effect for successfully gaining sample persons' cooperation is very low; however, for some estimates of the CAPI variables the variance located at the interviewer level is even lower. The difference between the interviewer effects on a sample persons' propensity to be successfully contacted and the means of the substantive survey variables is greater than the difference between the interviewer effects on a sample persons' propensity to participate in PIAAC and the means of the substantive survey variables.

I assume the low interviewer effect on a sample persons' propensity to cooperate to be a positive effect of the intense interviewer trainings conducted in PIAAC Germany. In the interviewer trainings, a special focus was set on achieving high response rates in PIAAC Germany to comply with international standards. In addition, successfully making contact with sample persons was also an important issue during the interviewer trainings, but still, the interviewer effect on a sample persons' propensity to be successfully contacted is relatively high. Looking at Figure 7.2, a combined graph of Figure 4.1 and Figure 4.2 presented in Chapter 4, it is obvious that the variation between interviewers' contact rates is very low compared to the variation of interviewers' cooperation rates. However, there are a few interviewers with rather low contact rates who might be the reason for the high interviewer effect on a sample persons' propensity to be contacted as they appear to be different compared to the rest of the interviewers who show a lower variability in their contact rates.

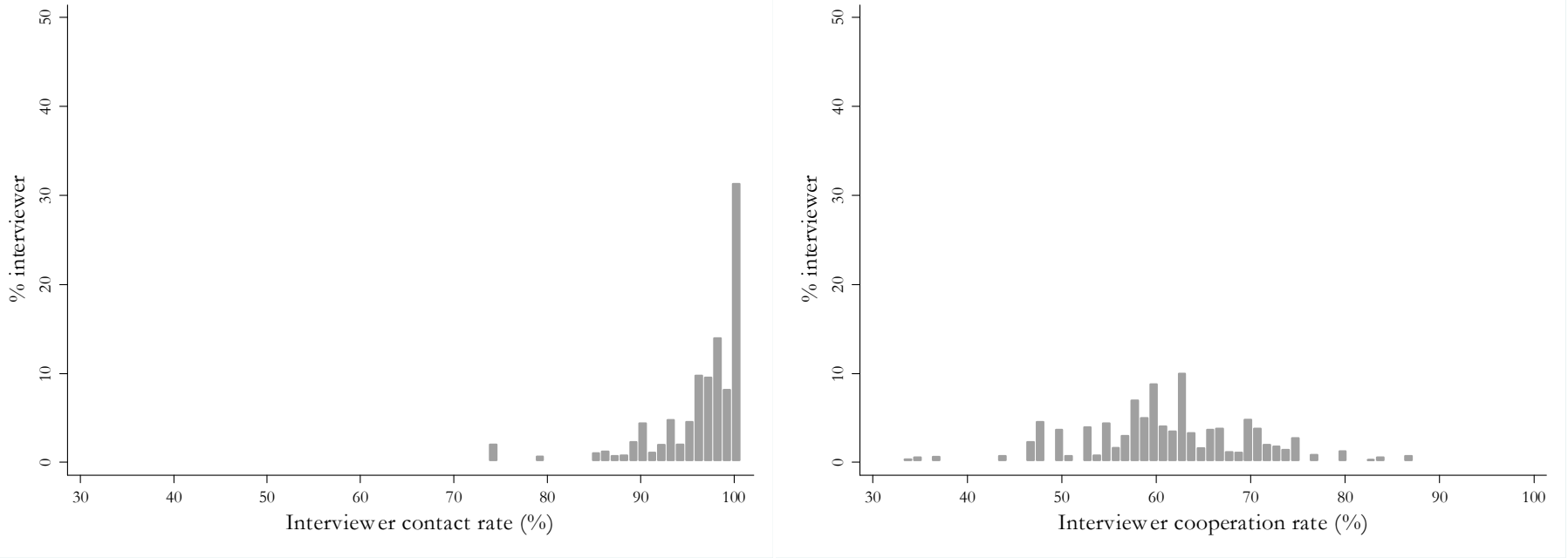


Figure 7.2 Contact and cooperation rates by interviewers, PIAAC 2012

Notes. Data based on 115 interviewers. Interviewer contact rate: percentage of cases for which interviewers successfully made contact by total number of sample persons assigned to an interviewer. Interviewer cooperation rate: percent of cases for which interviewers successfully gained sample persons' cooperation by total number of sample persons.

7.3.2 Is there a relationship between interviewers' success in recruiting sample persons and the means of substantive survey variables?

To answer the second research question, *Is there a relationship between interviewers' success in recruiting sample persons and the means of substantive survey variables?* I estimate correlations between sample persons' predicted probabilities of beeing successfully contacted and successfully cooperating and the predicted sample means of substantive survey variables at the interviewer level. More specifically, I focus on whether I find *significant correlations* between the sample persons' predicted probabilities of beeing successfully contacted and successfully cooperating and the predicted sample means of the substantive survey variables under research by interviewer or whether *no significant correlation* is obtained.

Since I conducted the analyses for a large number of substantive survey variables (38 variables) a detailed discussion of each correlation would be impossible. Instead, I interpret only some example results in the next paragraphs. More specifically, I present scatterplots for the relationship between the sample persons' predicted probabilities of beeing successfully contacted and successfully cooperating and the predicted sample mean scores of the two variables literacy and numeracy from the CASI part of the survey interview by interviewer. The scatterplots showing the relationship between the sample persons' predicted probabilities of beeing successfully contacted to successfully cooperating with the predicted sample means of the substantive survey variables from the interviewer-administered interview part at the interviewer level can be found in Appendix H. The results of all correlations, Spearman's rho, and the corresponding p-value at the 5%-level are presented in Table 7.1.

Table 7.1 Relationship between the predicted probabilities for successful contact and successful cooperation and the predicted means of CAPI and CASI variables by interviewer, Spearman's rho and p-values, three-level logistic and three-level linear regression, PIAAC 2012

	Survey variable	Successful contact		Successful cooperation	
		Spearman's rho	p-value	Spearman's rho	p-value
CAPI	H_Q01a	0.04	0.67	0.06	0.57
	H_Q01b	-0.24	0.01**	-0.20	0.04*
	H_Q01c	0.10	0.32	-0.08	0.43
	H_Q01d	-0.03	0.76	0.00	0.96
	H_Q01e	-0.43	0.00**	-0.05	0.60
	H_Q01f	-0.10	0.28	0.04	0.71
	H_Q01g	0.24	0.01**	0.08	0.42
	H_Q01h	-0.08	0.39	-0.02	0.80
	H_Q02a	-0.39	0.00**	-0.11	0.25
	H_Q02b	-0.24	0.01**	-0.13	0.18
	H_Q02c	-0.10	0.28	-0.13	0.18
	H_Q02d	0.06	0.55	0.07	0.46
	H_Q03b	0.08	0.39	-0.07	0.47
	H_Q03c	0.11	0.25	-0.12	0.23
	H_Q03d	0.02	0.80	-0.06	0.56
	H_Q03f	-0.12	0.20	-0.10	0.29
	H_Q03g	0.08	0.42	-0.01	0.88
	H_Q03h	-0.22	0.02*	-0.21	0.03*
	H_Q05a	-0.47	0.00**	-0.14	0.15
	H_Q05c	-0.32	0.00**	-0.07	0.49
	H_Q05d	-0.46	0.00**	-0.01	0.88
	H_Q05e	-0.28	0.00**	0.10	0.31
	H_Q05f	-0.39	0.00**	0.07	0.47
	H_Q05g	-0.22	0.02*	-0.22	0.02*
	H_Q05h	-0.27	0.01**	-0.20	0.04*
	I_Q04b	-0.18	0.07	-0.03	0.74
	I_Q04d	-0.24	0.01**	-0.18	0.07
	I_Q04h	-0.32	0.00**	-0.03	0.76
	I_Q04j	-0.11	0.24	-0.09	0.37
	I_Q04l	-0.13	0.18	-0.19	0.05
	I_Q04m	-0.25	0.01**	-0.14	0.15
	I_Q05f	0.48	0.00**	0.02	0.83
	I_Q06a	-0.26	0.01**	-0.04	0.65
	I_Q07a	-0.24	0.01**	-0.08	0.44
	I_Q07b	-0.21	0.03*	-0.06	0.54
	J_Q08	-0.27	0.01**	-0.12	0.23
CASI	Literacy	-0.20	0.04*	-0.13	0.17
	Numeracy	-0.19	0.05*	-0.12	0.24

Notes. Contact: successful contact. Cooperation: successful cooperation. Predicted means of CAPI variables range from 1 to 5. Question text for core questionnaire variables can be found in Appendix F. Predicted mean scores of literacy and numeracy range from 0 to 500. Sample items for literacy and numeracy can be found in Appendix A. Number of interviewers = 107. Number of PSUs = 240. Contact and Cooperation: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, unemployment rate). CAPI and CASI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Contact and cooperation: Contact: Number of sample persons = 7,902. Cooperation: Number of sample persons = 7,450. CAPI and CASI: Number of respondents = 4,132 (3,613 for H_Q05c). CAPI = computer-assisted personal interviewing. CASI = computer-assisted self-interviewing.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

The relationship between the sample persons' predicted probabilities of beeing successfully contacted and the predicted mean scores of the two substantive survey variables of the self-completion part of the survey interview at the interviewer level is very similar: I find a weak but significant correlation between the sample persons' predicted probabilities of beeing successfully contacted and the predicted literacy mean scores at the interviewer level (Spearman's' rho = -0.20, p = .04; see Figure 7.3 and Table 7.1) as well as between sample persons' predicted probabilities of beeing successfully contacted and the predicted numeracy mean scores at the interviewer level (Spearman's' rho = -0.19, p = .05; see Figure 7.3 and Table 7.1).

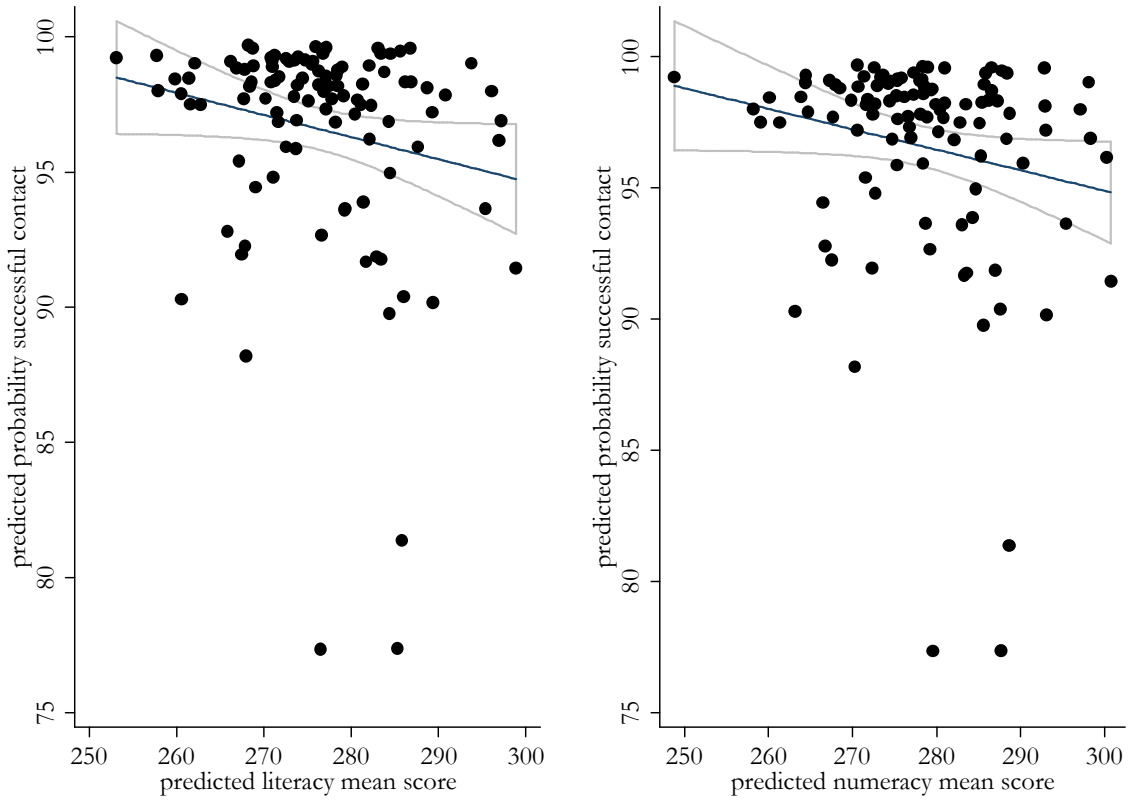


Figure 7.3 Relationship between the predicted probability for successful contact and the predicted mean score of the CASI variables literacy and numeracy by interviewer, PIAAC 2012

Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful contact and CASI variables literacy and numeracy. Significant = $p < 0.05$. Contact: Successful contact. Predicted mean scores of literacy and numeracy range from 0 to 500. Sample items for literacy and numeracy can be found in Appendix A. Number of interviewers = 107. Number of PSUs = 240. Contact: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). Literacy and numeracy: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Contact: Number of sample persons = 7,902. Literacy and numeracy: Number of respondents = 4,132.

The relationships between sample persons' predicted probabilities of beeing successfully contacted and the predicted means of the substantive survey variables of the interviewer-administered part of the survey interview per interviewer, are found to be very similar to the ones found for the self-administered interview part: I find significant correlations between the sample persons' predicted probabilities of beeing successfully contacted and the predicted means of 21 of 36 substantive CAPI survey variables at the interviewer level (see in Table 7.1 and Appendix H). However, for 15 of the 36 substantive CAPI survey variables, I find no relationship between the sample persons' predicted probabilities of beeing successfully contacted and the predicted means of the respective variable by interviewer (see Table 7.1).

In contrast to the results regarding successfully making contact with sample persons, the relationship between the sample persons' predicted probabilities to successfully cooperate and the predicted mean scores of the two substantive survey variables of the self-completion part of the survey interview by interviewer is *not* found to be significant (see Figure 7.4 and Table 7.1).

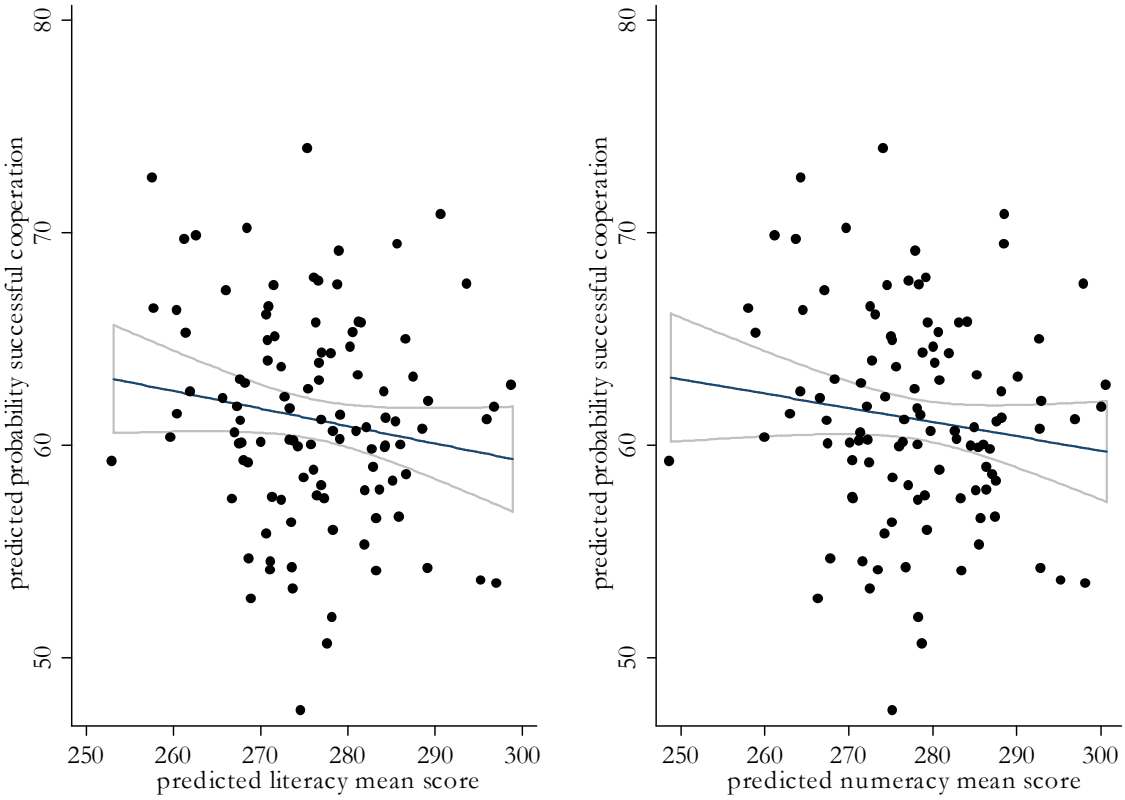


Figure 7.4 Relationship between the predicted probability for successful cooperation and the predicted proficiency mean score of the CASI variables literacy and numeracy by interviewer, PIAAC 2012

Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful cooperation and CASI variables literacy and numeracy. Significant = $p < 0.05$. Cooperation: Successful cooperation. Predicted mean scores of literacy and numeracy range from 0 to 500. Sample items for literacy and numeracy can be found in Appendix A. Number of interviewers = 107. Number of PSUs = 240. Cooperation: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). Literacy and numeracy: Controlled for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Cooperation: Number of sample persons = 7,450. Literacy and numeracy: Number of respondents = 4,132.

Overall, I find the relationship between the sample persons' predicted probabilities of beeing successfully contacted and the predicted means of the substantive survey variables at the interviewer level to be significant for 23 out of 38 correlations. For the correlations between the sample persons' predicted probabilities of successfully cooperating and the predicted means of the substantive survey variables by interviewers, I find four out of 38 correlations to be significant. There are two major conclusions of these findings with respect to my research question *Is there a relationship between interviewers' success in recruiting sample persons and the means of substantive survey variables?*: When calculating 38 correlations, one would expect to find 1.9 significant relationships by chance. Thus, my results *first* indicate that there is a significant

relationship between interviewers' success in making contact with sample persons and the estimates of the variables in the data they collect during the survey interview and, *second*, a weak significant relationship between interviewers' success at gaining sample persons' cooperation and the estimates of the variables in the data they collect during the interview.

However, an alternative explanation of the significant relationships between the sample persons' predicted probabilities of beeing successfully contacted and successfully cooperating and the predicted means of the substantive survey variables at the interviewer level could be a *selection effect* introduced during the recruitment of the sample persons by the interviewers. More specifically, interviewers who are more successful during the recruitment process might recruit a set of respondents that differ systematically in the estimates of the survey variables under research from the *mean* respondent (e.g. successful interviewers might recruit more sample persons that are typically hard to reach). Although I control for a large number of sample composition characteristics, this might not be efficient in reducing this possible selection effect.

7.3.3 Are there interviewer characteristics that explain interviewer variance for both survey unit nonresponse and estimates of substantive survey variables?

To answer the third research question in this chapter *Are there interviewer characteristics that explain interviewer variance for both survey unit nonresponse and estimates of substantive survey variables?*, I tested whether there are interviewer characteristics that significantly explain interviewer variance of the propensity of a sample person to be successfully contacted *and* interviewer variance in the estimates of substantive survey variables. More specifically, for each dependent variable, I compared the amount of interviewer variance between the model including the sample composition control variables and the model that includes an interviewer characteristic and the sample composition control variables by means of likelihood-ratio tests (Rabe-Hesketh & Skrondal, 2012).

The findings can be summarized as follows: Some of the interviewer characteristics significantly explain parts of the interviewer variance in the estimates of the survey variables (see

Table 7.2 for an overview and see Chapter 6 for a detailed description of the results). More specifically, the following interviewer characteristics explain interviewer variance in the estimates of the substantive survey variables under research: Interviewers' in-study and overall work experience, expectations with regard to the overall response rate, indicators on how to conduct standardized survey interviews, indicators on how to achieve response, trust, data protection concerns and three out of the four audio indicators (correctly collected formal criteria, number of incorrect skipped questions, number of incorrect read questions).

Table 7.2 Significant explained interviewer variance by interviewer characteristic for which a significant effect on successful contact and the estimates of the CAPI and CASI variables at the 1% and 5%-significance-level was found, three-level logistic and three-level-linear regression, PIAAC 2012

	Contact	CAPI									CASI	
		H_Q01a	H_Q01f	H_Q01h	H_Q02d	H_Q03b	H_Q03c	H_Q03g	H_Q05c	I_Q04l	Literacy	Numeracy
Experience												
In-study experience				x					x		x	x
Overall work experience				x				x		x		
Working hours												
Expectations											x	x
Expected response rate: overall											x	x
Reasons for working as an interviewer												
How to conduct standardized survey interviews									x			x
How to achieve response				x								
Trust										x		
Data protection concerns		x										
Audio indicators												
Collected formal criteria				x						x		
Number of incorrect skipped questions		x		x				x	x			
Number of incorrect read questions		x		x				x	x			

Notes. Contact: Successful contact. CAPI variables: Means. Question text for CAPI variables can be found in Appendix F. Literacy and numeracy: Mean scores. Sample items for literacy and numeracy can be found in Appendix A. No significant effects at the 1%- or 5%-significance level for the following interviewer characteristics: Age, gender, education, occupation status, expectations (response rate income variable), social desirability, audio indicator: Collecting the permission to record by respondent. No significant effects at the 1%- or 5%-significance level for the following dependent variables: Contact, H_Q02d, H_Q03b, H_Q03c. Contact: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, unemployment rate). CAPI and CASI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Only one interviewer characteristic included per model. Contact: Number of interviewers = 115. Number of PSUs = 251. Number of sample persons = 7,902. CAPI and CASI: Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132 (3,613 for H_Q05c).

In addition, my results show that none of the available interviewer characteristics significantly explain interviewer variance of a sample persons' propensity to be successfully contacted or take part in PIAAC Germany (see also Chapter 4). In consequence, none of the available interviewer characteristics significantly explain interviewer effects on both unit nonresponse and estimates of substantive survey variables.

I already discussed above and in Chapter 4 that there is little variation in the contact rates achieved by interviewers. In other words, interviewers seem to be very homogeneous in successfully making contact with sample persons (see Figure 7.2). Thus, it is not surprising that no interviewer characteristics have been identified to explain the interviewer effect on the propensity of a sample person to be successfully contacted. However, some outliers might be the reason for the high interviewer effect on contact. I assume that either the data needed to explain interviewer effects during the contact phase is not available or that specific circumstances could be the reason why some interviewers have lower contact rates than most of the other interviewers.

7.4 Conclusion

In this chapter, I studied the relationship between interviewer effects on survey unit nonresponse and estimates of substantive survey variables using data from the German implementation of PIAAC Germany. I started with a comparison of the variance introduced by interviewers during the recruitment process and the survey interview process. Afterwards, I tested whether I would find a significant relationship between the sample persons' predicted probabilities to be successfully contacted and successfully cooperate and the predicted means of a large number of substantive survey variables collected in the survey interview at the interviewer level. In the last section, I compared which interviewer characteristics explain interviewer variance for both, survey unit nonresponse and estimates of substantive survey variables. As I already discussed my findings in the results section, I will briefly repeat my major findings and

give some suggestions for countering interviewer effects within the next paragraphs. Overarching conclusions as well as implications for further research are presented in Chapter 8.

First, I find a large interviewer effect on a sample persons' propensity to be successfully contacted, almost no interviewer effect on a sample persons' propensity to participate in PIAAC, and a moderate interviewer effect on the estimates of substantive survey variables. I assume the relatively large interviewer effect during the contact phase to be caused by outliers: The variation between the interviewers regarding their contact rates is rather low, but some interviewers seem to be different as they have lower contact rates compared to most of the interviewers. Additionally, the low interviewer effect I find on sample persons' propensity to participate in PIAAC might be a positive result of the intense interviewer trainings. Thus, I suggest intensifying interviewer trainings regarding all aspects of a survey process to prepare interviewers for all survey parts in the best possible way. In practice, this is not always possible due to cost and timing constraints. However, regular re-trainings of the interviewers could prevent interviewer effects on all aspects of survey error (for an overview see Lessler et al., 2008).

