methods, data, analyses | Vol. 8(2), 2014, pp. 199-222 DOI: 10.12758/mda.2014.008

Interviewer Behavior and Interviewer Characteristics in PIAAC Germany

Daniela Ackermann-Piek[1,2] *& Natascha Massing*[2]

[1] *University of Mannheim*
[2] *GESIS – Leibniz Institute for the Social Sciences*

Abstract

Interviewers are the first in line when it comes to data collection. Therefore, it is important that they perform their tasks diligently, so that the data they collect are comparable and that errors are minimized. This paper analyzes how interviewers conducted interviews for the *Programme for the International Assessment of Adult Competencies* (PIAAC) and which kinds of mistakes they made. We approached these questions with audio interview recordings collected during the fieldwork of PIAAC in Germany (carried out in 2011/2012), as well as with an interviewer survey conducted with the German PIAAC interviewers. First, we introduce the data and the coding scheme used to evaluate interviewers' behavior with audio recordings. Subsequently, we describe the interviewers' actual behavior with regard to standardized interviewing techniques and investigate whether interviewer characteristics are associated with data quality. Our results demonstrate that interviewers do deviate from the expected behavior in all the aspects we examined. However, we identified only few associations with interviewers' background characteristics.

Keywords: PIAAC Germany, interviewer behavior, data quality, audio recordings, interviewer survey, interviewer characteristics

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. (1992), Loosveldt (2008), Schaeffer, Dykema, and Maynard (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, Kreuter, & Eckman, 2012), nonresponse (e.g., Blom, de Leeuw, & Hox, 2011; Durrant, D'Addio & Steele, 2013; Jäckle, Lynn, Sinibaldi, & Tipping, 2013), measurement (Durrant, Groves, Staetsky, & Steele, 2010; Rice, 1929), and coding or editing of survey responses (e.g., Campanelli, Thompson, Moon, & Staples, 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.[1]

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 expected a) to read aloud questions, as well as instructions, as they are scripted, b) provide adequate

1 For more information regarding nonresponse in PIAAC Germany, see Helmschrott and Martin (in this volume).

Direct correspondence to
Daniela Ackermann-Piek, University of Mannheim, 68131 Mannheim, Germany
E-mail: ackermann-piek@uni-mannheim.de

Acknowledgment: We thank GESIS – Leibniz-Institute for the Social Sciences for its support. Furthermore, we thank the funders of PIAAC in Germany: The Federal Ministry of Education and Research, with the participation of the Federal Ministry of Labor and Social Affairs. We also thank Prof. Annelies Blom and the reviewers for many valuable comments and suggestions as well as Jana Kaiser for proof reading. Finally, we thank the German survey agency TNS Infratest for its cooperation in administrating the interviewer survey of PIAAC and all the interviewers who participated in the PIAAC interviewer survey.

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.[2] 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. (2009) 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 are 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 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.

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[3] and then describe both data sources in more detail.

---

2 This relationship is usually linear (e.g. Jäckle et al., 2013) or, rarely, U-shaped (Singer, Frankel, & Glassman, 1983)

3 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, 2013) and the German PIAAC technical report (Zabal et al., 2014).

2.1 PIAAC Germany and the Role of Interviewers

PIAAC is an international survey, initiated by the OECD (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.[4] 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 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.[5]

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 were of high quality, specific and comprehensive technical standards and guidelines were defined by the international Consortium (OECD 2010) 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.[6] 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

4 This included two main fieldwork phases as well as several re-issue phases.

5 For a definition of a completed case in PIAAC, see OECD (2010).

6 All standards and guidelines related to interviewers are described in detail in Massing, Ackermann, Martin, Zabal, and Rammstedt (2013).

questionnaire that was collected and 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; Couper & Groves, 1992; Fowler & Mangione, 1990; Japec, 2008). Furthermore, German PIAAC interviewers were carefully selected.[7]

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.

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, Sturgis, & Williams, 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 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.[8] 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 1). These cases could not be matched

7 The selection criteria are described in detail in Zabal et al. (2014).

8 The source questionnaire is presented in Blom and Korbmacher (2013).

Table 1 Overview of the Interviewer Questionnaire

	n	Percent
Interviewer received questionnaire	128	100.0
Interviewer returned questionnaire	115	89.8
Questionnaire contained interviewer ID	100	78.1

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

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.

A summary of the interviewers' background characteristics, collected through the interviewer survey, is provided in Table 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.

Table 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
	46 – 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
	> 40 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.

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 2010). 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 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 (for an overview see 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[9] 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.[10] 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.

---

9 Coding the background questionnaire took about one hour per recording and was conducted directly from the recordings, using the software Audacity (Mazzoni & Dannenberg, 2012).

10 The PIAAC technical standards and guidelines only required this part of the interview to be reviewed via recordings.

Table 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 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.[11] We will explain these codes by using the example of a question wording,

11 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.

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

provided in Figure 2, to illustrate deviations from standardized interviewing techniques.

We coded each single incidence of an incorrectly skipped question as incorrect interviewer behavior. With respect to the question wording provided in Figure 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 were 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 the error introduced by the interviewer and asked to go back, to change the information that applied to her situation.

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 2 to explain the differences. For the wording presented in Figure 2, the interviewer might simply leave out the

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 2 Example of a question from the PIAAC background questionnaire

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 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 incor-

rect 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 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.