Second, I show that – at the interviewer level – there is a significant relationship between the sample persons' predicted probability of being contacted and the predicted proficiency mean scores of the self-administered part of the survey interview as well as the predicted means of most of the substantive survey variables from the interviewer-administered interview part. Next to that, I find almost no significant relationship between sample persons' predicted probability of successful cooperation and the predicted means of the substantive survey variables by interviewer. These results lead either to the conclusion that there is a significant relationship between interviewers' success in making contact or cooperation with sample persons and the measures they collect during the survey interview, or this could be the result of a selection effect (more successful interviewers might recruit a set of respondents that differ systematically in the estimates of the survey variables under research from the "mean" respondent). In order to detect any inconsistencies in the early stages of the data collection process, the relationship of

interviewers' success during the recruitment process and estimates of key variables on the interviewer level should be closely monitored during field work for each survey. For example, if some interviewers differ significantly from other interviewers, they could receive re-trainings, with regard to how to successfully make contact and how to adhere to standardized interviewing techniques.

Third, I show that some of the interviewer characteristics explain parts of the interviewer variance identified in the estimates of substantive survey variables. Nevertheless, although my analyses are based on a very rich source of interviewer characteristics, none of the interviewer characteristics available can significantly explain any of the interviewer variance of a sample persons' propensity to be successfully contacted. As mentioned above, interviewers seem to be very homogeneous in successfully making contact with sample persons, which is one reason why I find no interviewer characteristics to explain the interviewer effect on the propensity of a sample person to be successfully contacted. Furthermore, as there are many studies that were not able to explain interviewer effects during the contact phase by means of available interviewer characteristics, there might be other interviewer characteristics that should be studied than the ones collected so far in studies on interviewer effects (see also West & Blom, 2017). In addition, I assume the explanatory power of interviewer characteristics to be study-specific. Thus, if no interviewer variance can be significantly explained by means of available interviewer characteristics, researchers should either focus on reporting interviewer effects or control for as many interviewer characteristics as possible in their analyses to account for possible interviewer effects. However, the consequence with regard to my research question is that I cannot determine whether there are interviewer characteristics that can explain parts of the interviewer variance during the recruitment process and for the estimates of substantive survey variables.

8 Conclusion

According to the TSE framework described by Groves et al. (2009b), errors can occur at each step of the survey life cycle. Interviewers have manifold tasks administering a survey and thus they are relevant in minimizing errors at each step they are involved in. In the present dissertation, I focus on two error sources: Survey unit nonresponse error and measurement error, both, of which can be caused by differences between interviewers. My motivation is to, *identify* interviewer effects and to *explain* interviewer effects for both error sources to shed more light on how interviewer effects could be countered to produce data of high quality. I base my analyses on data from PIAAC Germany as well as on rich data on interviewer characteristics. After a summary of my main findings, I make recommendations on how interviewer effects could be reduced prior to and during the data collection process and finish with some suggestions for further research.

I start with answering the question of whether I find interviewer effects on survey unit nonresponse and on estimates of substantive survey variables. Overall, the results of my dissertation show that there are interviewer effects on survey unit nonresponse as well as on estimates of substantive survey variables in PIAAC Germany. More specifically, In Chapter 4, I report that 20% of the variance is located at the interviewer level for successfully making contact with sample persons. Instead, the interviewer variance located at the interviewer level for successfully gaining sample persons' cooperation is very low (2%). In addition, in Chapter 5, I show that interviewers deviate from standardized interviewing techniques in all aspects I assess: Collect formal information, collect informed consent from respondents, skip and read questions correctly. As standardized interviewing techniques are used as a strategy to reduce interviewer effects during interviewer-administered interviews, and interviewers are found to deviate from these techniques, it is not surprising that I find interviewer effects on the estimates of a large number of substantive survey variables (Chapter 6). Nevertheless, compared to the interviewer

effects in the contact phase, they are moderate (ICCs: 0% - 7%). The PIAAC interview consists of two separate parts, the interviewer-administered survey interview (CAPI) and a self-administered cognitive test (CASI): I find the interviewer effects to be slightly higher for the interviewer-administered CAPI interview part compared to the self-administered CASI interview part. However, the difference is not very high (CAPI: ICC 5.3% on average, CASI: ICC 3.8% on average), which is in accordance to findings from the literature, where several studies found that self-administered interview parts reduce interviewer effects but do not eliminate them (Couper & Rowe, 1996; Kroh, 2004; O'Muircheartaigh & Campanelli, 1999; West & Peytcheva, 2014). I also find no difference in the amount of the interviewer effect between the types of the substantive variables. All variables under investigation were theoretically preselected to be prone to interviewer effects and my results show that interviewer effects are not limited to a special type of question or interview part.

After answering the questions of whether there are interviewer effects on unit nonresponse and on estimates of substantive survey variables, I intend to explain the identified interviewer effects by means of interviewer characteristics. I am in the fortunate position of being able to draw on a very rich source of interviewer characteristics to explain the interviewer effects I find on survey unit nonresponse and on the estimates of substantive survey variables. However, I find that there is no clear pattern of the explanatory power of the available interviewer characteristics for the interviewer effects on unit nonresponse and on the estimates of the substantive survey variables. More precisely, none of the available interviewer characteristics can explain interviewer variance on survey unit nonresponse (Chapter 4), and I find some interviewer characteristics which explain the interviewer effects on the estimates of substantive survey variables (Chapter 6): Interviewer experience significantly explains interviewer variance for both interview parts and interviewers' actual behavior with regard to how often they incorrectly skipped and read questions explains interviewer variance for the majority of the estimates of the variables from the CAPI part of the survey interview. Other interviewer characteristics explain interviewer variance

for the estimates of only one of the substantive survey variables under research, such as interviewers' expectation regarding their own response rate, indicators on how important different aspects on how to achieve response, working hours per week, indicators with regard to reasons for working as an interviewer and indicators regarding how to conduct standardized survey interviews, trust in people and data protection concerns. In addition, for interviewers' socio-demographic characteristics, as well as expectation with regard to the response rate of the income variable and social desirability, my results show that no interviewer variance is significantly explained at all for the estimates of any substantive survey variable. Furthermore, I find no association between interviewers' socio-demographic characteristics and their actual behavior regarding standardized interviewing techniques during the survey interviews (Chapter 5).

In Chapter 7, I conduct combined analyses of interviewer effects on the different aspects of the survey life cycle: I aim to answer the question of whether there is a relationship between interviewers' success in recruiting sample persons and the data they collect from their respondents. My results show, that there is a significant relationship between the sample persons' predicted probabilities of beeing successfully contacted and the predicted means of the majority of substantive survey variables under research at the interviewer level. This implies that interviewers differ systematically with regard to the means of the variables in the data they collect which then again might result in biased data. However, I find that there is no relationship between the predicted probabilities of successfully gaining sample persons' cooperation and the predicted means of the majority of substantive survey variables under research by interviewers.

There is one major limitation to using data from the German implementation of PIAAC to study interviewer effects: It is not possible to fully disentangle sample composition and interviewer effects, because the sample persons were not randomly assigned to interviewers. In my analyses of interviewer effects on survey unit nonresponse as well as on estimates of substantive survey variables, I show that a non-negligible amount of variance is located at the

PSU level. This PSU variance can be significantly reduced by including sample composition characteristics. Thus, in all models, I control for the non-random assignment of sample persons to interviewers by including all available sample composition characteristics in the models.

In order to reduce interviewer effects on the different aspects of error sources described in the TSE framework, there are several activities that could be put in place before and during data collection according to my results of this chapter. Within the next paragraphs, I discuss four major implications for the practice.

First, I conclude that it is worth reporting interviewer effects in interviewer-administered surveys for each step of the survey process where interviewers are involved. No clear pattern of the explanatory power of interviewer characteristics has been identified so far between different studies for the several steps of the survey process, which is in accordance with my key findings of my dissertation. Thus, I suggest that all available interviewer characteristics should be used to explain interviewer effects at each respective survey step on a *study-specific* basis. Subsequently, the interviewer characteristics that significantly reduce interviewer variance for the respective survey steps could be provided in data sets (e.g. scientific use files), to ensure that researchers can increase the quality of their substantive research by controlling for interviewer effects in their analyses.

The second overall conclusion of my findings is that I assume that the intense 5-day in-person interviewer trainings conducted with the German PIAAC interviewers might have reduced the differences between interviewers in some respects. More specifically: The very low interviewer effect on unit nonresponse, no significant relationship between the predicted probabilities of successfully gaining sample persons' cooperation and the predicted means of the majority of substantive survey variables at the interviewer level, and no effects of interviewers' socio-demographic characteristics could be a positive outcome of interviewer`s awareness that they affect the quality of the data. However, there are still some topics that could be emphasized more strongly during interviewer trainings to further reduce interviewer effects during the contact

phase as well as during the survey interview process, such as contact strategies, privacy issues or probing techniques. Furthermore, in Germany, interviewers typically undergo intensive interviewer trainings when they start working as an interviewer. Additional intensive study-specific trainings are only provided very seldom and interviewers typically receive short briefings or information material before they start working for a new survey. Nevertheless, my findings show that there is no clear pattern of the explanatory power of interviewer characteristics, such as socio-demographic characteristics, attitudes and behaviors. Thus, I assume that it makes sense to focus on study-specific topics during interviewer trainings to prepare interviewers in the best possible way for each survey individually.

The third inference I draw from the results in my dissertation is that it is worth monitoring interviewers' behavior during the interview. Interviewers' actual behavior regrading standardized interviewing techniques has been shown to significantly explain some of the interviewer variance in the estimates of the variables from the CAPI part of the interview. Thus, I recommend implementing quality control checks to detect deviations of interviewers' behavior with regard to standardized interviewing techniques on a regular basis. Only when deviations become obvious early during the field period, can individual and item-based re-trainings for interviewers be implemented to receive positive effects on data quality. However, more research is needed to show how such quality control activities could be implemented practically.

Forth, I find that detecting problems with interviewers' behavior is not only worthwhile during the interview process, but also during the recruitment phase. I find a significant relationship between predicted probabilities in the contact phase and predicted values of the survey data by interviewer, which could either be an actual relationship or the result of selection effects during the recruitment phase. I conclude that reducing differences between interviewers already during the recruitment process could have positive effects on the overall data quality. Thus, I recommend closely monitoring interviewers work with regard to all parts of the survey

process where interviewers are involved in order to detect any problems early during the data collection process and to intervene if necessary.

In my dissertation, I conduct numerous analyses for a large number of variables to provide further insights in explaining interviewer effects on survey unit nonresponse and estimates of substantive survey variables. Still, there are a lot of remaining open questions and new questions come up from the results of my analyses which reveal issues to be studied in further research.

One important area for further research is to explain differences in how often interviewers' deviate from standardized interviewing techniques. More data is needed to set up powerful models to study this research question. Next to interviewers` socio-demographic characteristics, their attitudes, expectations, personality, workload and many other characteristics could be tested in predefined survey experiments to learn more on how to prevent interviewers' deviations from standardized interviewing techniques and resulting interviewer effects on estimates of substantive survey variables.

Furthermore, to date, no systematic has been identified in the explanatory power of interviewer characteristics on different aspects of a survey life cycle described in the TSE framework. This is mainly the case, because surveys that try to explain interviewer effects by means of interviewer characteristics differ considerably. In future research, a well-planned comparative survey assessing the explanatory power of interviewer characteristics for interviewer effects across different studies might help to gain more insight here.

In addition, analyses of multiple error sources could be extended following the suggestions of Brunton-Smith et al. (2012), who analyzed whether the success in contacting sample persons and gaining sample persons' cooperation is correlated with interviewer variance in the estimates of survey variables. This approach could be used to develop strategies for a responsive design with regard to interviewer effects: By closely monitoring interviewer variance components for multiple error sources described in the TSE framework, inconsistencies could be detected early during data collection. More specifically, to date, interviewers success during the recruitment

process was of major research interest but could be expanded to include measures of data quality. Interviewers who are then found to be conspicuous, could receive individualized re-trainings to reduce interviewer effects and improve the overall data quality. However, experimental studies are needed to develop responsive interviewer monitoring strategies which are, on the one hand effective, but, on the other hand, feasible and easy to implement by survey practitioners.

References

Ackermann-Piek, D., & Massing, N. (2014). Interviewer behavior and interviewer characteristics in PIAAC Germany. *Methods, Data, Analyses, 8*(2), 199-222. doi:10.12758/mda.2014.008

Aquilino, W. S. (1993). Effects of spouse presence during the interview on survey responses concerning marriage. *Public Opinion Quarterly, 57*(3), 358-376.

Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use & Misuse, 35*(6-8), 845-867. doi:10.3109/10826080009148424

Beerten, R. (1999). The effect of interviewer and area characteristics on survey response rates: an exploratory analysis. *Survey Methodology Bulletin, 45*, 7-15.

Beullens, K., Loosveldt, G., & Dries, T. (2013). *Explaining interviewer effects on interviewer fieldwork performance indicators*. Paper presented at the 5th Conference of the European Survey Research Association, Ljubljana.

Biemer, P. P. (2010). Overview of design issues: Total survey error. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2 ed., pp. 27-57). Bingley, UK: Emerald Group Publishing Limited.

Biemer, P. P., & Lyberg, L. (2003). *Introduction to survey quality*. New York, NY: Wiley.

Billiet, J., & Loosveldt, G. (1988). Improvement of the quality of responses to factual survey question by interviewer training. *Public Opinion Quarterly, 52*(2), 190-211. doi:10.1086/269094

Blohm, M., Hox, J. J., & Koch, A. (2007). The influence of interviewers' contact behavior on the contact and cooperation rate in face-to-face houshold surveys. *International Journal of Public Opinion Research, 19*(1), 97-111. doi:10.1093/ijpor/edh120

Blom, A. G. (2009). Nonresponse Bias Adjustments: What can Process Data Contribute? *ISER Working Paper Series, 21*.

Blom, A. G., Ackermann, D., Korbmacher, J. M., Krieger, U., & Massing, N. (2013). *Comparing interviewer characteristics and explaining interviewer effects on nonresponse across three German face-to-face surveys*. Paper presented at the Workshop explaining interviewer effects in interviewer-mediated surveys, Mannheim.

Blom, A. G., de Leeuw, E. D., & Hox, J. J. (2011). Interviewer effects on nonresponse in the European Social Survey. *Journal of Official Statistics, 27*(2), 359-377.

Blom, A. G., & Korbmacher, J. M. (2013). Measuring interviewer characteristics pertinent to social surveys: A conceptual framework. *Survey Methods: Insights from the Field, 1*(1). doi:10.13094/SMIF-2013-00001

Brehm, J. O. (1993). *The phantom respondents: Opinion surveys and political representation.* Ann Arbor, MI: The University of Michigan Press.

Brenner, M. (1982). Response-effects of "role-restricted" characteristics of the interviewer. In W. Dijkstra & J. van der Zouwen (Eds.), *Response behavior in the survey-interview* (pp. 131-165). London, UK: Acad. Press.

Brüderl, J., Huyer-May, B., & Schmiedeberg, C. (2013). Interviewer Behavior and the Quality of Social Network Data. In P. Winker, N. Menold, & R. Porst (Eds.), *Interviewers' deviations in surveys impact, reasons, detection and prevetion* (pp. 147-160). Frankfurt am Main: Peter Lang.

Brunton-Smith, I., Sturgis, P., & Williams, J. (2012). Is success in obtaining contact and cooperation correlated with the magnitude of interviewer variance? *Public Opinion Quarterly, 76*(2), 265-286. doi:10.1093/poq/nfr067

Campanelli, P., & O'Muircheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey nonresponse. *Quality & Quantity, 33*(1), 59-76. doi:10.1023/A:1004357711258

Campanelli, P., Thompson, K., Moon, N., & Staples, T. (1997). The quality of occupational coding in the United Kingdom. In L. Lyberg (Ed.), *Survey measurement and process quality* (pp. 437-456). New York, NY: Wiley.

Cannell, C. F., Marquis, K. H., & Laurent, A. (1977). *A summary of studies of interviewing methodology*: DHEW.

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology, 12*, 389-437. doi:10.2307/270748.

Cleary, P. D., Mechanic, D., & Weiss, N. (1981). The effect of interviewer characteristics on responses to a mental health interview. *Journal of Health and Social Behavior, 22*(2), 189-193.

Collins, M. (1980). Interviewer variability: A review of the problem. *Journal of the Market Research Society, 22*(2), 77-95.

Conrad, F. G., Bromme, J. S., Benkí, J. R., Kreuter, F., Groves, R. M., Vannette, D., & McClain, C. (2013). Interviewer speech and the success of survey invitations. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*(Series A), 191-210.

Couper. (2013). Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods, 7*(3), 145-156.

Couper, & Rowe. (1996). Evaluation of a computer-assisted self-interview component in a computer-assisted personal interview survey. *Public Opinion Quarterly, 60*(1), 89-105. doi:Doi 10.1086/297740

Couper, M. P., & Groves, R. M. (1992). The role of the interviewer in survey participation. *Survey Methodology, 18*(2), 263-277.

de Heer, W. (1999). International response trends: Results of an international survey. *Journal of Official Statistics, 15*(2), 129-142.

de Leeuw, E. D. (1999). How do successful and less successful interviewers differ in tactics for combating survey nonresponse? *Bulletin de Méthodologie Sociologique, 62*(1), 29-62.

de Leeuw, E. D., Hox, J. J., Snijkers, G., & de Heer, W. (1998). Interviewer opinions, attitudes, and strategies regarding survey participation and their effect on response. In A. Koch & R. Porst (Eds.), *Nonresponse in survey research* (pp. 239-262). Mannheim: ZUMA Nachrichten SPEZIAL 4.

Durbin, J., & Stuart, A. (1951). Differences in response rates of experienced and inexperienced Interviewers. *Journal of the Royal Statistical Society, 114*(2), 163-206.

Durrant, G. B., D'Addio, J., & Steele, F. (2013). Analysing interviewer call record data by using a multilevel discrete-time event history modelling approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 176*(1), 251-269. doi:10.1111/j.1467-985X.2012.01073.x

Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly, 74*(1), 1-36. doi:10.1093/poq/nfp098

Eckman, S. (2013). Do different listers make the same housing unit frame? Variability in housing unit listing. *Journal of Official Statistics, 29*(2), 249-259. doi:10.2478/jos-2013-0021

Eckman, S., & Kreuter, F. (2011). Confirmation bias in housing unit listings. *Public Opinion Quarterly, 75*(1), 139-150. doi:10.1093/poq/nfq066

Ehrlich, J. S., & Riesman, D. (1961). Age and authority in the interview. *Public Opinion Quarterly, 25*(1), 39-56.

Federal Institute for Research on Building Urban Affairs and Spatial Development. (2015). *Indicators and maps on territorial and urban development*.

Feldman, J. J., Hyman, H., & Hart, C. W. (1951-52). A field study of interviewer effects on the quality of survey data. *Public Opinion Quarterly, 15*(4), 734-761.

Flores-Macias, F., & Lawson, C. (2008). Effects of interviewer gender on survey responses: Findings from a household survey in Mexico. *International Journal of Public Opinion Research, 20*(1), 100-110. doi:10.1093/ijpor/edn007

Fowler, F. J. J., & Mangione, T. W. (1990). *Standardized survey interviewing. Minimizing interviewer-related error* (Vol. 18). Newbury Park, CA: Sage Publications.

Freeman, J., & Butler, E. W. (1976). Some sources of interviewer variance in surveys. *The Public Opinion Quarterly, 40*(1), 79-91.

Freeman, J., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*(2), 195-202.

Gfroerer, J., Eyerman, J., & Chromy, J. (2002). *Redesigning an ongoing oational household survey: Methodological issues*. DHHS Publication No. SMA 03-3768. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Groves, R. M. (2004). *Survey error and survey costs*. Hoboken, NJ: John Wiley & Sons.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York; NY: Wiley.

Groves, R. M., Fowler, F. J. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (1992). Survey interviewing. In R. M. Groves, F. J. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, & R. Tourangeau (Eds.), *Survey Methodology* (pp. 269-301): John Wiley & Sons.

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

Groves, R. M., Fowler, F. J. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009b). *Survey Methodology* (Vol. 2). Hoboken, NJ: John Wiley & Sons.

Groves, R. M., & Lyberg, L. (2010). Total survey error past, present and future. *Public Opinion Quarterly, 74*(5), 849-879. doi:10.1093/poq/nfq065

Groves, R. M., & McGonagle, K. A. (2001). A theory-guided interviewer training protocol regarding survey participation. *Journal of Official Statistics, 17*(2), 249-265.

Groves, R. M., O'Hare, B. C., Gould-Smith, D., Benkí, J. R., & Maher, P. (2008). Telephone interviewer voice characteristics and the survey participation desion. In J. M. Lepkowski, C. Tucker, M. J. Brick, E. D. De Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 385-400). New York, NY: John Wiley & Sons.

Haan, M., Ongena, Y. P., & Huiskes, M. (2013). Interviewers' question: Rewording not always a bad thing. In P. Winker, N. Menold, & R. Porst (Eds.), *Interviewers' deviations in surveys impact, reasons, detection and prevetion* (pp. 173-194). Frankfurt am Main: Peter Lang.

Helmschrott, S., & Martin, S. (2014). Nonresponse in PIAAC Germany. *Methods, Data, Analyses, 8*(2), 243-266. doi:10.12758/mda.2014.009

Houtkoop-Steenstra, H. (2000). *Interaction and the standardized survey interview: The living questionnaire*. Cambridge: Cambridge University Press.

Hox, J. J. (2010). *Multilevel analysis - Techniques and applications* (2 ed.). New York, NY: Routledge. Taylor & Francis Group.

Hox, J. J., & de Leeuw, E. D. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 103-119). New York, NY: John Wiley & Sons.

Hox, J. J., de Leeuw, E. D., & Kreft, I. G. G. (1991). The effect of interviewer and respondent characteristics on the quality of survey data: A multilevel model. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 439-461). New York, NY: John Wiley & Sons.

Hyman, H. H., & Cobb, W., J. (1954). *Interviewing in social research*. Chicago, IL: University of Chicago Press.

Jäckle, A., Lynn, P., Sinibaldi, J., & Tipping, S. (2013). The effect of interviewer experience, attitudes, personality and skills on respondent co-operation with face-to-face surveys. *Survey Research Methods, 7*(1), 1-15.

Japec, L. (2005). *Quality issues in interview surveys*. Stockholm: Stockholm University.

Japec, L. (2008). Interviewer error and interviewer burden. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. De Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 187-211). Hoboken, NJ: John Wiley & Sons.

Kroh, M. (2004). Intervieweffekte bei der Erhebung des Körpergewichts: Die Qualität von umfragebasierten Gewichtsangaben *DIW Discussion Papers, No. 439*.

Landis, J. R., Sullivan, D., & Sheley, J. (1973). Feminist attitudes as related to sex of the interviewer. *The Pacific Sociological Review, 16*(3), 305-314. doi:10.2307/1388489

Lehtonen, R. (1996). Interviewer attitudes and unit nonresponse in two different interview schemes. In S. Laaksonen (Ed.), *In international perspectives on nonresponse: Proceedings of the sixth international workshop on household survey nonresponse* (pp. 130-140). Helsinki: Statistics Finland.

Lemay, M., & Durand, C. (2002). The effect of interviewer attitude on survey cooperation. *Bulletin de Méthodologie Sociologique, 76*, 27-44.

Lessler, J. T., Eyerman, J., & Wang, K. (2008). Interviewer training. In E. D. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology*. New York, NY: Taylor & Francis Group.

Lipps, O., & Pollien, A. (2011). Effects of interviewer experience on components of nonresponse in the European Social Survey. *Field Methods, 23*(2), 156-172. doi:10.1177/1525822x10387770

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of American Statistical Association, 83*(404).

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.

Lohr, S., L. (1999). *Sampling: Design and analyses*. Pacific Grove, CA: Duxbury Press.

Loosveldt, G. (2008). Face-to-face interviews. In E. D. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 201-220). New York, NY: Taylor & Francis Group.

Loosveldt, G., & Beullens, K. (2014). A Procedure to Assess Interviewer Effects on Nonresponse Bias. *SAGE Open, 4*(1), 1-12. doi:10.1177/2158244014526211

Lord, V. B., Friday, P. C., & Brennan, P., K. (2005). The effects of interviewer characteristics on arrestees' responses to drug-related questions. *Applied Pychology in Criminal Justice, 1*(1), 36-55.

Lui, M., & Stainback, K. (2013). Interviewer gender effects on survey responses to marriage-related questions. *Public Opinion Quarterly, 77*(2), 606-618. doi:10.1093/poq/nft019

Lyberg, L. (2012). Survey quality. *Survey Methodology, 38*(2), 107-130.

Mangione, T. W., Fowler, F. J. J., & Louis, T. A. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics, 8*(3), 293-307.

Massing, N., & Ackermann, D. (2013). *Interviewer characteristics & interviewer effects in PIAAC Germany*. Paper presented at the 5th Conference of the European Survey Research Association, Ljubljana.

Massing, N., Ackermann, D., Martin, S., Zabal, A., & Rammstedt, B. (2013). Controlling interviewers` work in PIAAC - the Programme for the International Assessment of Adult Competencies. In P. Winker, N. Menold, & R. Porst (Eds.), *Interviewers` deviations in surveys impact, reasons, detection and prevetion* (pp. 117-130). Frankfurt am Main: Peter Lang.

Matschinger, H., Bernert, S., & Angermayer, M. C. (2005). An analysis of interviewer effects on screening questions in a computer assisted personal mental health interview. *Journal of Official Statistics, 21*(4), 657-674.

Maynard, D. W., & Schaeffer, N. C. (2002). Standardization and its discontents. In R. M. Groves, G. Kalton, J. N. K. Rao, N. Schwarz, & C. Skinner (Eds.), *Standardization and tracit knowledge: Interaction and practice in the survey interview* (pp. 3-45). New York, NY: John Wiley & Sons.

Mazzoni, D., & Dannenberg, R. (2012). Audacity. Retrieved March 2014 from http://audacity.sourceforge.net/?lang=de.

Mneimneh, Z. N. (2012). *Interview privacy and social conformity effects on socially desirable reporting behavior: Importance of cultural, individual, question, design and implementation factors (Doctoral dissertation).* Ann Arbor, MI: University of Michigan-Ann Arbor Program in Survey Methdology.

Mneimneh, Z. N., Heeringa, S. G., Tourangeau, R., & Elliott, M. R. (2014). Bridging psychometrics and survey methodology: Can mixed rasch models identify socially desirable reporting behavior? *Journal of Survey Statistics and Methodology, 2*(3), 257-282. doi:10.1093/jssam/smu008

Moorman, P. G., Newman, B., Millikan, R. C., Tse, C.-K., & Sandler, D. P. (1999). Participation rates in a ase-control study: The impact of age, race, and race of interviewer. *Annals of Epidemiology, 9*, 188-195.

Morton-Williams, J. (1993). *Interviewer approaches.* Cambridge: Dartmouth Publishing Company.

Moum, T. (1998). Mode of administration and interviewer effects in self-reported symtoms of anxiety and depression. *Social Indicators Research, 45*(1-3), 279-318.

Nealon, J. (1983). The effects of male vs. female telefone interviewers. *U.S. Department of Agriculture.*

O'Muircheartaigh, C. (1976). Response errors in an attitudional sample survey. *Quality & Quantity, 10*, 97-115. doi:10.1007/BF00144162

O'Muircheartaigh, C., & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 161*(1), 63-77.

O'Muircheartaigh, C., & Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 162*(3), 437-446. doi:10.1111/1467-985x.00147

OECD. (2013a). PIAAC Literacy - sample items. Retrieved 06/27/2017 from http://www.oecd.org/skills/piaac/Literacy%20Sample%20Items.pdf

OECD. (2013b). PIAAC Numeracy - sample items. Retrieved 06/27/2017 from http://www.oecd.org/skills/piaac/Numeracy%20Sample%20Items.pdf.

OECD. (2014). PIAAC technical standards and guidelines. Retrieved 06/03/2014 from http://www.oecd.org/skills/piaac/documentation.htm.

OECD. (2016). Technical report of the Survey of Adult Skills: Reader's companion, second edition, OECD Publishing, Paris. Retrieved March 2014 from http://dx.doi.org/10.1787/9789264258075-en.

Oksenberg, L., & Cannell, C. F. (1988). Effects of interviewer vocal characteristics on nonresponse. In R. M. Groves, P. P. Biemer, L. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 257-269). New York, NY: John Wiley & Sons.

Oksenberg, L., Coleman, L., & Cannell, C. F. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly, 50*, 97-111.

Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly, 75*(1), 99-114. doi: 10.1093/poq/nfq067

Olson, K., Kirchner, A., & Smyth, J. (2016). Do interviewers with high cooperation rates behave differently? Interviewer cooperation rates and interview behaviors. *Survey Practice, 9*.

Ongena, & Dijkstra. (2006). Question-answer sequences in survey-interviews. *Quality & Quantity, 40*, 983-1011. doi:10.1007/s11135-005-5076-4

Ongena, Y. P. (2005). *Interviewer and respondent interaction in survey interviews*. Amsterdam: Vrije Universiteit.

Ongena, Y. P., & Dijkstra, W. (2006). Methods of behavior coding of survey interviews. *Journal of Official Statistics, 22*(3), 419-451.

Pannekoek, J. (1991). A mixed model for analysing measurement errors for dichotomous variables. In P. P. Biemer, R. M. Groves, L. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 517-530). Hoboken, NJ: John Wiley & Sons.

Pickery, J., & Loosveldt, G. (2002). A multilevel multinomial analysis of interviewer effects on various components of unit nonresponse. *Quality & Quantity, 36*, 427-437.

Pickery, J., Loosveldt, G., & Carton, A. (2001). The effects of interviewer and respondent characteristics on response behavior in panel surveys: A multilevel approach. *Sociological Methods & Research, 29*(4), 509-523. doi:10.1177/0049124101029004004

Porst, R. (2008). *Fragebogen: Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudional modeling using Stata* (3 ed.). Lakeway Drive, College Station, TX: Stata Press.

Rammstedt, B., Ackermann, D., Helmschrott, S., Klaukien, A., Maehler, B. D., Martin, S., . . . Zabal, A. (2013). *Grundlegende Kompetenzen Erwachsender im internationalen Vergleich: Ergebnisse von PIAAC 2012*. Münster: Waxmann.

Rammstedt, B., Martin, S., Zabal, A., Konradt, I., Maehler, B. D., Perry, A. g. K., . . . Helmschrott, S. (2016). *Programme for the International Assessment of Adult Competencies (PIAAC), Germany - Reduzierte Version*.

Rammstedt, B., Martin, S., Zabal, A., Massing, N., Ackermann-Piek, D., Helmschrott, S., . . . Maehler, B. D. (2014). *Programme for the International Assessment of Adult Competencies (PIAAC), Germany – Master Dataset. Unpublished data. Mannheim: GESIS - Leibniz Institute for the Social Science*.

Rice, S. (1929). Contagious bias in the interview: A methodological note. *American Journal of Sociology, 35*(3), 420-423.

Schaeffer, N. C. (1991). Conversation with a purpose - or conversation? Interaction in the standardized interview. In P. P. Biemer, R. M. Groves, L. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 369-391). Hoboken, NJ: John Wiley & Sons.

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

Schaeffer, N. C., Garbarski, D., Freese, J., & Maynard, D. W. (2013). An interactional model of the call for survey participation: actions and reactions in the survey recruitment call. *Public Opinion Quarterly, 77*(1), 323-351. doi:10.1093/poq/nft006

Schaeffer, N. C., & Maynard, D. W. (1996). From paradigm to prototype and back again. In N. Schwarz & S. Sudman (Eds.), *Answering questions* (pp. 65-88). San Francisco, CA: Jossey-Bass Publishers.

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61*(4), 576-602. doi:0033-362X/97/6104-0002$02.50

Schober, M. F., & Conrad, F. G. (2002). A collaborative view of standardized survey interviews. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouven (Eds.), *Standardization and tracit knowledge: interaction and practice in the survey interview* (pp. 67-94). New York, NY: John Wiley & Sons.

Singer, E., Frankel, M., & Glassman, M. B. (1983). The effect of interviewer characteristics and expactations on response. *Public Opinion Quarterly, 47*(1), 68-83.

Sinibaldi, J., Durrant, G. B., & Kreuter, F. (2013). Evaluating the measurement error of interviewer observed paradata. *Public Opinion Quarterly, 77*(S1), 173-193. doi:10.1093/poq/nfs062

Smit, J. H., Dijkstra, W., & Van der Zouven, J. (1997). Suggestive interviewer behaviour in surveys: an experimental study. *Journal of Official Statistics, 13*(1), 19-28.

Snijkers, G., Hox, J. J., & de Leeuw, E. D. (1999). Interviewers' tactics for fighting survey nonresponse. *Journal of Official Statistics, 15*(2), 185-198.

Stember, H., & Hyman, H. (1949-50). Interviewer effects in the classification of responses *Public Opinion Quarterly, 13*(4), 669-682.

Stokes, S. L., & Yeh, M.-Y. (1988). Searching for causes of interviewer effects in telephone surveys. In R. M. Groves (Ed.), *Telephone survey methodology* (pp. 357-373). Ney York, NY: John Wiley and Sons.

Sturgis, P., & Campanelli, P. (1998). The scope for reducing refusals in household surveys: An investigation based in transcripts of tape-recorded doorstep interactions. *Journal of the Market Research Society, 40*(2), 121-139.

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys*. Chicago, IL: Aldine Publishing Company.

Sykes, W., & Collins, M. (1992). Anatomy of the survey interview. *Journal of Official Statistics, 8*(3), 277-291.

Tourangeau, R., Couper, M. P., & Steiger, D. M. (2003). Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. *Computers in Human Behavior, 19*(1), 1-24.

Tourangeau, R., Kreuter, F., & Eckman, S. (2012). Motivated underreporting in screening interviews. *Public Opinion Quarterly, 76*(3), 453-469. doi:10.1093/poq/nfs033

Tourangeau, R., Rasinski, K., Jobe, J. B., Smith, T. W., & Pratt, W. F. (1997). Surces of error in a survey on sexual behavior. *Journal of Official Statistics, 13*(4), 341-365.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and questions context. *Public Opinion Quarterly, 60*, 275-304.

Turner, C. F., Forsyth, B. H., O'Reilly, J. M., Cooley, P. C., Smith, T. K., Rogers, S. M., & Miller, H. G. (1998). Automated self-interviewing and the survey measurement of sensitive behaviors. In M. P. Couper, R. Baker, J. Bethlehem, C. Clark, J. Martin, W. Nicholls II, & J. O'Reilly (Eds.), *Computer assisted survey information collection* (pp. 455-473). New York, NY: Wiley.

Vassallo, R., Durrant, G. B., Smith, P. W. F., & Goldstein, H. (2015). Interviewer effects on non-response propensity in longitudinal surveys: a multilevel modelling approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 178*(1), 83-99.

Wang, K., Kott, P., & Moore, A. (2013). *Assessing the relationship between interviewer effects and NSDUH data quality*. Retrieved from Durham, NC:

Webster, C. (1996). Hispanic and Anglo interviewer and respondent ethnicity and gender: The impact on survey response quality. *Journal of Marketing Research, 33*(1), 62-72.

Weisband, S., & Kiesler, S. (1996). *Self disclosure on computer forms: Meta-analysis and implications.* Paper presented at the SIGCHI conference on human factors in computing systems.

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J., B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology, 2*(178). doi:10.3389/fpsyg.2011.00178

West, B. T., & Blom, A. G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology, 5*(2), 175-211. doi:10.1093/jssam/smw024

West, B. T., Kreuter, F., & Ursula, J. (2013). "Interviewer" effects in face-to-face surveys: A function of sampling, measurement error, or nonresponse? *Journal of Official Statistics, 29*(2), 277-297. doi:10.2478/jos-2013-0023

West, B. T., & Olson, K. (2010). How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly, 74*(5), 1004–1026. doi:10.1093/poq/nfq061

West, B. T., & Peytcheva, E. (2014). Can interviewer behaviors during ACASI affect data quality? *Survey Practice, 7*(5).

Winkler, N., Kroh, M., & Spiess, M. (2006). Entwicklung einer deutschen Kurzskala zur zweidimensionalen Messung von sozialer Erwünschtheit. *Discussion Papers DIW Berlin (Deutsches Institut für Wirtschaftsforschung)*, 1-31.

Zabal, A., Martin, S., Massing, N., Ackermann, D., Helmschrott, S., Barkow, I., & Rammstedt, B. (2014). *PIAAC Germany 2012: Technical report*. Münster: Waxmann.

Appendix

A: PIAAC sample items

Figure A.1 PIAAC literacy sample items, OECD (2013a).

OECD Skills Surveys

OECD
BETTER POLICIES FOR BETTER LIVES

PIAAC **Literacy – sample items**

Literacy is the ability to understand and use information from written texts in a variety of contexts to achieve goals and develop knowledge and potential. This is a core requirement for developing higher-order skills and for positive economic and social outcomes. Previous studies have shown reading literacy to be closely linked to positive outcomes at work, to social participation, and to lifelong learning.

Literacy - Sample Items

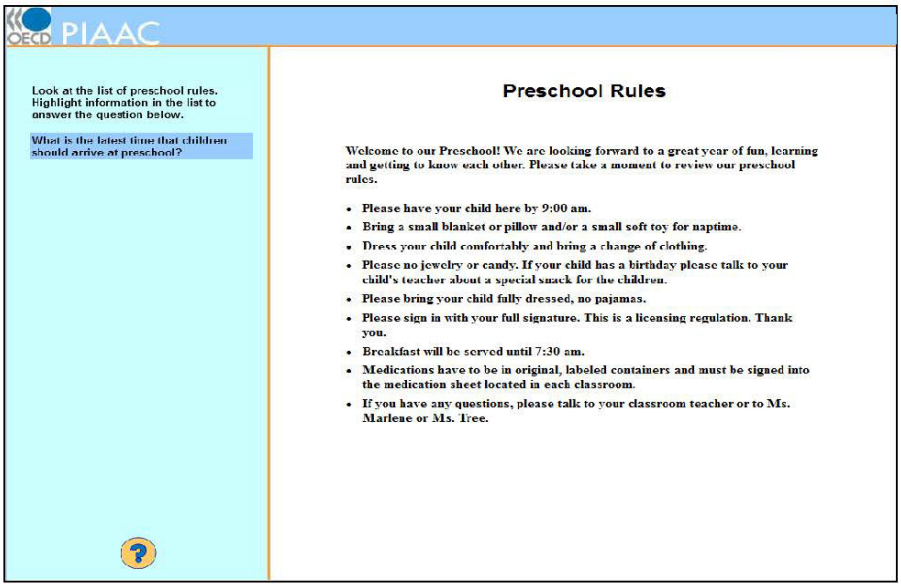
Two examples of the literacy items used in the Survey of Adult Skills (PIAAC) are presented below. Both use print-based stimuli. The sample problem-solving items presented separately give an idea of the type of "digital" stimulus material used.

The items are presented in the form delivered by the computer-based version of the assessment. To answer the questions, respondents highlight words and phrases or click on the appropriate location on the screen using a mouse.

Sample Item 1: Preschool rules

Preschool rules represents an item of average difficulty and focuses on the following aspects of the literacy construct:

Cognitive process	Access and identify
Context	Personal
Medium	Print



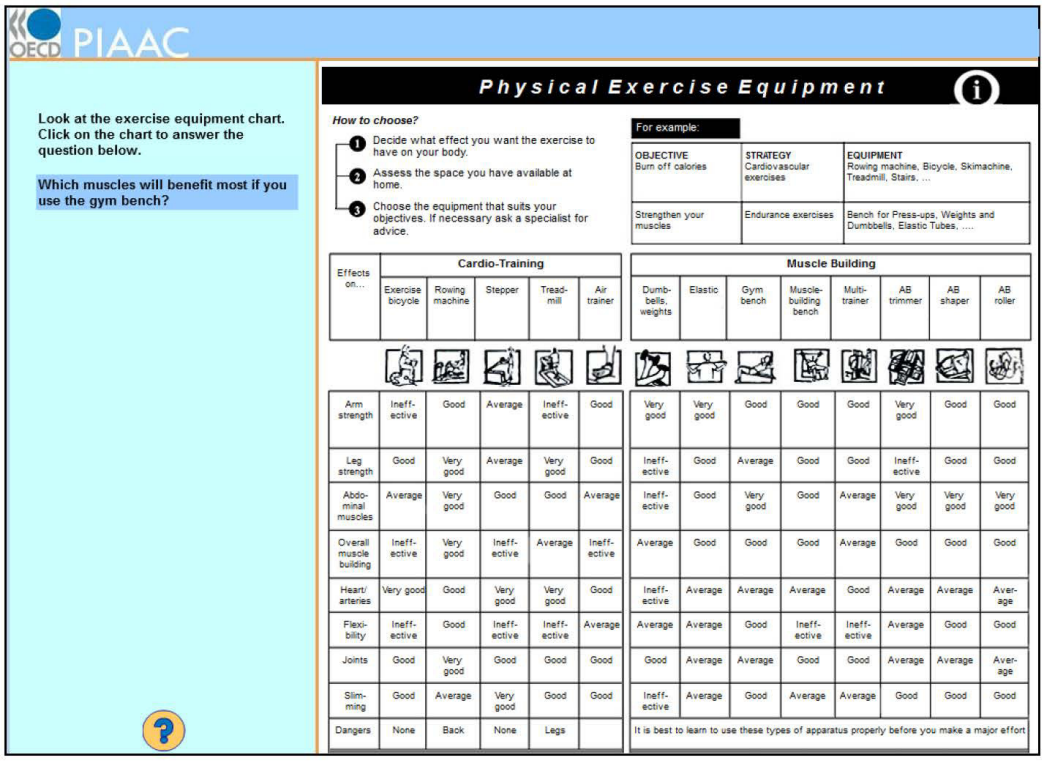
Sample Items 2 and 3: Physical Exercise Equipment

In many cases, several questions are associated with the same stimulus material. In the case of the stimulus relating to physical exercise equipment, there are two associated questions or test items.

The first item represents a relatively easy item and focuses on the following aspects of the literacy construct:

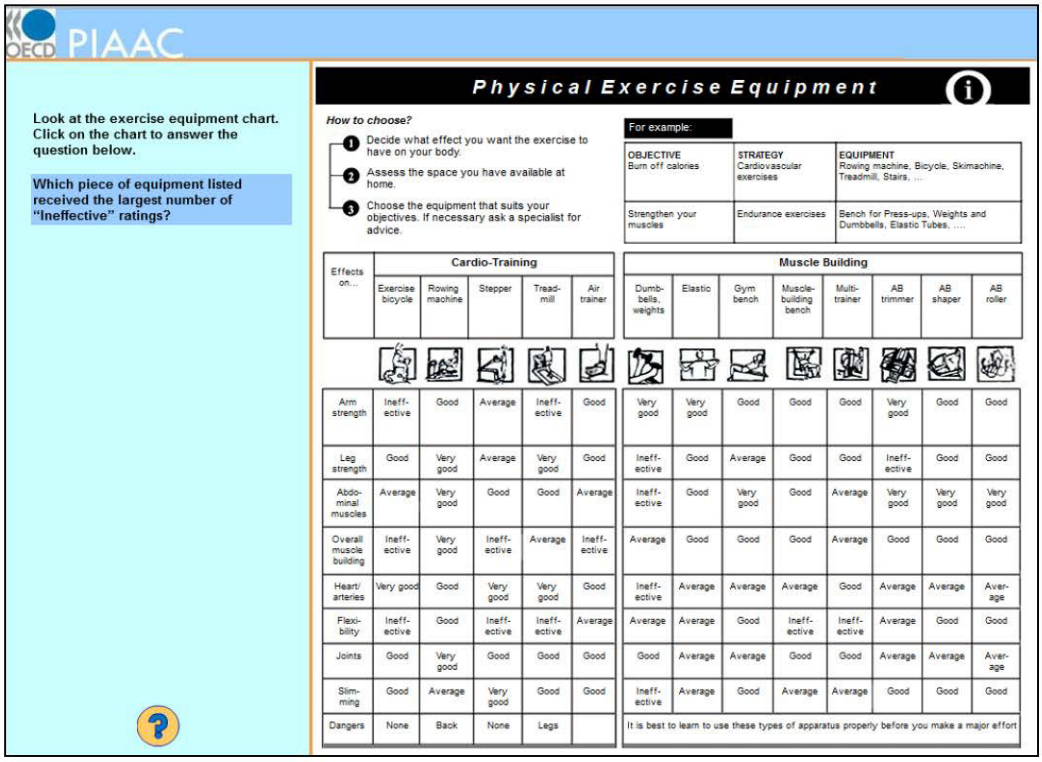
Cognitive process	Access and identify
Context	Personal
Medium	Print

Respondents answer the question by clicking on the cell in the chart that contains information about exercise equipment. Each of the cells and all of the images are "clickable" and multiple cells can be selected.



The second item represents a relatively easy item and focuses on the following aspects of the literacy construct:

Cognitive process	Integrate and interpret
Context	Personal
Medium	Print



OECD PIAAC

Look at the exercise equipment chart. Click on the chart to answer the question below.

Which piece of equipment listed received the largest number of "Ineffective" ratings?

Physical Exercise Equipment

How to choose?

1. Decide what effect you want the exercise to have on your body.
2. Assess the space you have available at home.
3. Choose the equipment that suits your objectives. If necessary ask a specialist for advice.

For example:

OBJECTIVE	STRATEGY	EQUIPMENT
Burn off calories	Cardiovascular exercises	Rowing machine, Bicycle, Skimachine, Treadmill, Stairs, ...
Strengthen your muscles	Endurance exercises	Bench for Press-ups, Weights and Dumbbells, Elastic Tubes, ...

Effects on...	Cardio-Training					Muscle Building							
	Exercise bicycle	Rowing machine	Stepper	Tread-mill	Air trainer	Dumb-bells, weights	Elastic	Gym bench	Muscle-building bench	Multi-trainer	AB trimmer	AB shaper	AB roller
Arm strength	Ineff-ective	Good	Average	Ineff-ective	Good	Very good	Very good	Good	Good	Good	Very good	Good	Good
Leg strength	Good	Very good	Average	Very good	Good	Ineff-ective	Good	Average	Good	Good	Ineff-ective	Good	Good
Abdo-minal muscles	Average	Very good	Good	Good	Average	Ineff-ective	Good	Very good	Good	Average	Very good	Very good	Very good
Overall muscle building	Ineff-ective	Very good	Ineff-ective	Average	Ineff-ective	Average	Good	Good	Good	Average	Good	Good	Good
Heart/ arteries	Very good	Good	Very good	Very good	Good	Ineff-ective	Average	Average	Average	Good	Average	Average	Aver-age
Flexi-bility	Ineff-ective	Good	Ineff-ective	Ineff-ective	Average	Average	Average	Good	Ineff-ective	Ineff-ective	Average	Good	Good
Joints	Good	Very good	Good	Good	Good	Good	Average	Average	Good	Good	Average	Average	Aver-age
Slim-ming	Good	Average	Very good	Good	Good	Ineff-ective	Average	Good	Average	Average	Good	Good	Good
Dangers	None	Back	None	Legs		It is best to learn to use these types of apparatus properly before you make a major effort							

3

Figure A.2 PIAAC numeracy sample items, OECD (2013b).

OECD Skills Surveys

OECD
BETTER POLICIES FOR BETTER LIVES

PIAAC Numeracy – sample items

In the Survey of adult skills (PIAAC). numeracy is defined as the ability to use, apply, interpret, and communicate mathematical information and ideas. It is an essential skill in an age when individuals encounter an increasing amount and wide range of quantitative and mathematical information in their daily lives. Numeracy is a skill parallel to reading literacy, and it is important to assess how these competencies interact, since they are distributed differently across subgroups of the population.

The items are presented in the form delivered by the computer-based version of the assessment. To answer the questions, respondents need to click in the appropriate box, and/or type figures in the space provided.