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 calculated 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 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

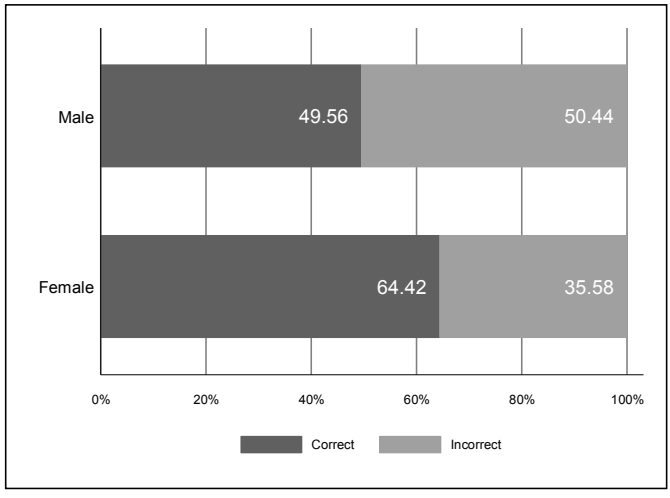
Note. Data based on 107 interviewers and 217 recordings.

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 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.

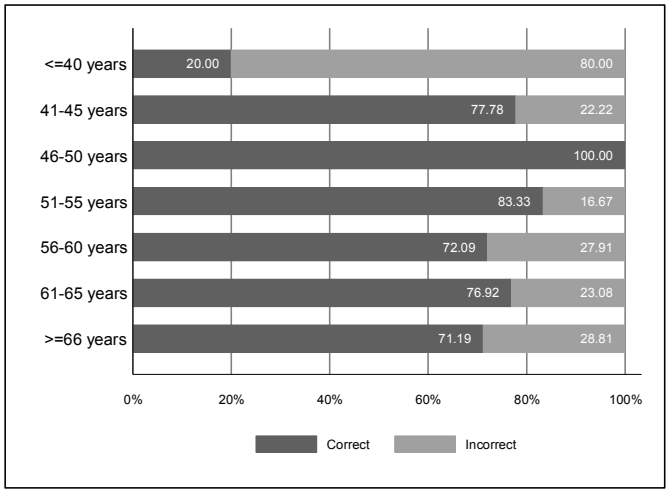
Crossing admin I with interviewers' characteristics revealed that there were significantly fewer mistakes in recording the date of the interview, the interviewer ID, as well as the respondent ID in interviews conducted by female interviewers, compared to interviews conducted by their male colleagues (Figure 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, $p = 0.048$).

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 4). However, we found significantly more mistakes



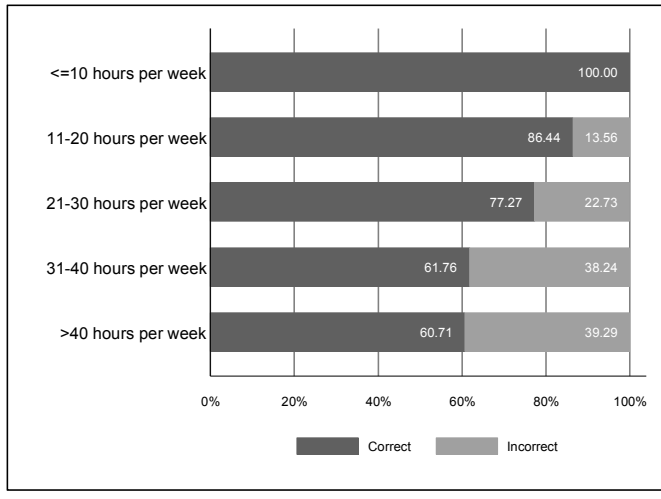
Notes. $\chi^2 = 4.8742$. $p = 0.027$. Data based on 107 interviewers and 217 recordings.

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

in interviews conducted by interviewers with longer working hours per week (Figure 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 = 0.5882, $p = 0.001$).

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). 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.[12]

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 6). Examples of how interviewers reworded

12 Results available from corresponding author upon request.

Table 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.

Table 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.

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.[13]

13 Results available from corresponding author upon request.

Table 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.

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 7). No significant association was found for any of the five tested interviewer characteristics.[14]

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 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[15] – 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

14 Results available from corresponding author upon request.

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

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 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 are 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 homoge-

neous. 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 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.

References

Beullens, K., Loosveldt, G., & Dries, T. (2013). *Explaining interviewer effects on interviewer fieldwork performance indicators*. Paper presented at the 5th Annual 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.

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

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

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.

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., 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: John Wiley and Sons.

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology* (Vol. 17, pp. 389-437). San Francisco, CA: Jossey-Bass Publishers.

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

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

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

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

Gfroerer, J., Eyerman, J., & Chromy, J. (2002). *Redesigning an ongoing National 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., Fowler, F. J. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey interviewing*. 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

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 prevention* (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.

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. (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.

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.

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 prevention* (pp. 117-130). Frankfurt am Main: Peter Lang.

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.

OECD. (2010). PIAAC technical standards and guidelines (December 2010). Retrieved March 2014 from http://www.oecd.org/site/piaac/surveyofadultskills.htm.

OECD. (2013). Technical report of the Survey of Adult Skills (PIAAC). Retrieved March 2014 from http://www.oecd.org/site/piaac/surveyofadultskills.htm.

OECD. (2014). Retrieved November 2014 from http://www.oecd.org/site/piaac/.

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

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

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

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., & 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.

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.

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

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

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

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