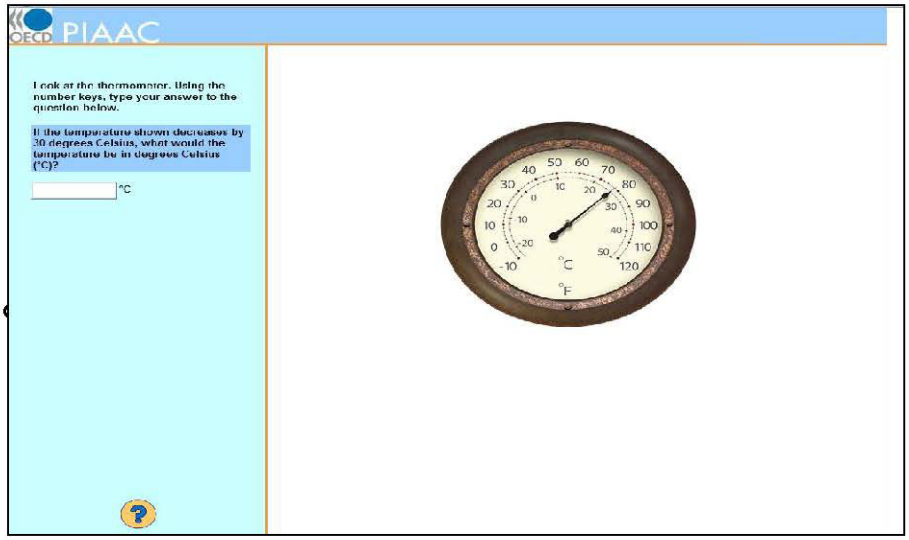
Numeracy - Sample Items

Sample Item 1: Thermometer

This item (of low difficulty) focuses on the following aspects of the numeracy construct:

Content	Dimension and shape
Process	Act upon, use (measure)
Context	Every day or work

Respondents are asked to type in a numerical response based on the graphic provided.

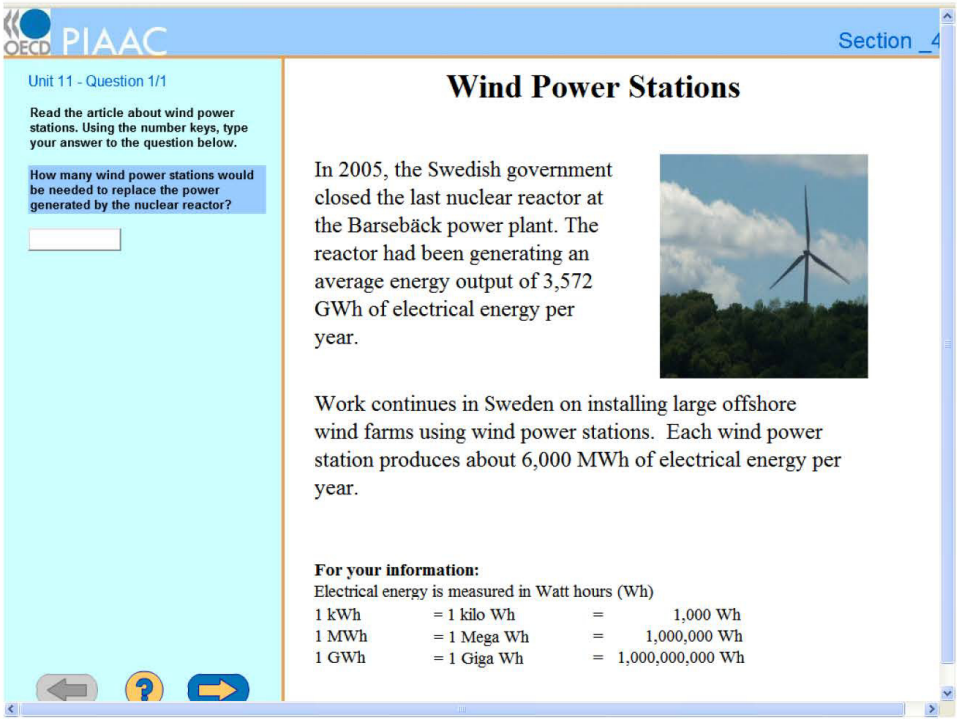


1

Sample Item 2: Wind power stations

This sample item (of medium difficulty) focuses on the following aspects of the numeracy construct:

Content	Quantity and Number
Process	Act upon, use (compute)
Context	Community and society

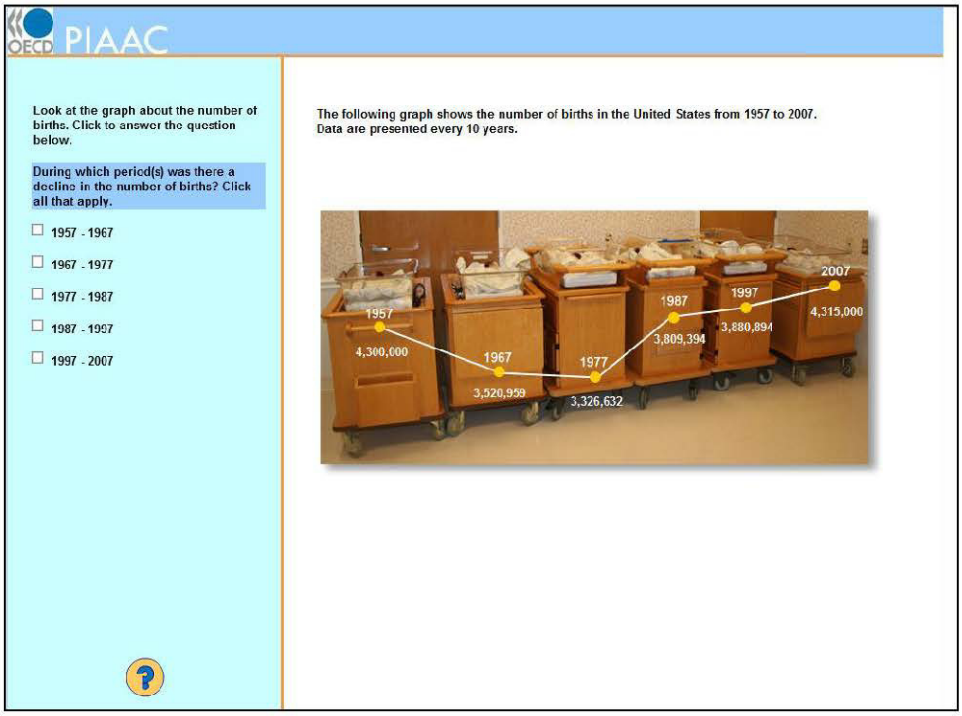


Sample Item 3: Births in the United States

This item (of medium difficulty) focuses on the following aspects of the numeracy construct:

Content	Data and chance
Process	Interpret, evaluate
Context	Community and society

Respondents are asked to respond by clicking on one or more of the time periods provided in the left pane on the screen.



B: Audio recordings: Results of interviewers' actual behavior with regard to standardized interviewing techniques

Table B.1 Interviewers' actual behavior: Collecting formal criteria

Collected formal criteria: date of interview, interviewer ID, respondent ID		
	n	**Percent**
Incorrect	51	47.66
Partly correct, partly incorrect	18	16.82
Correct	38	35.51
Total	107	100.00

Note. Table based on 107 interviewers and 217 recordings.

Table B.2 Interviewers' actual behavior: Collecting the permission to record the interview from respondent

Collected permission to record interview from respondent		
	n	**Percent**
Incorrect	72	67.29
Partly correct, partly incorrect	15	14.02
Correct	20	18.69
Total	107	100.00

Note. Table based on 107 interviewers and 217 recordings.

Table B.3 Interviewers' actual behavior: Number of incorrectly skipped questions

Number of incorrectly skipped questions		
	n	**Percent**
0	41	38.32
0.5	21	19.63
1	13	12.15
1.5	6	5.61
2	5	4.67
2.5	3	2.80
3	3	2.80
3.5	2	1.87
4	1	0.93
4.5	3	2.80
5	1	0.93
6	2	1.87
6.5	2	1.87
7	1	0.93
8.5	1	0.93
9.5	1	0.93
14.5	1	0.93
Total	107	100.00

Notes. Table based on 107 interviewers and 217 recordings. Odd numbers result from the fact that the number of incorrectly skipped questions can differ between the audio recorded interviews per interviewer.

Table B.4 Interviewers' actual behavior: Number of incorrectly read questions

Number of incorrectly read questions		
	n	**Percent**
0	7	6.54
0.5	10	9.35
1	6	5.61
1.5	7	6.54
2	9	8.41
2.5	1	0.93
3	2	1.87
3.5	8	7.48
4	2	1.87
4.5	3	2.80
5	3	2.80
5.5	3	2.80
6	2	1.87
6.5	4	3.74
7	2	1.87
8	1	0.93
8.5	1	0.93
9	4	3.74
9.5	1	0.93
10	1	0.93
10.5	1	0.93
11	4	3.74
11.5	1	0.93
12	2	1.87
12.5	3	2.80
13	1	0.93
15	1	0.93
15.5	1	0.93
16	1	0.93
17	2	1.87
18	1	0.93
18.5	1	0.93
19	1	0.93
20.5	2	1.87
21	1	0.93
21.5	1	0.93
22	1	0.93
26	1	0.93
29.5	1	0.93
32.5	1	0.93
40	1	0.93
47	1	0.93
Total	107	100.00

Notes. Table based on 107 interviewers and 217 recordings. Odd numbers result from the fact that the number of incorrectly read questions can differ between the audio recorded interviews per interviewer.

Table B.5 Interviewers' actual behavior: Statistics for the four indicators about interviewers' actual behavior

	Mean	SE	Min	Max
Collected date of interview, interviewer ID, respondent ID	0.44	0.45	0	1
Collected permission to record interview from respondent	0.26	0.40	0	1
Number of incorrectly skipped questions	1.43	2.35	0	15
Number of incorrectly read questions	7.57	8.64	0	47

Notes. Table based on 107 interviewers and 217 recordings. SE = standard error. Min = minimum. Max = maximum.

C: Interviewer survey: Original German version and English translation of the questionnaire

Topic	German	English
Introduction	Vielen Dank, dass Sie sich für unseren Fragebogen einige Minuten Zeit nehmen. Sie helfen uns damit sehr, die Befragungssituation besser zu verstehen. Ihre Antworten dienen **NICHT** einer Bewertung Ihrer Qualifikation oder Leistung. GESIS wird die Daten der Befragung anonymisieren und nur in anonymisierter und zusammengefasster Form an TNS Infratest weitergeben. **Damit lässt sich aus den Daten und Ergebnissen nicht erkennen, von welcher Person die Angaben gemacht worden sind.** Die Teilnahme sowie die Beantwortung einzelner Fragen ist natürlich freiwillig. Selbstverständlich werden alle Auflagen des Datenschutzes strengstens eingehalten. Weitere Informationen zum Datenschutz finden Sie im beiliegenden Datenschutzblatt. Bitte tragen Sie hier Ihre PIAAC-Interviewernummer ein! PIAAC-Interviewernummer: ☐ Bitte beziehen Sie die Fragen auf Ihre Tätigkeit als Interviewer/in insgesamt. Wenn es nur um Ihre Tätigkeit für PIAAC geht, ist dies aus dem Fragetext ersichtlich.	Thank you very much for taking a few minutes to complete our questionnaire. Your participation helps us immensely to better understand the interview situation. Your answers do **NOT** serve as an assessment of your performance. GESIS will anonymize the survey data and transmit it to TNS Infratest only in an anonymous and summarized form. **Thus, it is not possible to identify a person from the data and results.** The participation as well as the answering of individual questions is, of course, voluntary. Of course all data protection regulations are strictly adhered to. Further information on data protection can be found in the enclosed data protection sheet. Please fill in your PIAAC interviewer number! PIAAC interviewer number: ☐ *Please refer the questions to your overall job as an interviewer. If a question only relates to your work for PIAAC, this will be apparent from the question text.*
Job as an interviewer	**Fragetext** Wie lange arbeiten Sie insgesamt als Interviewer/in? **Antworten** • Weniger als 2 Jahre • 2 bis 5 Jahre • 6 bis 10 Jahre • 11 bis 15 Jahre • Mehr als 15 Jahre • Weiß nicht	**Question text** For how long in total have you been working as an interviewer? **Answers** • Less than 2 years • 2 to 5 years • 6 to 10 years • 11 to 15 years • More than 15 years • Don't know

Fragetext Wie viele Stunden pro Woche arbeiten Sie derzeit ungefähr als Interviewer/in?	**Question text** How many hours a week do you currently approximately work as an interviewer?
Hilfetext Bitte zählen Sie Ihre Stunden als Interviewer/in für TNS Infratest und mögliche andere Arbeitgeber zusammen.	**Help text** Please count your hours as an interviewer for TNS Infratest and, if possible, other employers.
Antwortfeld ____ Stunden	**Answer field** ____ hours
• Weiß nicht	• don't know
Fragetext Es gibt unterschiedliche Beweggründe als Interviewer/in zu arbeiten: Wie wichtig sind Ihnen die folgenden Punkte?	**Question text** There are different reasons for working as an interviewer. How important are the following aspects to you?
Hilfetext Antworten Sie bitte in jeder Zeile anhand der folgenden Skala. Der Wert 1 bedeutet: überhaupt nicht wichtig, der Wert 7 bedeutet: sehr wichtig. Mit den Werten zwischen 1 und 7 können Sie Ihre Meinung abstufen.	**Help text** Please provide an answer in each row using the following scale. Value 1 means: not important at all, value 7 means: very important. With the values between 1 and 7 you can grade your opinion.
Items • Bezahlung • Interessante Tätigkeit • Gelegenheit, unter die Leute zu kommen • Einblicke in die sozialen Lebensumstände anderer Menschen bekommen • Mitwirkung an wissenschaftlicher Forschung • Mitwirkung an Forschung, die der Gesellschaft dient • Möglichkeit der freien Zeiteinteilung	**Items** • Payment • Interesting work • Opportunity to interact with people • Gaining insight into other people's social circumstances • Involvement in scientific research • Involvement in research that serves society • Possibility to determine own working hours
Antwortskala • 1 Überhaupt nicht wichtig • 2 • 3	**Answer scale** • 1 not important at all • 2 • 3

• 4 • 5 • 6 • 7 Sehr wichtig • Weiß nicht	• 4 • 5 • 6 • 7 very important • don’t know
Fragetext Es folgen nun einige Fragen zu schwierigen Zielpersonen und Kontaktierungsversuchen. Wir möchten gerne von Ihnen wissen, wie Sie in den folgenden Situationen reagieren. Die Aussage trifft … auf mich zu.	**Question text** Below follows a series of statements about difficult respondents and contact attempts. We would like to know from you, how you react in the following situations. The statement applies to me ……
Hilfetext 1. Bitte beziehen Sie Fragen auf Ihre Tätigkeit als Interviewer/in insgesamt d.h. NICHT nur auf PIAAC. 2. Antworten Sie bitte in jeder Zeile anhand der folgenden Skala!	**Help text** 1. Please relate the questions to your activity as an interviewer in general, that is, NOT only to PIAAC. 2. Please provide an answer in each row using the following scale!
Items • Wenn der/die Befragte eine Frage nicht versteht, erkläre ich, was eigentlich mit der Frage gemeint ist. • Wenn der/die Befragte Schwierigkeiten mit einer Frage hat, helfe ich nicht, sondern lese den genauen Wortlaut der Frage noch mal vor. • Wenn ich merke, dass der/die Befragte Schwierigkeiten hat, mir zuzuhören, kürze ich lange Fragetexte ab. • Wenn ich merke, dass der/die Befragte Schwierigkeiten hat, die Frage zu verstehen, spreche ich langsamer. • Wenn ich merke, dass der/die Befragte es eilig hat, spreche ich schneller. • Wenn ich vom bisherigen Interviewverlauf her weiß, wie eine Antwort lauten wird, ergänze ich die Antwort. • Wenn ich merke, dass der/die Befragte kein Hochdeutsch spricht, spreche ich auch im regionalen Dialekt. • Ich halte mich immer genau an die Intervieweranweisungen, auch wenn ich sie nicht für sinnvoll erachte. • Wenn der/die Befragte mich beim Vorlesen einer Frage unterbricht,	**Items** • If the respondent doesn't understand a question, I explain what is actually meant with the question. • If the respondent has difficulties with a question, I don't help, but read out the exact wording again. • If I notice that the respondent has difficulties listening to me, I shorten long question texts. • If I notice that the respondent has difficulties understanding the question, I speak more slowly. • If I notice that the respondent is in a hurry, I speak faster. • If I know from the course of the interview what an answer will be, I complete the answer myself. • If I notice that the respondent doesn't speak High German, I also speak regional dialect. • I always exactly stick to the interviewer instructions, even if I don’t consider them sensible. • If the respondent interrupts me when reading a question to give an answer, I always read the question to the end anyhow.

um eine Antwort zu geben, lese ich die Frage trotzdem immer bis zu Ende vor. • Bei Studien, bei denen die Befragten selbstständig Fragen ausfüllen oder Aufgaben bearbeiten sollen, helfe ich auch dann nicht, wenn mich der/die Befragte um Hilfe bittet.	• In the case of studies in which respondents are asked to fill out questions independently or to process tasks, I do not help the respondent although they might ask me for help.
Antworten • voll und ganz • eher • eher nicht • über-haupt nicht • Weiß nicht	**Answers** • perfectly • somewhat • not really • not at all • don't know
Fragetext Zielpersonen reagieren oft ganz unterschiedlich auf die Bitte, an einer Studie teilzunehmen: Manche sind spontan dazu bereit, andere sind sich unsicher oder verweigern sofort. Bitte nennen Sie uns zu den folgenden Aussagen Ihre Meinung als erfahrene/r Interviewer/in. Ich stimme der Aussage…	**Question text** Sample persons have different reactions to the request to participate in a study: Some agree spontaneously, others hesitate or refuse immediately. In the following statements, please tell us your opinion as an experienced interviewer. I agree with the statement ...
Hilfetext Antworten Sie bitte in jeder Zeile anhand der folgenden Skala!	**Help text** Please provide an answer in each row using the following scale!
Items • Schwer motivierbare Zielpersonen sollten immer zur Teilnahme überredet werden. • Mit genug Aufwand kann sogar die am schwersten motivierbare Zielperson zur Teilnahme überredet werden. • Ein/e Interviewer/in sollte die Privatsphäre der Zielpersonen respektieren. • Wenn eine Zielperson der Befragung ablehnend gegenübersteht, sollte eine Verweigerung akzeptiert werden. • Man sollte immer die Freiwilligkeit der Befragung hervorheben. • Es ist nicht sinnvoll, schwer motivierbare Zielpersonen wiederholt zu kontaktieren. • Wenn man sie zur rechten Zeit erwischt, werden die meisten Leute	**Items** • Reluctant respondents should always be persuaded to participate. • With enough effort, even the most reluctant respondent can be persuaded to participate. • An interviewer should respect the privacy of the respondent. • If a respondent is reluctant, a refusal should be accepted. • One should always emphasize the voluntary nature of participation. • It does not make sense to contact reluctant target persons repeatedly. • If you catch them at the right time, most people will agree to participate. • Respondents that were persuaded after great effort do not provide reliable answers. • If one is convinced of the survey, even hard-to-motivate sample persons can be gained to cooperate.

	teilnehmen. • Befragte, die nur mit großem Aufwand zur Teilnahme überredet wurden, liefern keine zuverlässigen Antworten. • Wenn man von der Befragung überzeugt ist, können sogar schwer motivierbare Zielpersonen leicht zur Teilnahme gewonnen werden. **Antworten** • voll und ganz zu • eher zu • eher nicht zu • über-haupt nicht zu • Weiß nicht	 **Answers** • strongly • somewhat • not really • not at all • don't know
General attitudes and behaviour	**Fragetext** Wir möchten gerne von Ihnen wissen, inwieweit Sie den folgenden Aussagen zustimmen oder nicht zustimmen. **Hilfetext** Antworten Sie bitte in jeder Zeile anhand der folgenden Skala! **Items** • Es gibt nur wenige Menschen, denen man voll vertrauen kann. • Wenn man nicht aufpasst, wird man von anderen ausgenutzt.	**Question text** We would like to know from you, to what extent you agree or disagree with the following statements. **Help text** Please provide an answer in each row using the following scale! **Items** • There are few people whom you can fully trust. • If you are not careful, you will be taken advantage of by others.
Data protection	**Fragetext** Wie besorgt sind Sie um den Schutz Ihrer persönlichen Daten? **Hilfetext** Bitte nur eine Antwort ankreuzen! **Antworten** • Sehr besorgt • Ziemlich besorgt • Ein wenig besorgt • Gar nicht besorgt • Weiß nicht	**Question Text** How concerned are you about the safety of your personal data? **Help text** Please tick one answer only! **Answers** • Very concerned • Quite concerned • A little concerned • Not concerned at all • Don't know

Expectations about the PIAAC Main Study	**Fragetext** Haben Sie bereits in der Feldstudie als Interviewer/in für PIAAC gearbeitet? **Antworten** • Ja • Nein	**Question Text** Have you worked as an interviewer for the PIAAC Field Test? **Answers** • Yes • No
	Fragetext Bei PIAAC erhalten die Zielpersonen vorab und unabhängig von ihrer tatsächlichen Teilnahme einen kleinen Block als Geschenk. Ihnen wird zudem eine Entschädigung für den Zeitaufwand in Höhe von 50,- € in bar im Anschreiben angekündigt. Was erwarten Sie, wie viel Prozent Ihrer Zielpersonen werden insgesamt einer Durchführung des Interviews zustimmen? **Antwortfeld** ____ % Erwartete Teilnahmebereitschaft • Weiß nicht	**Question Text** In PIAAC, the sample persons receive a small block as a gift in advance, irrespective of their actual participation. In addition, they are notified of a compensation for the amount for time spent on PIAAC of 50, - € in cash in the advance letter. What do you expect, which percentage of your sample persons will agree to the interview. **Answer field** ____ % Expected response rate in percent • don't know
	Fragetext In sozialwissenschaftlichen Studien wird sehr oft auch nach dem Einkommen der Befragten gefragt. Was denken Sie, wie viele Ihrer Befragten (in Prozent) werden in der PIAAC-Studie Auskunft über ihr Einkommen geben? **Antwortfeld** ____ % Erwartete Auskunftsbereitschaft • Weiß nicht	**Question Text** Social surveys very often ask about respondents' income. How many of your respondents (in percent) in PIAAC do you expect will provide information about their income? **Answer field** ____ % Expected willingness to provide information • don't know
Personal details	**Fragetext** Sind Sie männlich oder weiblich? **Antworten** • Männlich	**Question Text** Are you male or female? **Answers** • male

• Weiblich	• female
Fragetext Wie alt sind Sie? **Antworten** • Bis einschließlich 40 Jahre • 41 bis 45 Jahre • 46 bis 50 Jahre • 51 bis 55 Jahre • 56 bis 60 Jahre • 61 bis 65 Jahre • 66 Jahre und älter	**Question text** How old are you? **Answers** • Up to and including 40 years • 41 to 45 years • 46 to 50 years • 51 to 55 years • 56 to 60 years • 61 to 65 years • 66 years and older
Fragetext Üben Sie derzeit neben Ihrer Tätigkeit als Interviewer/in noch eine andere Erwerbstätigkeit aus? Sind Sie außerdem… **Hilfetext** Bitte **alles** Zutreffende ankreuzen! **Antworten** • Vollzeit erwerbstätig • Teilzeit erwerbstätig • Geringfügig oder unregelmäßig erwerbstätig • Arbeitslos • Student/in • In Berufsausbildung oder Umschulung • In Rente / Pension • In Altersteilzeit mit Arbeitszeit Null • Erwerbsunfähig • In Elternzeit / Erziehungszeit • Hausfrau / Hausmann • Sonstiges • Nichts davon	**Question text** Apart from your job as an interviewer do you have any other job? Are you... **Help text** Please tick all that apply! **Answers** • full-time employed • part-time employed • marginally or irregularly employed • unemployed • a student • in vocational training or occupational re-training • retired • in retirement with no working time • disabled • on parental leave • a homemaker • other • none of these

Fragetext
Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?

Hilfetext
Bitte nur den **höchsten** Schulabschluss ankreuzen!

Antworten
• Hauptschulabschluss (Volksschulabschluss)
• Realschulabschluss (Mittlere Reife)
• Abgang von der Polytechnischen Oberschule nach der 8. Klasse (nach 1965)
• Abgang von der Polytechnischen Oberschule nach der 10. Klasse (vor 1965: 8. Klasse)
• Fachhochschulreife, Abschluss Fachoberschule
• Allgemeine oder fachgebundene Hochschulreife / Abitur (Gymnasium bzw. EOS, EOS mit Lehre)
• Einen anderen Schulabschluss, und zwar:_____

Question text
Which is your highest level of education?

Help text
Please tick your highest level of education only!

Answers
• Secondary school leaving certificate
(lowest formal qualification of Germany's tripartite secondary school system, after 8 or 9 years of schooling)
• Secondary school certificate
(intermediary secondary qualification, after 10 years of schooling)
• Leaving from polytechnic secondary school after the 8th grade (after 1965)
(lowest formal qualification of Germany's tripartite secondary school system, after 8 or 9 years of schooling)
• Leaving from polytechnic secondary school after the 10th grade (before 1965: 8th grade)
(intermediary secondary qualification, after 10 years of schooling)
• Advanced technical college certificate
(certificate fulfilling entrance requirements to study at a university of applied science)
• General or subject-specific university entrance diploma / university-entrance diploma
(higher qualification, entitling holders to study at a university)
• Other school leaving certificate, namely:_____

D: Interviewer survey: Results from the interviewer survey

Table D.1 Interviewer survey: Factor matrix for items about reasons for working as an interviewer, including eigenvalues, variances and Cronbach´s Alphas

	Factor1	Factor2	Factor3	Factor4
Payment	-0.08	-0.06	**0.46**	-0.10
Interest	0.39	**0.44**	0.15	-0.05
Socialize	-0.01	**0.54**	0.09	0.16
Insight	-0.00	**0.57**	-0.14	0.00
Research	**0.91**	-0.03	-0.02	0.09
Society	**0.94**	0.03	-0.01	-0.04
Time	0.12	0.09	**0.49**	0.06
Eigenvalues	2.80	0.40	0.36	0.04
Variance	2.60	1.86	1.08	0.33
Alpha	0.94	0.66	0.42	

Table D.2 Interviewer survey: Factor matrix for items about standardized interviewing techniques, including eigenvalues, variances and Cronbach´s Alphas for original five factor solution

	Factor1	Factor2	Factor3	Factor4	Factor5
Explaining	**0.34**	-0.02	0.01	0.04	0.29
Wording	**0.65**	0.05	-0.08	0.05	0.07
Shortening	**0.61**	0.15	0.05	0.02	0.00
Slow	-0.03	0.03	0.02	-0.04	**0.34**
Fast	0.27	0.03	0.12	0.30	-0.20
Completing	**0.47**	-0.00	0.18	-0.13	-0.08
Dialect	0.06	-0.01	-0.05	**0.43**	0.02
Instructions	0.16	**0.62**	0.01	-0.01	-0.06
Interrupt	0.10	**0.44**	0.21	0.06	0.08
Selfadmin	0.04	**0.49**	-0.04	-0.05	0.02
Eigenvalues	2.54	0.50	0.22	0.13	0.02
Variance	2.30	1.68	1.12	0.86	0.84
Alpha	0.70	0.66			

Table D.3 Interviewer survey: Factor matrix for items about standardized interviewing techniques, including eigenvalues, variances and Cronbach´s Alphas for the selected factor solotion

	Factor1	Factor2	Factor3	Factor4	Factor5
Explaining	**0.34**	-0.02	0.01	0.04	0.29
Wording	**0.65**	0.05	-0.08	0.05	0.07
Shortening	**0.61**	0.15	0.05	0.02	0.00
Slow	-0.03	0.03	0.02	-0.04	**0.34**
Fast	0.27	0.03	0.12	0.30	-0.20
Completing	**0.47**	-0.00	0.18	-0.13	-0.08
Dialect	0.06	-0.01	-0.05	**0.43**	0.02
Instructions	0.16	**0.62**	0.01	-0.01	-0.06
Interrupt	0.10	**0.44**	0.21	0.06	0.08
Selfadmin	0.04	**0.49**	-0.04	-0.05	0.02
Eigenvalues	2.54	0.50	0.22	0.13	0.02
Variance	2.30	1.68	1.12	0.86	0.84
Alpha	0.55	0.61	0.66	n/a	n/a

Note. n/a = not applicable.

Table D.4 Interviewer survey: Cronbach´s Alphas for items about how to achieve response

	Factor1	Factor2	Factor3	Factor4	Factor5
Persuasion	**0.41**	-0.12	0.05	-0.01	0.26
Effort	0.00	**0.30**	-0.02	0.03	**0.34**
Privacy	0.01	-0.03	**0.43**	-0.20	-0.01
Refusal	0.32	0.04	**0.40**	0.05	-0.04
Voluntary	-0.03	0.00	**0.52**	0.06	0.04
Contacting	**0.52**	0.15	-0.03	0.03	-0.03
Righttime	-0.05	**0.59**	0.02	-0.02	-0.05
Answers	0.18	-0.11	-0.09	**0.28**	0.03
Confident	0.07	**0.55**	-0.02	-0.03	0.15
Eigenvalues	1.17	0.81	0.54	0.13	0.01
Variance	0.96	0.93	0.79	0.59	0.44
Alpha	0.47	0.57	0.53		

Table D.5 Interviewer survey: Cronbach´s Alphas for items about social desirability

	Factor1	Factor2
Impression	**0.62**	-0.02
Judgement	**0.66**	-0.01
Preference	**0.36**	0.07
Change	-0.02	**0.53**
Honesty	-0.09	**0.47**
Player	-0.13	**-0.40**
Eigenvalues	1.22	0.43
Variance	1.11	0.86
Alpha	0.60	0.52

Table D.6 Interviewer survey: Statistics for indicators on interviewers' attitudes, behavior and personality

		Mean	SE	Min	Max
Reasons for working as an interviewer	science	2.05	1.15	1	7
	people	2.36	0.99	1	6
	formal	1.96	0.87	1	6
How to conduct standardized survey interviews	tailorContent	3.00	0.73	1	4
	tailorTime	3.30	0.50	2	4
	tailorNo	3.50	0.55	1	4
How to achieve response	motivate	2.58	0.71	1	4
	diligent	2.54	0.57	1	4
	voluntary	1.54	0.48	1	3
Trust	trust	3.14	0.86	1	5
Social desirability	self	1.80	0.41	1	3
	others	2.21	0.35	1	3

Notes. SE =standard error. Min = minimum. Max = maximum.

E: Results of multilevel-logistic regressions for succsessful contact and sucsessful cooperation

Table E.1 Model statistics of successful contact and successful cooperation, intercept-only two-level and three-level logistic regression for interviewer and PSU levels, model 0 and model 1, PIAAC 2012

	2-level interviewer		2-level PSU		3-level: interviewer & PSU	
	Contact	**Cooperation**	**Contact**	**Cooperation**	**Contact**	**Cooperation**
$\sigma^2_{u_{0jk}}$ PSU			1.740	0.069	1.256	0.008
			(0.373)	(0.021)	(0.379)	(0.021)
$\sigma^2_{v_{00k}}$ interviewer	1.443	0.073			1.155	0.070
	(0.328)	(0.020)			(0.381)	(0.022)
ICC PSU			34.59	2.04	22.04	0.22
ICC PSU within interviewer					42.29	2.29
ICC interviewer	30.49	2.17			20.25	2.07
Number of interviewers	115	115			115	115
Number of PSUs			251	251	251	251
Number of cases	7902	7450	7902	7450	7902	7450
BIC	2272	9930	2270	9950	2234	9939
AIC	2259	9916	2256	9936	2213	9918

Notes. Contact: Successful contact. Cooperation: Successful cooperation. SE for variance components shown in parentheses. Variance sample person is fixed at π ²/3 (3.290). PSU = primary sample unit. ICC = intra-class correlation coefficient. AIC = Akaike information criterion. BIC = Bayesian information criterion.

Table E.2 Odds rations and model statistics of successful contact, three-level logistic regression, model 1 – model 19, PIAAC 2012

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Sample person characteristics														
Socio-demographic characteristics														
Age			-0.15**	-4.39	-0.15**	-4.39	-0.15**	-4.39	-0.15**	-4.39	-0.15**	-4.39	-0.15**	-4.39
Age squared			0.00**	4.62	0.00**	4.62	0.00**	4.62	0.00**	4.62	0.00**	4.63	0.00**	4.63
Gender: Male			-0.38**	-2.93	-0.38**	-2.94	-0.38**	-2.93	-0.38**	-2.93	-0.38**	-2.93	-0.38**	-2.92
Nationality: German			0.41*	2.14	0.42*	2.14	0.42*	2.14	0.41*	2.14	0.41*	2.14	0.41*	2.14
Area statistics														
PSU size (refrence:0-4,999)														
5,000-19,999			-0.45	-0.92	-0.43	-0.88	-0.44	-0.90	-0.45	-0.92	-0.45	-0.92	-0.46	-0.94
20,000-99,999			-1.65**	-3.25	-1.62**	-3.16	-1.64**	-3.21	-1.65**	-3.27	-1.65**	-3.22	-1.66**	-3.27
100,000-499,999			-1.55*	-2.46	-1.53*	-2.41	-1.54*	-2.43	-1.56*	-2.49	-1.56*	-2.45	-1.56*	-2.48
500,000-99,999,999			-2.49**	-3.53	-2.47**	-3.50	-2.48**	-3.51	-2.50**	-3.56	-2.49**	-3.51	-2.51**	-3.57
Region (reference: North)														
West			-0.34	-0.90	-0.34	-0.90	-0.31	-0.80	-0.34	-0.90	-0.34	-0.91	-0.35	-0.92
South			-0.47	-1.04	-0.44	-0.98	-0.44	-0.98	-0.48	-1.07	-0.43	-0.96	-0.51	-1.13
East			-0.06	-0.12	-0.08	-0.15	-0.03	-0.06	-0.10	-0.21	-0.04	-0.09	-0.13	-0.26
Share of Germans			-0.02	-0.69	-0.02	-0.69	-0.02	-0.71	-0.02	-0.65	-0.02	-0.68	-0.02	-0.66
Share of single households			-0.03	-1.38	-0.03	-1.39	-0.03	-1.39	-0.03	-1.34	-0.03	-1.38	-0.03	-1.34
Share of unemployed			-0.03	-0.47	-0.03	-0.46	-0.03	-0.46	-0.03	-0.51	-0.03	-0.42	-0.03	-0.48
Interviewer characteristics														
Socio-demographic characteristics														
Age (reference: <=45 years)														
45-60 years					0.38	0.87								
61+ years					0.15	0.35								
Gender: Male							0.11	0.42						
Education: Abitur									0.16	0.63				
Occupational status														
Full- or part-time employed											-0.38	-1.08		
Retired													-0.17	-0.61
Experience														
In-study experience														
Work experience: 6-10 years (reference: <=5)														
Work experience: 11+ years (reference: <=5)														
Working hours (reference: <=15 hours)														
16-30 hours														
30+ hours														
Expectations														
Expected response rate: overall														
Expected response rate: income variable														
Reasons for working as an interviewer														
Science														
People														
Formal														
How to conduct standardized survey interviews														
TailorContent														
TailorTime														
TailorNo														
Slow														
Dialect														
How to achieve response														
Motivate														
Diligent														
Voluntary														
Answers														
Trust														
Social desirability														
Self														
Other														
No/few data protection concerns (reference: high concerns)														
$\sigma^2_{u_{0jk}}$ PSU	1.26	0.38	0.09	0.22	0.09	0.23	0.09	0.22	0.07	0.23	0.11	0.22	0.08	0.22
$\sigma^2_{v_{00k}}$ interviewer	1.15	0.38	0.94	0.28	0.91	0.28	0.93	0.28	0.95	0.28	0.90	0.28	0.94	0.28
ICC PSU	22.04		2.06		2.08		2.18		1.64		2.55		1.89	
ICC PSU within interviewer	42.29		23.76		23.23		23.83		23.59		23.50		23.74	
ICC interviewer	20.25		21.70		21.15		21.64		21.96		20.96		21.85	
BIC	2234		2234		2251		2242		2242		2241		2242	
AIC	2213		2115		2118		2117		2117		2116		2117	

Table E.2 Continued

	Model 8		Model 9		Model 10		Model 11		Model 12		Model 13	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Sample person characteristics												
Socio-demographic characteristics												
Age	-0.15**	-4.39	-0.15**	-4.40	-0.15**	-4.38	-0.15**	-4.40	-0.15**	-4.40	-0.15**	-4.40
Age squared	0.00**	4.62	0.00**	4.63	0.00**	4.61	0.00**	4.63	0.00**	4.63	0.00**	4.63
Gender: Male	-0.38**	-2.93	-0.38**	-2.93	-0.38**	-2.94	-0.38**	-2.91	-0.38**	-2.93	-0.38**	-2.92
Nationality: German	0.41*	2.14	0.41*	2.13	0.41*	2.13	0.41*	2.13	0.41*	2.13	0.42*	2.15
Area statistics												
PSU size (refrence:0-4,999)												
5,000-19,999	-0.44	-0.90	-0.46	-0.93	-0.47	-0.95	-0.46	-0.93	-0.43	-0.88	-0.47	-0.96
20,000-99,999	-1.65**	-3.23	-1.67**	-3.25	-1.65**	-3.21	-1.65**	-3.23	-1.65**	-3.22	-1.70**	-3.35
100,000-499,999	-1.55*	-2.44	-1.57*	-2.47	-1.49*	-2.33	-1.56*	-2.47	-1.56*	-2.45	-1.60*	-2.54
500,000-99,999,999	-2.49**	-3.52	-2.52**	-3.50	-2.47**	-3.47	-2.46**	-3.47	-2.50**	-3.51	-2.54**	-3.61
Region (reference: North)												
West	-0.32	-0.80	-0.39	-1.02	-0.45	-1.14	-0.38	-1.00	-0.35	-0.93	-0.32	-0.84
South	-0.48	-1.06	-0.54	-1.17	-0.49	-1.11	-0.52	-1.16	-0.48	-1.06	-0.51	-1.12
East	-0.04	-0.09	-0.10	-0.20	0.01	0.01	-0.12	-0.25	-0.01	-0.02	-0.09	-0.18
Share of Germans	-0.02	-0.72	-0.02	-0.74	-0.02	-0.79	-0.02	-0.69	-0.02	-0.70	-0.02	-0.72
Share of single households	-0.03	-1.40	-0.03	-1.38	-0.04	-1.59	-0.03	-1.29	-0.03	-1.38	-0.03	-1.37
Share of unemployed	-0.03	-0.45	-0.03	-0.45	-0.03	-0.45	-0.03	-0.54	-0.03	-0.46	-0.03	-0.45
Interviewer characteristics												
Socio-demographic characteristics												
Age (reference: <=45 years)												
45-60 years												
61+ years												
Gender: Male												
Education: Abitur												
Occupational status												
Full- or part-time employed												
Retired												
Experience												
In-study experience	0.08	0.25										
Work experience: 6-10 years (reference: <=5)			0.15	0.41								
Work experience: 11+ years (reference: <=5)			-0.10	-0.35								
Working hours (reference: <=15 hours)												
16-30 hours					0.82+	1.89						
30+ hours					0.70	1.55						
Expectations												
Expected response rate: overall							0.01	1.30				
Expected response rate: income variable									0.01	1.01		
Reasons for working as an interviewer												
Science											0.15	1.10
People											-0.08	-0.51
Formal											-0.19	-1.28
How to conduct standardized survey interviews												
TailorContent												
TailorTime												
TailorNo												
Slow												
Dialect												
How to achieve response												
Motivate												
Diligent												
Voluntary												
Answers												
Trust												
Social desirability												
Self												
Other												
No/few data protection concerns (reference: high concerns)												
$\sigma^2_{u_{0jk}}$ PSU	0.09	0.22	0.09	0.22	0.12	0.22	0.11	0.22	0.12	0.22	0.07	0.22
$\sigma^2_{v_{00k}}$ interviewer	0.94	0.28	0.92	0.28	0.85	0.27	0.92	0.28	0.91	0.28	0.94	0.28
ICC PSU	2.12		2.19		2.84				2.68		1.62	
ICC PSU within interviewer	23.82		23.57		22.70		23.79		23.70		23.50	
ICC interviewer	21.69		21.39		19.86		21.22		21.02		21.88	
BIC	2243		2251		2248		2241		2242		2258	
AIC	2117		2119		2116		2115		2116		2119	

Table E.2 Continued

	Model 14		Model 15		Model 16		Model 17		Model 18		Model 19	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Sample person characteristics												
Socio-demographic characteristics												
Age	-0.15**	-4.40	-0.15**	-4.41	-0.15**	-4.38	-0.15**	-4.39	-0.15**	-4.39	-0.15**	-4.42
Age squared	0.00**	4.63	0.00**	4.64	0.00**	4.62	0.00**	4.63	0.00**	4.63	0.00**	4.65
Gender: Male	-0.38**	-2.92	-0.38**	-2.92	-0.38**	-2.93	-0.38**	-2.94	-0.38**	-2.93	-0.38**	-2.90
Nationality: German	0.42*	2.19	0.41*	2.11	0.42*	2.14	0.41*	2.13	0.41*	2.14	0.42*	2.18
Area statistics												
PSU size (refrence:0-4,999)												
5,000-19,999	-0.41	-0.86	-0.50	-1.01	-0.44	-0.91	-0.46	-0.94	-0.45	-0.93	-0.51	-1.03
20,000-99,999	-1.65**	-3.34	-1.69**	-3.31	-1.63**	-3.20	-1.66**	-3.27	-1.65**	-3.26	-1.67**	-3.16
100,000-499,999	-1.56*	-2.57	-1.62*	-2.56	-1.53*	-2.42	-1.56*	-2.48	-1.55*	-2.46	-1.62*	-2.52
500,000-99,999,999	-2.43**	-3.57	-2.58**	-3.64	-2.46**	-3.47	-2.51**	-3.57	-2.49**	-3.53	-2.47**	-3.40
Region (reference: North)												
West	-0.29	-0.75	-0.39	-1.04	-0.33	-0.86	-0.35	-0.92	-0.34	-0.90	-0.38	-0.85
South	-0.32	-0.70	-0.35	-0.78	-0.41	-0.91	-0.46	-1.03	-0.46	-1.02	0.03	0.05
East	-0.14	-0.30	0.01	0.02	-0.05	-0.10	-0.10	-0.20	-0.06	-0.12	-0.06	-0.13
Share of Germans	-0.01	-0.49	-0.02	-0.88	-0.02	-0.67	-0.02	-0.64	-0.02	-0.68	-0.01	-0.44
Share of single households	-0.03	-1.33	-0.03	-1.42	-0.03	-1.41	-0.03	-1.31	-0.03	-1.39	-0.03	-1.10
Share of unemployed	-0.02	-0.34	-0.02	-0.35	-0.03	-0.44	-0.03	-0.48	-0.03	-0.46	-0.02	-0.26
Interviewer characteristics												
Socio-demographic characteristics												
Age (reference: <=45 years)												
45-60 years											0.73+	1.67
61+ years											0.46	0.90
Gender: Male											0.18	0.65
Education: Abitur											0.20	0.72
Occupational status												
Full- or part-time employed											-0.38	-0.95
Retired											0.10	0.24
Experience												
In-study experience											0.09	0.23
Work experience: 6-10 years (reference: <=5)											0.14	0.38
Work experience: 11+ years (reference: <=5)											0.09	0.29
Working hours (reference: <=15 hours)												
16-30 hours											0.60	1.23
30+ hours											0.27	0.52
Expectations												
Expected response rate: overall											0.02	1.57
Expected response rate: income variable											0.00	0.04
Reasons for working as an interviewer												
Science											-0.01	-0.06
People											-0.06	-0.38
Formal											-0.17	-1.06
How to conduct standardized survey interviews												
TailorContent	-0.08	-0.42									-0.18	-0.77
TailorTime	-0.72*	-2.40									-0.56+	-1.70
TailorNo	0.37	1.55									0.33	1.29
Slow	0.25	0.70									0.41	1.14
Dialect	-0.16	-1.22									-0.21	-1.47
How to achieve response												
Motivate			0.11	0.57							0.22	1.11
Diligent			-0.29	-1.33							-0.36	-1.48
Voluntary			0.70*	2.35							0.62+	1.83
Answers			0.02	0.09							-0.07	-0.39
Trust					-0.10	-0.67					-0.19	-1.21
Social desirability												
Self							-0.10	-0.33			-0.01	-0.02
Other							0.15	0.42			0.49	1.26
No/few data protection concerns												
(reference: high concerns)									-0.08	-0.29	0.04	0.12
$\sigma^2_{u_{0jk}}$ PSU	0.03	0.23	0.10	0.22	0.10	0.22	0.07	0.22	0.09	0.22	0.16	0.24
$\sigma^2_{v_{00k}}$ interviewer	0.83	0.26	0.83	0.26	0.92	0.28	0.94	0.28	0.93	0.28	0.48	0.21
ICC PSU	0.64		2.29		2.30		1.68		2.02		4.19	
ICC PSU within interviewer	20.71		21.97		23.66		23.62		23.64		16.41	
ICC interviewer	20.13		19.68		21.37		21.94		21.62		12.22	
BIC	2270		2261		2242		2251		2243		2462	
AIC	2116		2115		2117		2119		2117		2141	

Notes. Contact: Successful contact. Standard errors in parentheses. Variance sample person is fixed at $\pi^2/3$ (3.290). All models control for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). Number of interviewers = 115. Number of PSUs = 251. Number of sample persons = 7,902. Coef. = coefficient. SE = standard error. ICC = intra-class correlation coefficient. PSU = primary sample unit. AIC = Akaike information criterion. BIC = Bayesian information criterion.

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

F: The core questionnaire (CAPI): Original German question text and English translation of substative variables under research

Item	English item text	German item text
Every day skill use: literacy - reading		
	I would now like to talk about your reading activities. Include any reading you might do on computer screens or other electronic displays. In everyday life, how often do you usually …	Nun möchte ich mit Ihnen über das Lesen im Alltag sprechen. Bitte berücksichtigen Sie dabei auch das Lesen am Bildschirm. In Ihrem Alltag, wie oft lesen Sie gewöhnlich ...
	1 Never 2 Less than one a month 3 Less than once a week but at least once a month 4 At least once a week but not every day 5 Every day	1 Nie 2 Seltener als einmal im Monat 3 Seltener als einmal pro Woche, aber mindestens einmal im Monat 4 Mindestens einmal pro Woche, aber nicht täglich 5 Täglich
H_Q01a	read directions or instructions?	Anleitungen oder Anweisungen?
H_Q01b	read letters, memos or e-mails?	Briefe, kurze Mittteilungen oder E-Mails?
H_Q01c	read articles in newspapers, magazines or newsletters?	Artikel in Zeitungen, Zeitschriften oder Newsletter?
H_Q01d	read articles in professional journals or scholarly publications?	Artikel in Fachzeitschriften oder wissenschaftliche Veröffentlichungen?
H_Q01e	read books, fiction or non-fiction?	Bücher, zum Beispiel Romane oder Sachbücher?
H_Q01f	read manuals or reference materials?	Handbücher oder Nachschlagewerke?
H_Q01g	read bills, invoices, bank statements or other financial statements?	Rechnungen, Bankauszüge oder Ähnliches?
H_Q01h	read diagrams, maps, or schematics?	Diagramme, Pläne, Karten oder Schaubilder?
Every day skill use: literacy - writing		
	The following questions are about your writing activities. Everyday writing includes any writing you might do on computers or other electronic devices. In everyday life, how often do you usually …	In den folgenden Fragen geht es um das Schreiben im Alltag. Bitte berücksichtigen Sie dabei auch das Schreiben am Computer oder an anderen elektronischen Geräten. In Ihrem Alltag, wie oft ...
	1 Never 2 Less than one a month 3 Less than once a week but at least once a month 4 At least once a week but not every day 5 Every day	1 Nie 2 Seltener als einmal im Monat 3 Seltener als einmal pro Woche, aber mindestens einmal im Monat 4 Mindestens einmal pro Woche, aber nicht täglich 5 Täglich
H_Q02a	write letters, memos or e-mails?	schreiben Sie normalerweise Briefe, kurze Mitteilungen oder E-Mails?
H_Q02b	write articles for newspapers, magazines or	schreiben Sie normalerweise Artikel für

	newsletters?	Zeitungen, Zeitschriften oder Newsletter?
H_Q02c	write reports?	schreiben Sie normalerweise Berichte?
H_Q02d	fill in forms?	füllen Sie normalerweise Formulare aus?

Every day skill use: numeracy

	The following questions are about activities that you undertake everyday numbers. In everyday life, how often do you usually …	In den folgenden Fragen geht es um Tätigkeiten im Alltag, die mit Zahlen, Mengenangaben, Statistik oder Mathematik zu tun haben. In Ihrem Alltag, wie oft ...
	1 Never 2 Less than one a month 3 Less than once a week but at least once a month 4 At least once a week but not every day 5 Every day	1 Nie 2 Seltener als einmal im Monat 3 Seltener als einmal pro Woche, aber mindestens einmal im Monat 4 Mindestens einmal pro Woche, aber nicht täglich 5 Täglich
H_Q03b	calculate prices, costs or budgets?	berechnen Sie normalerweise Preise, Kosten oder Budgets?
H_Q03c	use or calculate fractions, decimals or percentages?	verwenden oder berechnen Sie normalerweise Brüche, Dezimal- oder Prozentangaben?
H_Q03d	use a calculator - either hand-held or computer based?	verwenden Sie normalerweise einen Taschenrechner, egal ob als eigenständiges Gerät oder im Computer?
H_Q03f	prepare charts, graphs or tables?	erstellen Sie normalerweise Diagramme, Schaubilder oder Tabellen?
H_Q03g	use simple algebra or formulas?	verwenden Sie normalerweise einfache Formeln oder Ähnliches?
H_Q03h	use more advanced math or statistics such as calculus, complex algebra, trigonometry or use of regression techniques?	verwenden Sie normalerweise höhere Mathematik oder Statistik, wie zum Beispiel Analysis, komplexe Algebra, Trigonometrie oder Regressionsanalysen?

Every day skill use: ICT

	The following questions are about the use of computers or the internet everyday computers. This could be at home or in other places that offer internet services, like internet cafes or libraries. In everyday life, how often do you usually …	In den folgenden Fragen geht es um die Nutzung von Computern oder des Internets in Ihrem Alltag. Dies kann zu Hause sein oder dort, wo ein Internetzugang angeboten wird, wie z.B. in Internetcafés oder Büchereien. In Ihrem Alltag, wie oft ...
	1 Never 2 Less than one a month 3 Less than once a week but at least once a month 4 At least once a week but not every day 5 Every day	1 Nie 2 Seltener als einmal im Monat 3 Seltener als einmal pro Woche, aber mindestens einmal im Monat 4 Mindestens einmal pro Woche, aber nicht täglich 5 Täglich
H_Q05a	use email?	verwenden Sie normalerweise E-Mail?
H_Q05c	nutzen Sie normalerweise das Internet, um	use the internet in order to better

	Sachverhalte oder Themen besser zu verstehen, bei denen es z.B. um Ihre Gesundheit oder um Krankheiten geht, um finanzielle Dinge oder um die Umwelt?	understand issues related to, for example, your health or illnesses, financial matters, or environmental issues?
H_Q05d	conduct transactions on the internet, for example buying or selling products or services, or banking?	führen Sie normalerweise Transaktionen im Internet durch, z.B. Kauf oder Verkauf von Produkten oder Dienstleistungen oder Online-Banking?
H_Q05e	use spreadsheet software, for example Excel?	benutzen Sie normalerweise ein Tabellenkalkulationsprogramm, wie zum Beispiel Excel?
H_Q05f	use a word processor, for example Word?	benutzen Sie normalerweise ein Textverarbeitungsprogramm, wie zum Beispiel Word?
H_Q05g	use a programming language to program or write computer code?	benutzen Sie normalerweise eine Programmiersprache, um zu programmieren oder um Computercode zu schreiben?
H_Q05h	participate in real-time discussions on the internet, for example online conferences or chat groups?	nehmen Sie normalerweise an Echtzeitdiskussionen im Internet teil, z.B. an Onlinekonferenzen oder Chatgruppen?

About yourself: learning strategies

	I would now like to ask you some questions about how you deal with problems and tasks you encounter. To what extent do the following statements apply to you?	Nun möchte ich Ihnen einige Fragen dazu stellen, wie Sie mit Problemen und Aufgaben umgehen. In welchem Ausmaß treffen die folgenden Aussagen auf Sie zu?
	1 Not at all 2 Very little 3 To some extent 4 To a high extent 5 To a very high extent	1 Überhaupt nicht 2 In sehr geringem Maße 3 In gewissem Maße 4 In hohem Maße 5 In sehr hohem Maße
I_Q04b	When I hear or read about new ideas, I try to relate them to real life situations to which they might apply	Wenn ich von neuen Ideen höre oder lese, versuche ich sie auf passende Situationen in meinem Leben zu übertragen.
I_Q04d	I like learning new things	Ich lerne gerne Neues.
I_Q04h	When I come across something new, I try to relate it to what I already know	Um mit neuen Dingen umzugehen, versuche ich sie auf etwas zu beziehen, was ich bereits kenne.
I_Q04j	I like to get to the bottom of difficult things	Schwierigen Dingen gehe ich gerne auf den Grund.
I_Q04l	I like to figure out how different ideas fit together	Ich finde gerne heraus, wie verschiedene Ideen zusammenpassen.
I_Q04m	If I don't understand something, I look for additional information to make it clearer	Wenn ich etwas nicht verstehe, suche ich nach zusätzlichen Informationen, um mehr Klarheit zu gewinnen.

About yourself: cultural engagement

I_Q05f	In the last 12 months, how often, if at all, did you do voluntary work, including unpaid work for a charity, political party, trade union or other non-profit organisation?	In den letzten 12 Monaten, wie oft waren Sie - falls überhaupt - ehrenamtlich tätig, z.B. durch unbezahlte Arbeit für eine Wohltätigkeitsorganisation, eine politische Partei, eine Gewerkschaft oder eine

		sonstige gemeinnützige Organisation?
	1 Never 2 Less than one a month 3 Less than once a week but at least once a month 4 At least once a week but not every day 5 Every day	1 Nie 2 Seltener als einmal im Monat 3 Seltener als einmal pro Woche, aber mindestens einmal im Monat 4 Mindestens einmal pro Woche, aber nicht täglich 5 Täglich
About yourself: political efficacy and social trust		
	To what extent do you agree or disagree with the following statements? 1 Strongly agree 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Strongly disagree	Sagen Sie mir bitte, inwieweit Sie den folgenden Aussagen zustimmen oder nicht zustimmen. 1 Stimme voll und ganz zu 2 Stimme zu 3 Stimme teils zu, stimme teils nicht zu 4 Stimme nicht zu 5 Stimme überhaupt nicht zu
I_Q06a	People like me don't have any say about what the government does	Menschen wie ich haben keinerlei Einfluss darauf, was die Regierung macht.
I_Q07a	There are only a few people you can trust completely	Es gibt nur wenige Menschen, denen man voll vertrauen kann.
I_Q07b	If you are not careful, other people will take advantage of you	Wenn man nicht aufpasst, wird man von anderen ausgenutzt.
Background: number of books		
J_Q08	About how many books were there in your home when you were aged 16? Do not include magazines, newspapers or schoolbooks. To give an estimation, one meter of shelving is about 40 books. 1 10 books or less 2 11 to 25 books 3 26 to 100 books 4 101 to 200 books 5 201 to 500 books 6 More than 500 books	Wie viele Bücher gab es ungefähr bei Ihnen zu Hause, als Sie 16 Jahre alt waren? Zählen Sie bitte keine Zeitschriften, Zeitungen oder Schulbücher mit. Als Hinweis: auf einen Meter Regalbrett passen ungefähr 40 Bücher. 1 10 Bücher oder weniger 2 11 bis 25 Bücher 3 26 bis 100 Bücher 4 101 bis 200 Bücher 5 201 bis 500 Bücher 6 Mehr als 500 Bücher

Note. Questions introduction as well as question text differed slightly depending on the routing for each person (for more information see questionnaire documenatation in Rammstedt et al., 2016).

G: Results of multilevel-linear regressions for the estimates of substantive survey variables

Table G.1 Model statistics for the literacy and numeracy mean scores, intercept-only two-level linear regression for interviewer and PSU level, PIAAC 2012

	Interviewer		PSU	
	Literacy Coef. (SE)	**Numeracy** Coef. (SE)	**Literacy** Coef. (SE)	**Numeracy** Coef. (SE)
$\sigma^2_{e_{ijk}}$ respondent	1712.73	2097.31	1683.86	2066.34
	(38.17)	(46.74)	(38.11)	(46.77)
$\sigma^2_{u_{0jk}}$ PSU			139.27	152.21
			(22.23)	(25.64)
$\sigma^2_{v_{00k}}$ interviewer	112.57	121.80		
	(22.01)	(24.63)		
ICC PSU			7.64	6.86
ICC interviewer	6.17	5.49		
Number of Interviewers	107	107		
Number of PSUs			240	240
BIC	42649	43477	42651	43481
AIC	42630	43458	42632	43462

Notes. Literacy and numeracy mean scores range from 0 to 500. Sample items for literacy and numeracy can be found in Appendix A. Number of respondents = 4,132. CASI = computer assisted self-interviewing. AIC = Akaike information criterion. BIC = Bayesian information criterion. Coef. = coefficient. SE = standard error. PSU = primary sample unit. ICC = intra-class correlation coefficient.

Table G.2 Variance components including standard errors and ICCs for the estimates of the CAPI variables, excluding (model 1) and including (model 2) sample composition variables, three-level linear regression, PIAAC 2012

		Interviewer				PSU				Respondent				Interviewer		PSU		PSU within interviewer	
		Variances and standard errors												ICCs					
		Model 1		Model 2		Model 1		Model 2		Model 1		Model 2		Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
		Var	SE	Var	SE	Var	SE	Var	SE	Var	SE	Var	SE	ICC	ICC	ICC	ICC	ICC	ICC
CAPI	H_Q01a	0.06	0.01	0.06	0.01	0.02	0.01	0.02	0.01	1.31	0.03	1.29	0.03	4.27	4.14	1.65	1.41	5.93	5.55
	H_Q01b	0.04	0.01	0.03	0.01	0.01	0.01	0.00	0.01	1.16	0.03	1.00	0.02	3.18	2.92	1.16	0.07	4.34	2.99
	H_Q01c	0.03	0.01	0.03	0.01	0.00	0.00	0.00	0.00	1.10	0.02	1.03	0.02	2.94	2.66	0.00	0.00	2.94	2.66
	H_Q01d	0.05	0.01	0.03	0.01	0.00	0.00	0.00	0.00	1.52	0.03	1.39	0.03	3.00	2.37	0.00	0.00	3.00	2.37
	H_Q01e	0.02	0.02	0.00	0.00	0.04	0.02	0.03	0.01	1.99	0.05	1.72	0.04	0.95	0.00	2.15	1.47	3.10	1.47
	H_Q01f	0.06	0.01	0.05	0.01	0.01	0.01	0.00	0.01	1.25	0.03	1.18	0.03	4.33	4.29	0.44	0.23	4.78	4.52
	H_Q01g	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.91	0.02	0.83	0.02	1.33	0.91	0.05	0.00	1.39	0.91
	H_Q01h	0.08	0.02	0.07	0.01	0.02	0.01	0.01	0.01	1.27	0.03	1.12	0.03	5.89	5.49	1.22	1.15	7.12	6.64
	H_Q02a	0.01	0.01	0.01	0.01	0.02	0.01	0.00	0.01	1.45	0.03	1.10	0.02	0.88	0.66	1.43	0.24	2.31	0.90
	H_Q02b	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.01	0.23	0.01	0.21	0.00	0.26	0.56	0.47	0.56
	H_Q02c	0.02	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.82	0.02	0.75	0.02	2.37	1.66	0.25	0.00	2.62	1.66
	H_Q02d	0.05	0.01	0.05	0.01	0.00	0.01	0.00	0.01	0.70	0.02	0.69	0.02	6.30	6.38	0.52	0.37	6.82	6.75
	H_Q03b	0.13	0.02	0.12	0.02	0.00	0.01	0.00	0.00	1.66	0.04	1.59	0.04	7.31	6.86	0.05	0.00	7.37	6.86
	H_Q03c	0.11	0.02	0.10	0.02	0.00	0.00	0.00	0.00	1.65	0.04	1.49	0.03	6.34	6.16	0.00	0.00	6.34	6.16
	H_Q03d	0.04	0.01	0.04	0.01	0.01	0.01	0.00	0.01	1.46	0.03	1.34	0.03	2.91	2.60	0.56	0.17	3.47	2.77
	H_Q03f	0.02	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.81	0.02	0.69	0.02	2.92	2.10	0.89	0.64	3.81	2.74
	H_Q03g	0.10	0.02	0.09	0.02	0.02	0.01	0.02	0.01	1.48	0.03	1.28	0.03	6.48	6.64	0.94	1.17	7.43	7.81
	H_Q03h	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.01	0.54	0.01	0.17	0.00	0.00	0.00	0.17	0.00
	H_Q05a	0.01	0.01	0.01	0.01	0.02	0.01	0.00	0.00	1.24	0.03	1.11	0.03	1.15	0.65	1.28	0.00	2.43	0.65
	H_Q05c	0.05	0.01	0.05	0.01	0.02	0.01	0.00	0.01	1.01	0.02	0.95	0.02	4.80	4.59	1.49	0.21	6.29	4.80
	H_Q05d	0.00	0.01	0.00	0.00	0.03	0.01	0.01	0.01	1.49	0.04	1.31	0.03	0.16	0.00	1.68	1.03	1.84	1.03
	H_Q05e	0.03	0.01	0.02	0.01	0.00	0.00	0.00	0.00	1.14	0.03	1.07	0.03	2.45	2.18	0.00	0.00	2.45	2.18
	H_Q05f	0.02	0.01	0.01	0.01	0.05	0.02	0.05	0.02	1.36	0.03	1.25	0.03	1.11	1.14	3.56	3.56	4.67	4.70
	H_Q05g	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.01	0.37	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	H_Q05h	0.02	0.02	0.00	0.01	0.05	0.02	0.05	0.02	2.12	0.05	1.48	0.04	0.92	0.12	2.16	3.02	3.07	3.14
	I_Q04b	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.67	0.02	0.63	0.01	2.08	1.36	0.00	0.00	2.08	1.36
	I_Q04d	0.02	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.71	0.02	0.63	0.01	2.28	1.79	0.19	0.00	2.47	1.79
	I_Q04h	0.01	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.72	0.02	0.66	0.01	1.15	1.47	1.01	0.00	2.16	1.47
	I_Q04j	0.03	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.82	0.02	0.79	0.02	3.04	2.87	0.00	0.00	3.04	2.87
	I_Q04l	0.03	0.01	0.03	0.01	0.01	0.01	0.00	0.01	0.83	0.02	0.79	0.02	3.78	3.57	0.61	0.23	4.39	3.80
	I_Q04m	0.02	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.64	0.01	0.59	0.01	2.90	2.83	0.42	0.00	3.32	2.83
	I_Q05f	0.02	0.01	0.00	0.01	0.03	0.01	0.01	0.01	1.28	0.03	1.25	0.03	1.19	0.13	2.39	1.18	3.58	1.32
	I_Q06a	0.04	0.01	0.00	0.00	0.01	0.01	0.01	0.01	1.35	0.03	1.28	0.03	2.55	0.00	0.89	0.76	3.44	0.76
	I_Q07a	0.04	0.01	0.02	0.01	0.01	0.01	0.00	0.00	1.09	0.02	1.03	0.02	3.16	2.17	0.94	0.00	4.10	2.17
	I_Q07b	0.03	0.01	0.02	0.01	0.00	0.01	0.00	0.00	0.88	0.02	0.81	0.02	3.82	1.85	0.20	0.00	4.02	1.85
	J_Q08	0.09	0.02	0.04	0.01	0.01	0.02	0.00	0.00	1.83	0.04	1.28	0.03	4.81	2.67	0.72	0.00	5.53	2.67
CASI	Literacy	88.71	23.20	49.09	13.24	54.57	18.67	19.88	10.70	1680.53	38.19	1175.94	26.70	4.86	3.94	2.99	1.60	7.86	5.54
	Numeracy	95.67	26.13	54.13	15.09	60.71	22.00	24.91	12.63	2061.29	46.83	1376.39	31.23	4.31	3.72	2.74	1.71	7.05	5.43

Notes. Model 1 represents the intercept only model. Model 2 controlled for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Literacy and numeracy mean scores range from 0 to 500. Sample items for literacy and numeracy can be found in Appendix A. Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132 (3,613 for H_Q05c). PSU = primary sample unit. ICC = intra-class correlation

coefficient. Var = variance. SE = standard error. CAPI = computer-assisted personal interviewing. CASI = computer-assisted self-interviewing.

Table G.3 Regression coefficients and standard errors for the literacy and numeracy mean scores, three-level linear regression, models 2-23, PIAAC 2012

	Literacy		Numeracy	
	Coef.	**SE**	**Coef.**	**SE**
Interviewer characteristics				
Socio-demographic characteristics				
Age (reference: <=45 years)				
45-60 years				
61+ years				
Gender: Male				
Education: Abitur			3.36*	1.97
Occupational status				
Full- or part-time employed				
Retired				
Experience				
In-study experience	5.45**	2.21	7.34***	2.32
Work experience: 6-10 years (reference: <=5)				
Work experience: 11+ years (reference: <=5)	-3.92*	2.15		
Working hours (reference: <=15 hours)				
16-30 hours				
30+ hours				
Expectations				
Expected response rate: overall	-0.25***	0.07	-0.22***	0.08
Expected response rate: income variable				
Reasons for working as an interviewer				
Science				
People				
Formal				
How to conduct standardized survey interviews				
TailorContent				
TailorTime			-4.35*	2.34
TailorNo				
Slow	6.38**	(2.89)	7.16**	3.04
Dialect				
How to achieve response				
Motivate				
Diligent	-3.64**	1.64	-3.92**	1.75
Voluntary				
Answers				
Trust				
Social desirability				
Self				
Other				
Data protection concerns (reference: high concerns)				
Audio indicators (reference: incorrect)				
Formal criteria collected partly correct	-5.25**	2.64	-5.44*	2.82
Formal criteria collected correct				
Permission to record collected partly correct				
Permission to record collected correct				
Number of incorrect skipped questions				
Number of incorrect read questions				

Notes. Literacy and numeracy mean scores range from 0 to 500. Sample items for literacy and numeracy can be found in Appendix A. All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Only one interviewer characteristic included per model. Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132. Coef. = coefficient. SE = standard error.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table G.4 Regression coefficients and standard errors for the means of the CAPI variables, three-level linear regression, models 2-23, PIAAC 2012

	H_Q01a		H_Q01f		H_Q01h		H_Q02d	
	Coef.	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**
Interviewer characteristics								
Socio-demographic characteristics								
Age (reference: <=45 years)								
45-60 years								
61+ years								
Gender: Male								
Education: Abitur								
Occupational status								
Full- or part-time employed								
Retired								
Experience								
In-study experience					0.15**	0.08		
Work experience: 6-10 years (reference: <=5)								
Work experience: 11+ years (reference: <=5)					-0.19***	0.07		
Working hours (reference: <=15 hours)								
16-30 hours								
30+ hours	-0.19***	0.07						
Expectations								
Expected response rate: overall								
Expected response rate: income variable								
Reasons for working as an interviewer								
Science								
People								
Formal								
How to conduct standardized survey interviews								
TailorContent					0.09*	0.05		
TailorTime								
TailorNo								
Slow								
Dialect								
How to achieve response								
Motivate								
Diligent	-0.13**	0.05	-0.13***	0.05	-0.16***	0.05		
Voluntary								
Answers								
Trust								
Social desirability								
Self								
Other								
Data protection concerns (reference: high concerns)	0.14**	0.07						
Audio indicators (reference: incorrect)								
Formal criteria collected partly correct					-0.17*	0.09	-0.15**	0.07
Formal criteria collected correct	-0.12*	0.07			-0.14**	0.07		
Permission to record collected partly correct								
Permission to record collected correct	-0.14*	0.08						
Number of incorrect skipped questions	-0.05***	0.01			-0.03**	0.01		
Number of incorrect read questions	-0.01***	0.00			-0.01**	0.00		

Table Table G.4 continued

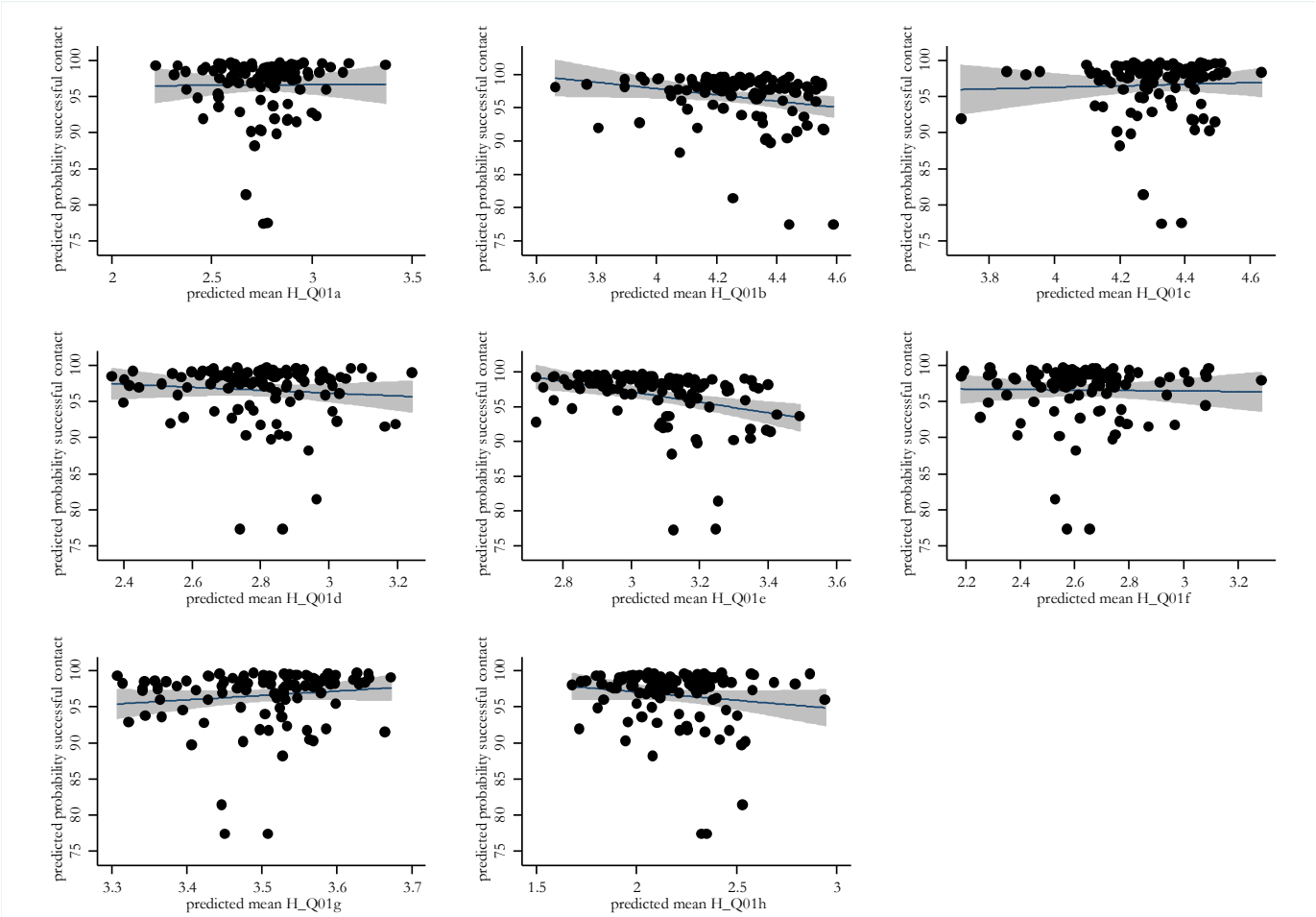
	H_Q03b		H_Q03c		H_Q03g		H_Q05c		I_Q04l	
	Coef.	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**
Interviewer characteristics										
Socio-demographic characteristics										
Age (reference: <=45 years)										
45-60 years	0.25*	0.15								
61+ years										
Gender: Male										
Education: Abitur					0.13*	0.07				
Occupational status										
Full- or part-time employed										
Retired									0.08*	0.05
Experience										
In-study experience							0.19***	0.06		
Work experience: 6-10 years (reference: <=5)									0.12**	0.06
Work experience: 11+ years (reference: <=5)	-0.17*	0.09			-0.23***	0.08				
Working hours (reference: <=15 hours)										
16-30 hours			0.23*	0.13	0.25*	0.13				
30+ hours										
Expectations										
Expected response rate: overall										
Expected response rate: income variable										
Reasons for working as an interviewer										
Science									-0.04**	0.02
People							-0.06*	0.03		
Formal										
How to conduct standardized survey interviews										
TailorContent			0.10*	0.06	0.10*	0.06				
TailorTime									0.10*	0.05
TailorNo							-0.12*	0.07		
Slow										
Dialect							0.09***	0.03		
How to achieve response										
Motivate										
Diligent			-0.12*	0.06	-0.11*	0.06				
Voluntary			0.13*	0.08	0.14*	0.08				
Answers										
Trust	0.09*	0.05							-0.07***	0.03
Social desirability										
Self										
Other										
Data protection concerns (reference: high concerns)										
Audio indicators (reference: incorrect)										
Formal criteria collected partly correct										
Formal criteria collected correct									0.12**	0.05
Permission to record collected partly correct										
Permission to record collected correct					-0.19**	0.09	0.12*	0.07	0.10*	0.06
Number of incorrect skipped questions	-0.03*	0.02			-0.04**	0.02	0.03**	0.01		
Number of incorrect read questions					-0.01***	0.00	0.01***	0.00		

Notes. Means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Only one interviewer characteristic included per model. Number of interviewers = 107. Number of PSUs = 240. Number of respondents = 4,132 (3,613 for H_Q05c). Coef. = coefficient. SE = standard error.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

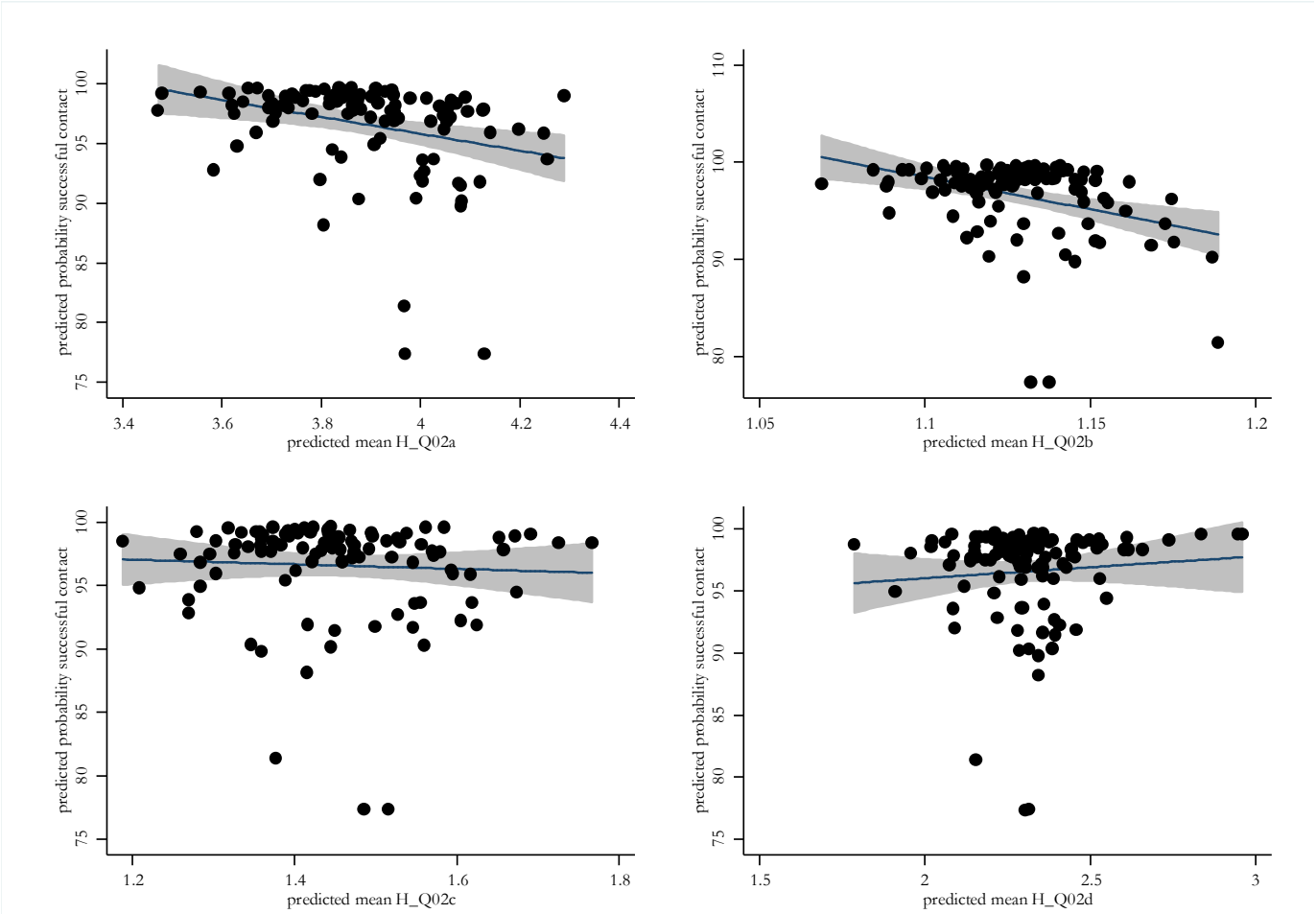
H: Relationship between the predicted probability for successful contact and successful cooperation and the predicted means of CAPI variables by interviewer

Figure H.1: Relationship between the predicted probability for successful contact and the predicted means of the CAPI variables of section H_Q01 by interviewer, PIAAC 2012



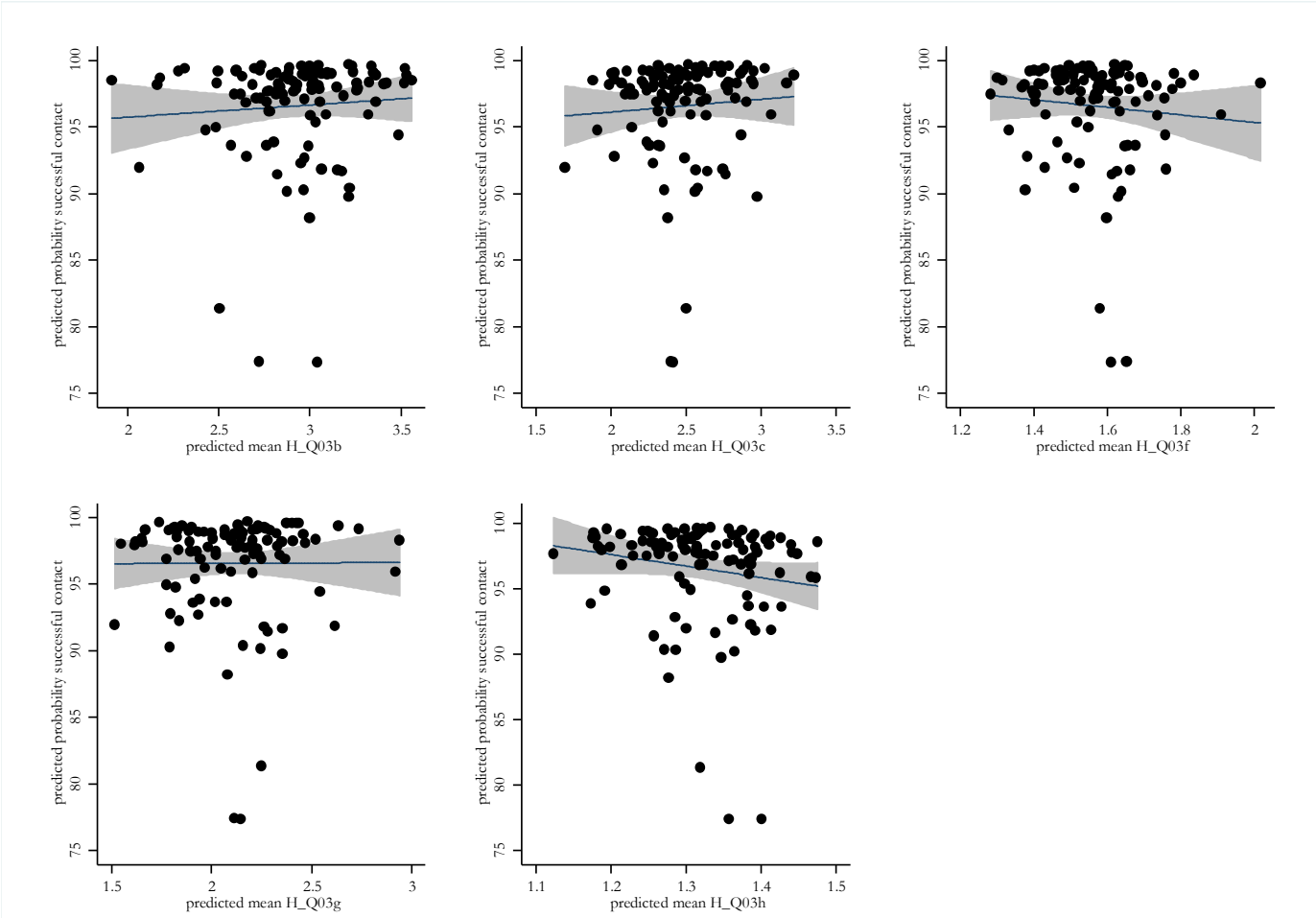
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful contact and CAPI variables of section H_Q01. Significant = $p < 0.05$. Contact: Successful contact. Predicted means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Contact: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Contact: Number of sample persons = 7,902. CAPI: Number of respondents = 4,132.

Figure H.2: Relationship between the predicted probability for successful contact and the predicted means of the CAPI variables of section H_Q02 by interviewer, PIAAC 2012



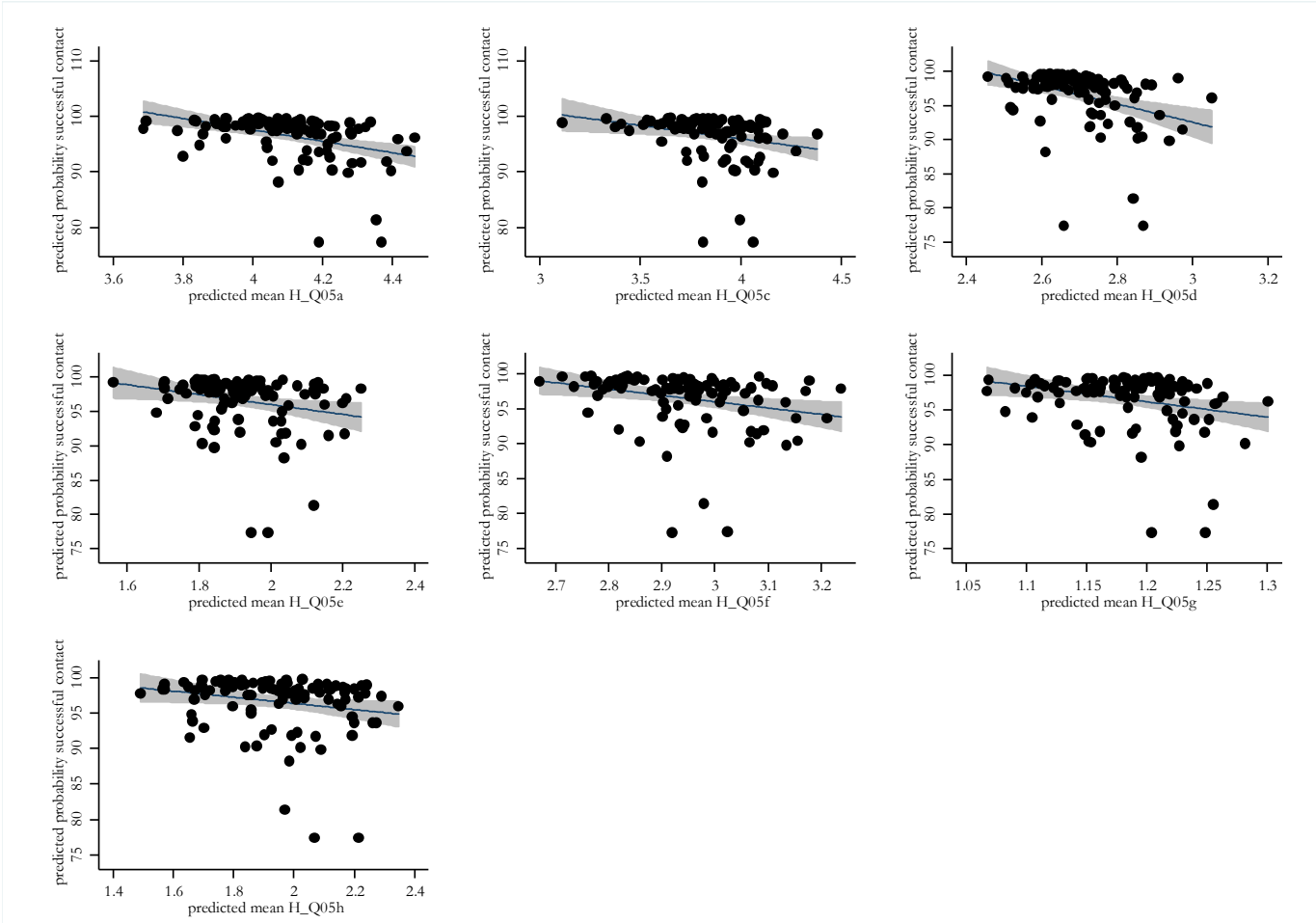
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful contact and CAPI variables of section H_Q01. Significant = $p < 0.05$. Contact: Successful contact. Predicted means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Contact: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Contact: Number of sample persons = 7,902. CAPI: Number of respondents = 4,132.

Figure H.3: Relationship between the predicted probability for successful contact and the predicted means of the CAPI variables of section H_Q03 by interviewer, PIAAC 2012



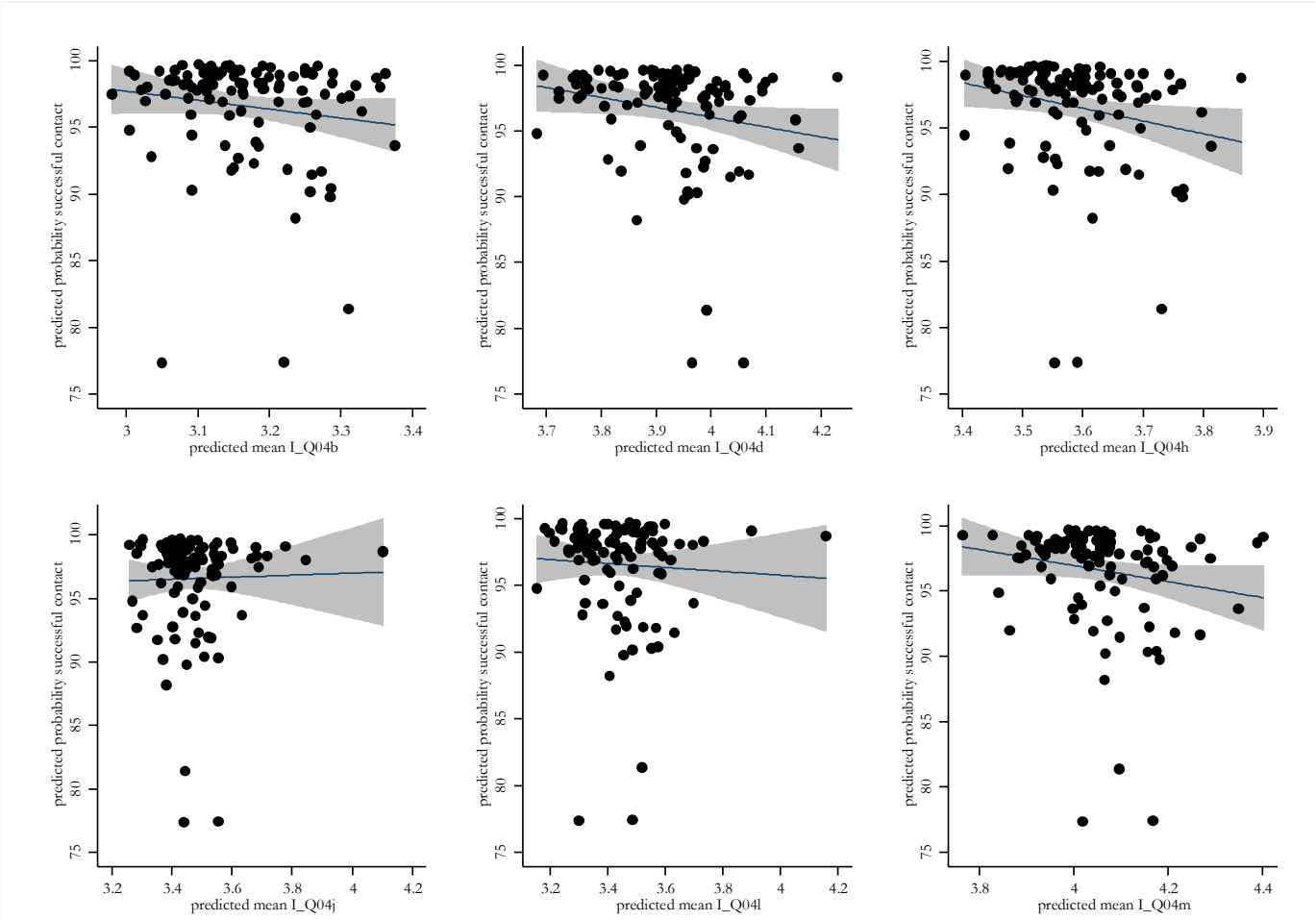
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful contact and CAPI variables of section H_Q01. Significant = $p < 0.05$. Contact: Successful contact. Predicted means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Contact: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Contact: Number of sample persons = 7,902. CAPI: Number of respondents = 4,132.

Figure H.4: Relationship between the predicted probability for successful contact and the predicted means of the CAPI variables of section H_Q05 by interviewer, PIAAC 2012



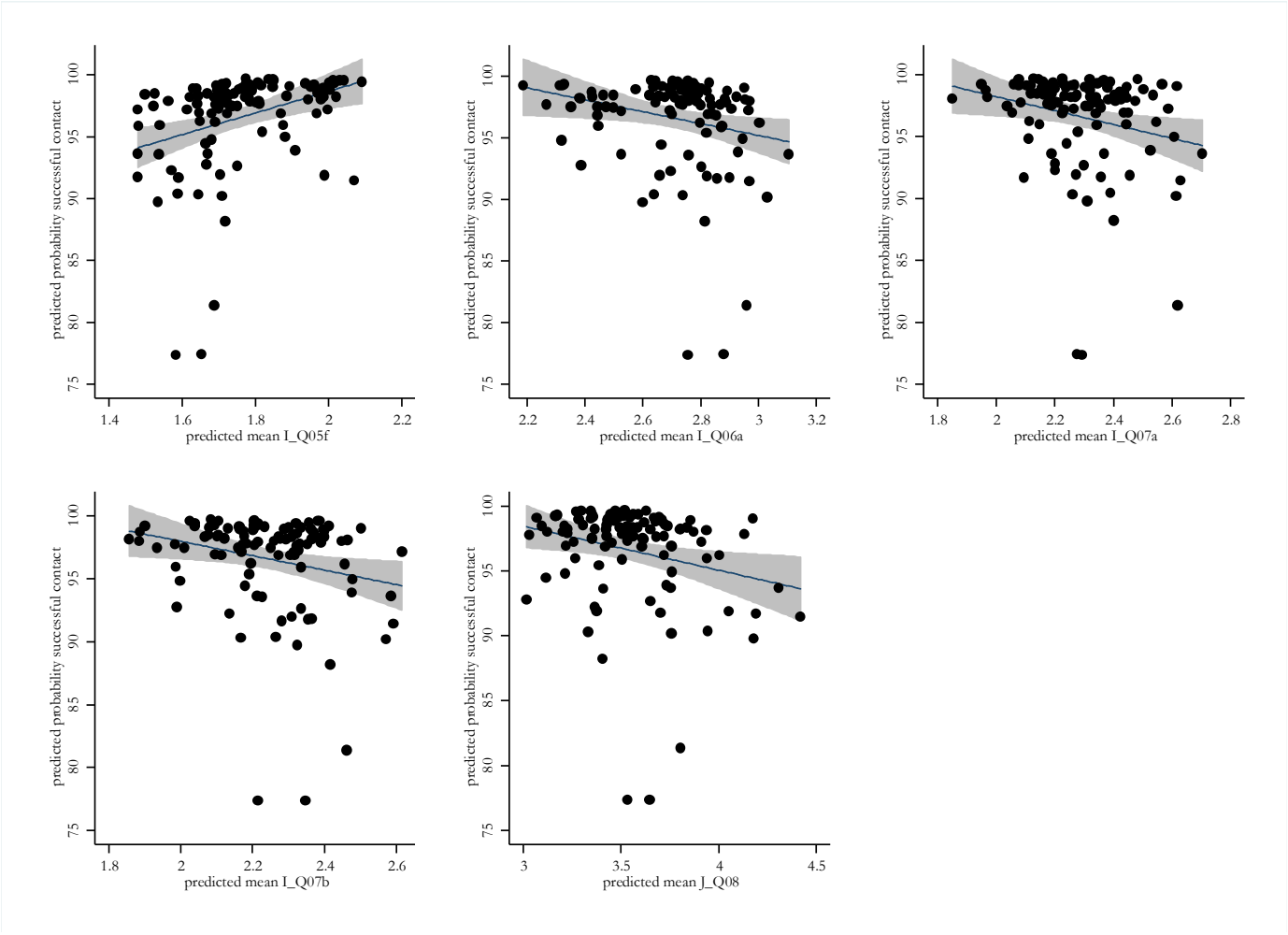
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful contact and CAPI variables of section H_Q01. Significant = $p < 0.05$. Contact: Successful contact. Predicted means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Contact: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Contact: Number of sample persons = 7,902. CAPI: Number of respondents = 4,132.

Figure H.5: Relationship between the predicted probability for successful contact and the predicted means of the CAPI variables of section I_Q04 by interviewer, PIAAC 2012



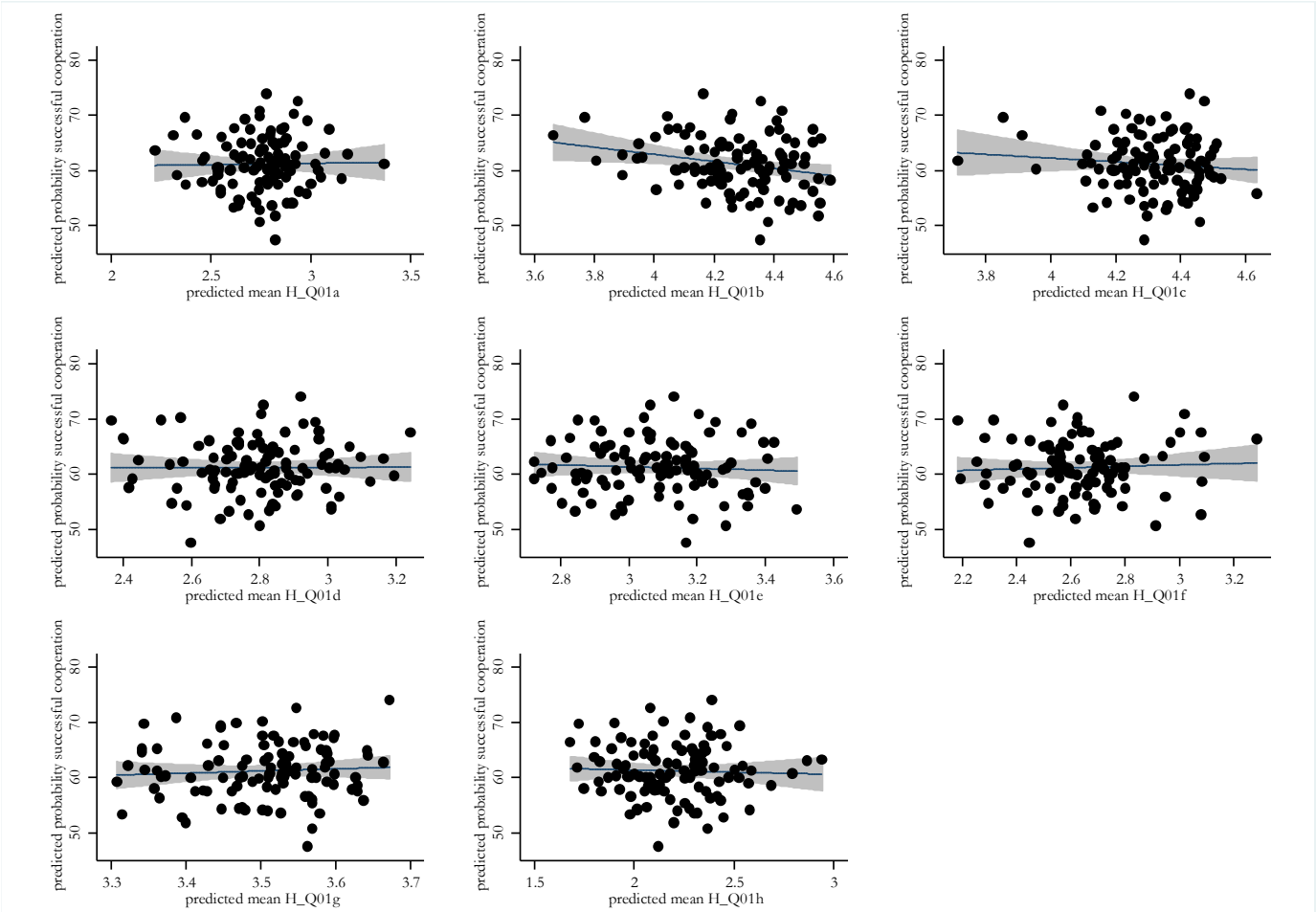
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful contact and CAPI variables of section H_Q01. Significant = $p < 0.05$. Contact: Successful contact. Predicted means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Contact: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Contact: Number of sample persons = 7,902. CAPI: Number of respondents = 4,132.

Figure H.6: Relationship between the predicted probability for successful contact and the predicted means of the CAPI variables I_Q05f, I_Q06a, I_Q07a, I_Q07b, J_Q08 by interviewer, PIAAC 2012



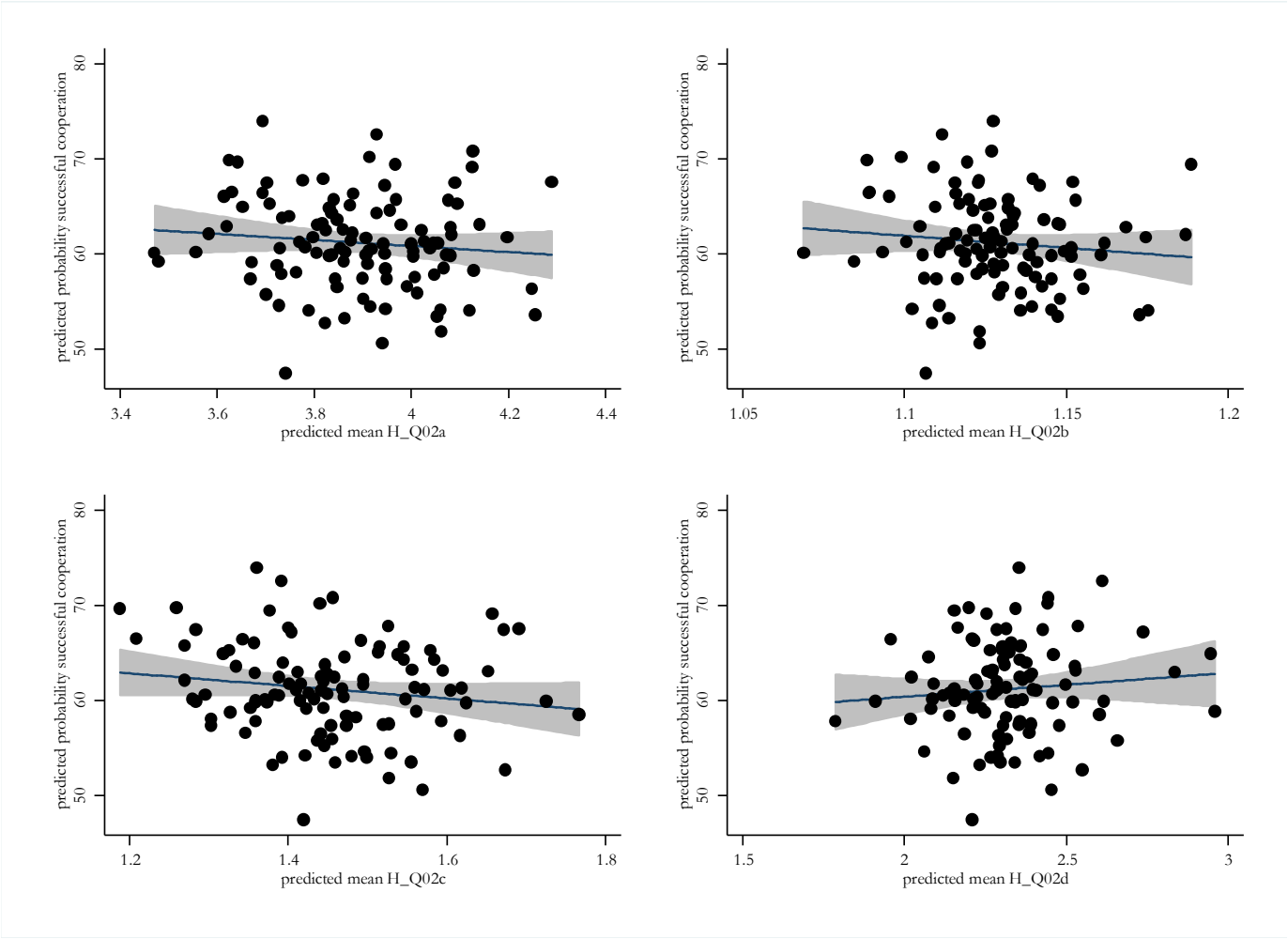
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful contact and CAPI variables of section H_Q01. Significant = $p < 0.05$. Contact: Successful contact. Predicted means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Contact: Number of interviewers = 107. Number of PSUs = 240. Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Contact: Number of sample persons = 7,902. CAPI: Number of respondents = 4,132.

Figure H.7: Relationship between the predicted probability for successful cooperation and the predicted means of the CAPI variables of section H_Q01 by interviewer, PIAAC 2012



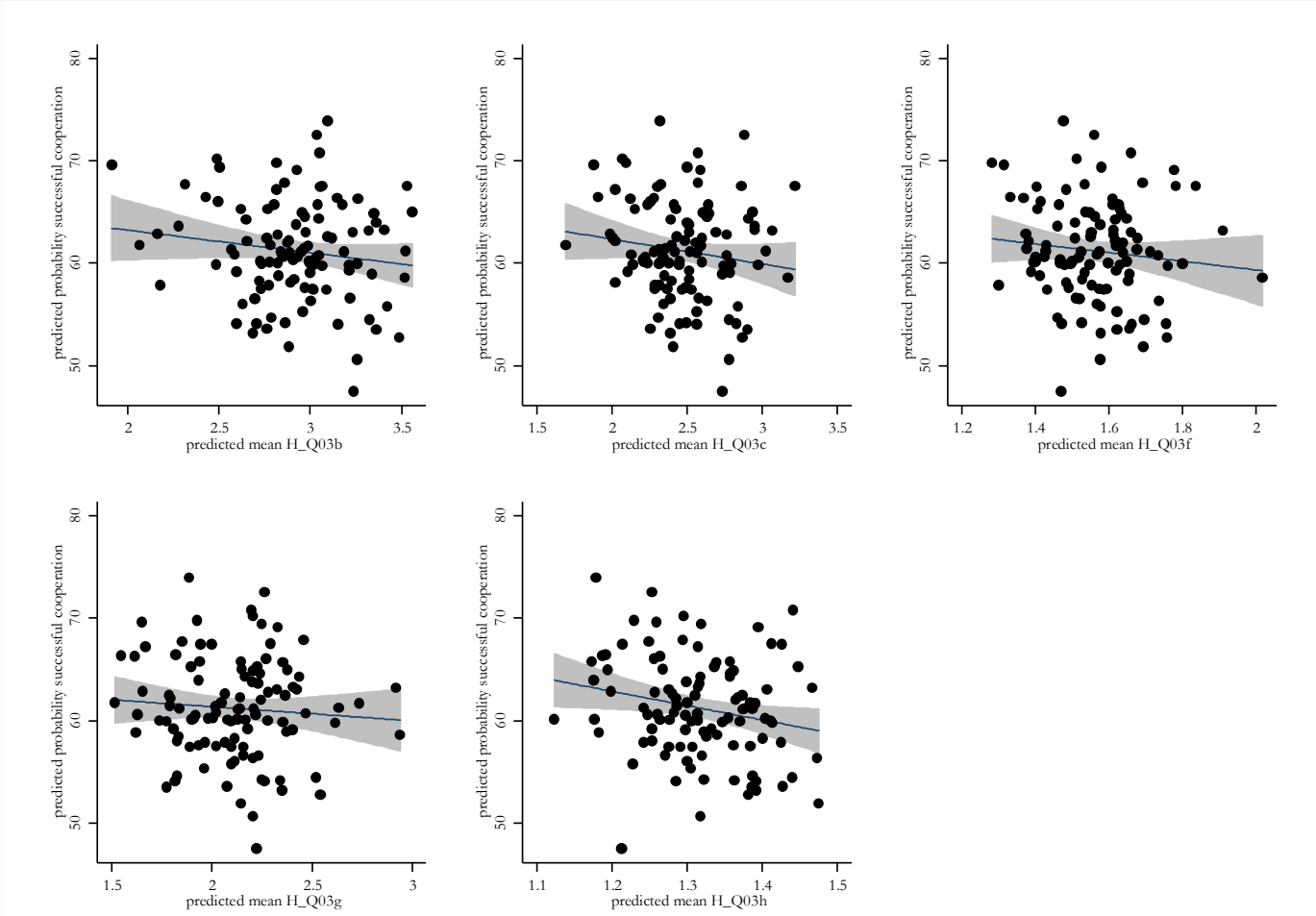
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful cooperation and CAPI variables of section H_Q01. Significant = $p < 0.05$. Cooperation: Successful cooperation. Means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Cooperation: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Cooperation: Number of sample persons = 7,450. CAPI: Number of respondents = 4,132.

Figure H.8: Relationship between the predicted probability for successful cooperation and the predicted means of the CAPI variables of section H_Q02 by interviewer, PIAAC 2012



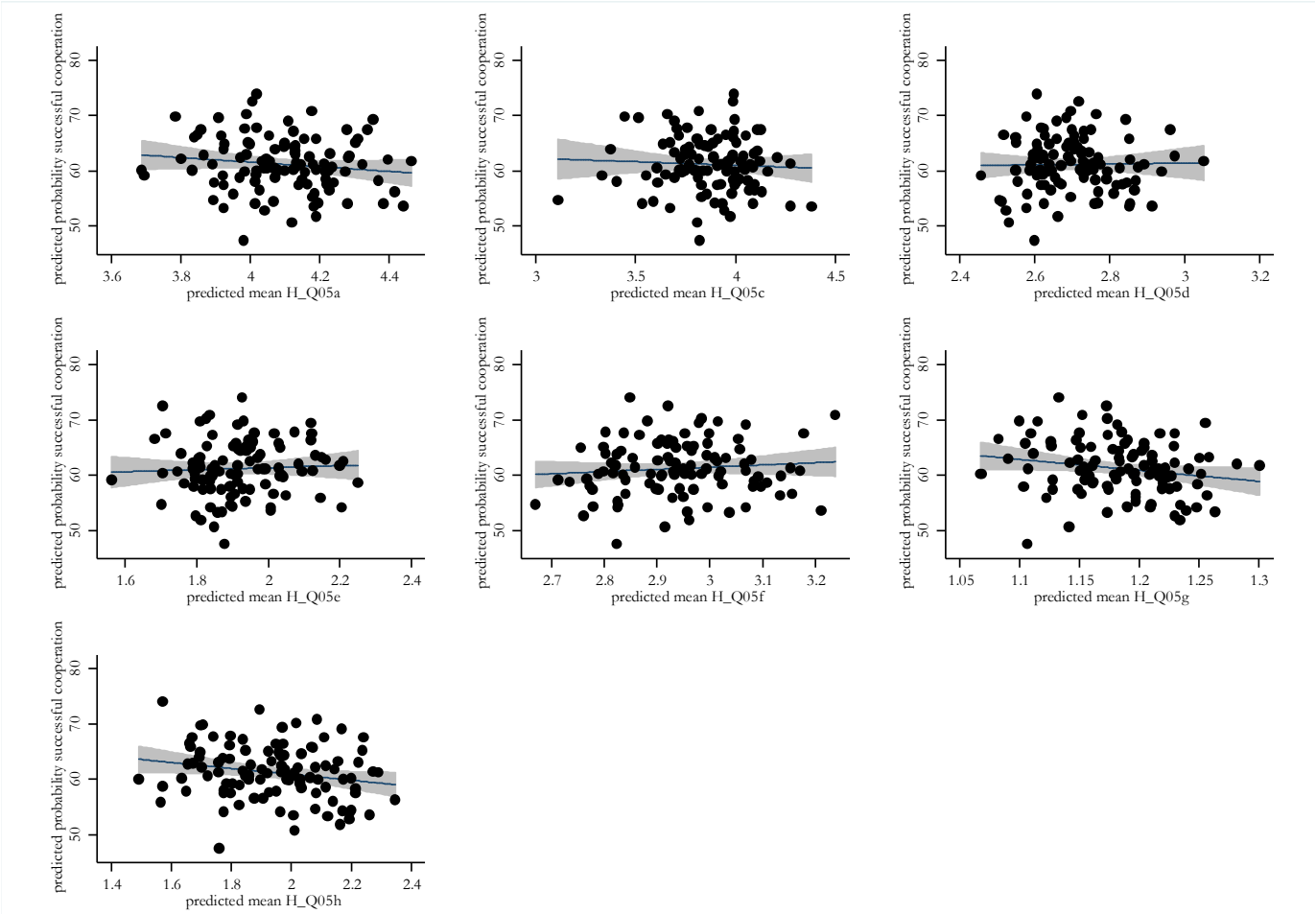
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful cooperation and CAPI variables of section H_Q01. Significant = $p < 0.05$. Cooperation: Successful cooperation. Means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Cooperation: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Cooperation: Number of sample persons = 7,450. CAPI: Number of respondents = 4,132.

Figure H.9: Relationship between the predicted probability for successful cooperation and the predicted means of the CAPI variables of section H_Q03 by interviewer, PIAAC 2012



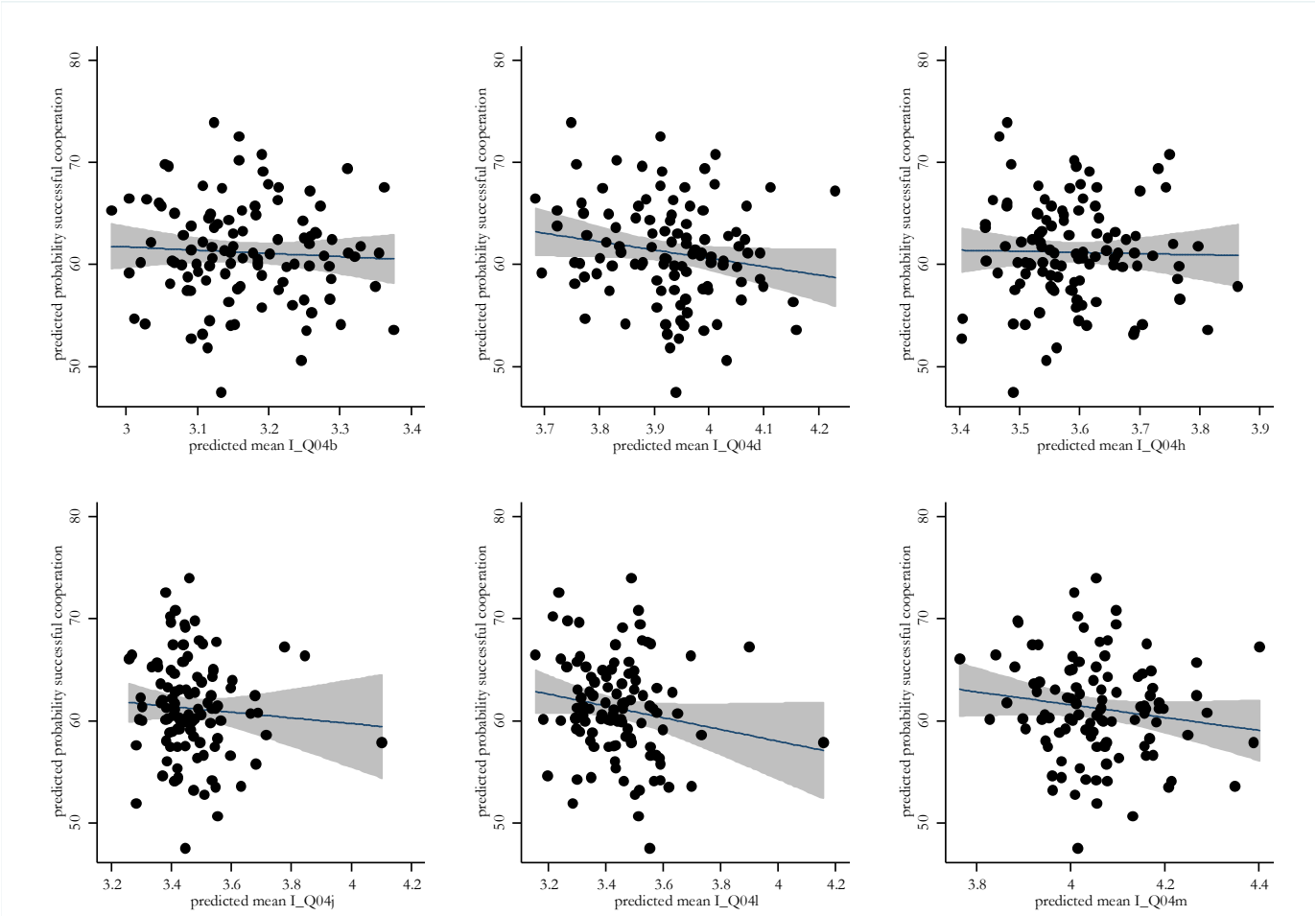
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful cooperation and CAPI variables of section H_Q01. Significant = $p < 0.05$. Cooperation: Successful cooperation. Means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Cooperation: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Cooperation: Number of sample persons = 7,450. CAPI: Number of respondents = 4,132.

Figure H.10: Relationship between the predicted probability for successful cooperation and the predicted means of the CAPI variables of section H_Q05 by interviewer, PIAAC 2012



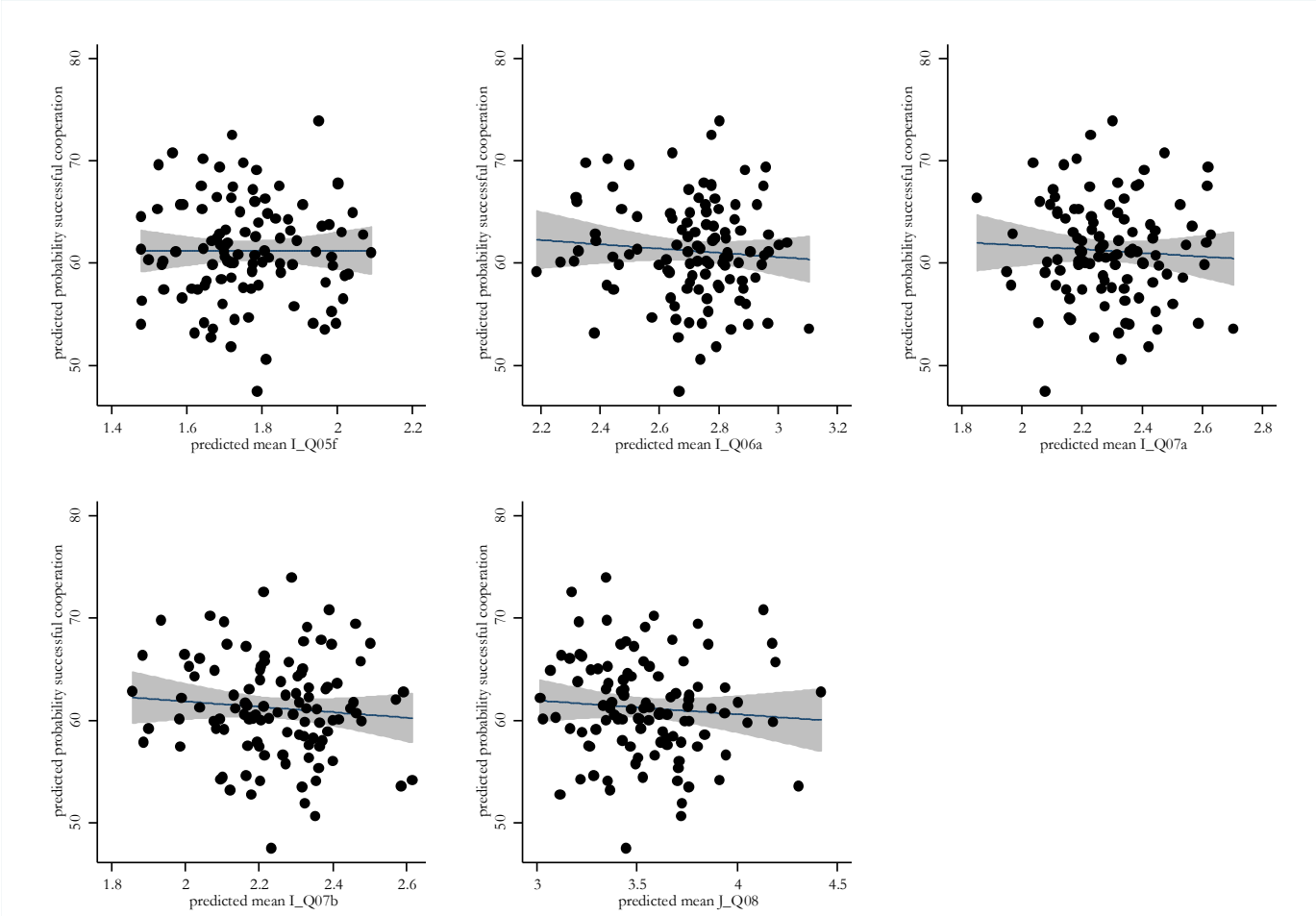
Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful cooperation and CAPI variables of section H_Q01. Significant = $p < 0.05$. Cooperation: Successful cooperation. Means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Cooperation: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Cooperation: Number of sample persons = 7,450. CAPI: Number of respondents = 4,132.

Figure H.11: Relationship between the predicted probability for successful cooperation and the predicted means of the CAPI variables of section I_Q04 by interviewer, PIAAC 2012



Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful cooperation and CAPI variables of section H_Q01. Significant = $p < 0.05$. Cooperation: Successful cooperation. Means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Cooperation: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Cooperation: Number of sample persons = 7,450. CAPI: Number of respondents = 4,132.

Figure H.12: Relationship between the predicted probability for successful cooperation and the predicted means of the CAPI variables I_Q05f, I_Q06a, I_Q07a, I_Q07b, J_Q08 by interviewer, PIAAC 2012



Notes. Scatterplot of Spearman's correlation with 95% confidence intervals for successful cooperation and CAPI variables of section H_Q01. Significant = $p < 0.05$. Cooperation: Successful cooperation. Means of CAPI variables range from 1 to 5. Question text for CAPI variables can be found in Appendix F. Number of interviewers = 107. Number of PSUs = 240. Cooperation: Controlled for sample composition characteristics (sample persons' age, gender, nationality, PSU size, region, share of Germans, share of single person households, and unemployment rate). CAPI: All models control for sample composition characteristics (respondents' age, gender, nationality, education, work status, health status, social background, PSU size, and region). Cooperation: Number of sample persons = 7,450. CAPI: Number of respondents = 4,132.