Interviewer Training Guidelines of Multinational Survey Programs: A Total Survey Error Perspective

Ackermann-Piek, Daniela; Silber, Henning; Daikeler, Jessica; Martin, Silke; Edwards, Brad

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

methods, data, analyses | Vol. 14(1), 2020, pp. 35-60 DOI: 10.12758/mda.2020.01

Interviewer Training Guidelines of Multinational Survey Programs: A Total Survey Error Perspective

Daniela Ackermann-Piek[1], *Henning Silber*[1], *Jessica Daikeler*[1], *Silke Martin*[1] *& Brad Edwards*[2]

[1] *GESIS – Leibniz Institute for the Social Sciences, Germany*

[2] *Westat, Rockville, Maryland, USA*

Abstract

Typically, interviewer training is implemented in order to minimize interviewer effects and ensure that interviewers are well prepared to administer the survey. Leading professional associations in the survey research landscape recommend the standardized implementation of interviewer training. Some large-scale multinational survey programs have produced their own training guidelines to ensure a comparable level of quality in the implementation of training across participating countries. However, the length, content, and methodology of interviewer training guidelines are very heterogeneous. In this paper, we provide a comparative overview of general and study-specific interviewer training guidelines of three multinational survey programs (ESS, PIAAC, SHARE). Using total survey error (TSE) as a conceptual framework, we map the general and study-specific training guidelines of the three multinational survey programs to components of the TSE to determine how they target the reduction of interviewer effects. Our results reveal that unit nonresponse error is covered by all guidelines; measurement error is covered by most guidelines; and coverage error, sampling error, and processing error are addressed either not at all or sparsely. We conclude, for example, that these guidelines could be an excellent starting point for new – small as well as large-scale – surveys to design their interviewer training, and that interviewer training guidelines should be made publicly available in order to provide a high level of transparency, thus enabling survey programs to learn from each other.

Keywords: interviewer training guidelines, interviewer effects, multinational survey programs, total survey error

Concerns about interviewer effects in interviewer-mediated surveys have accompanied generations of survey researchers. Most of the literature on interviewer effects focuses on the description and explanation of these effects after data collection (West & Blom, 2017). However, in order to ensure that interviewer-administered surveys produce high-quality data, it is essential that measures be taken to prevent, or minimize, interviewer effects. One such measure is the implementation of standardized interviewer training. In addition, interviewer training is typically used to ensure that interviewers are well prepared to adequately perform all the tasks they have during the survey implementation.

To date, findings on the effects of interviewer training on data quality are quite heterogeneous. Although most studies have shown large positive effects of interviewer training (e.g., Benson & Powell, 2015; Billiet & Loosveldt, 1988; Fowler Jr & Mangione, 1986; Fowler Jr., 1991; Mayer & O'Brien, 2001), some have found only small positive effects (e.g., Cantor, Allen, Schneider, Hagerty-Heller, & Yuan, 2004; McConaghy & Carey, 2005), and others have failed to identify any significant positive effects (e.g., Schnell & Trappmann, 2006). One reason for the heterogeneity of the findings on the effects of interviewer training may be that the training programs themselves are very heterogeneous (for a short overview, see Daikeler & Bosnjak, forthcoming; Lessler, Eyerman, & Wang, 2008). Thus, it is not surprising that general interviewer training have been proposed from multiple sources, all recommending a careful planning and standardized implementation of interviewer training when conducting large-scale interviewer-administered surveys (American Association for Public Opinion Research [AAPOR], n.d.; Alcser, Clemens, Holland, Guyer, & Hu, 2016; Daikeler, Silber, Bosnjak, Zabal, & Martin, 2017; Fowler Jr. & Mangione, 1990; International Organization for Standardization [ISO], 2012; Lessler et al., 2008). Interviewer training is part of the training concepts of large-scale interviewer-administered survey programs, such as the European Social Survey (ESS), the Survey of Health, Ageing and Retirement in Europe (SHARE), and the Programme for the International Assessment of Adult Competencies (PIAAC), and of the vast majority of large-scale national interviewer-administered surveys, such as the German Socio-Economic Panel (SOEP) and the U.S. National Health

Acknowledgments

The authors thank the participants of the workshop „Dashboard & Field Monitoring" held in Mannheim, 2018 for their thoughts which encouraged this research idea. We are grateful to Sabine Friedel and Verena Halbherr for helpful details about interviewer training in the ESS and SHARE and detailed cross-checks of the manuscript.

Direct correspondence to

Daniela Ackermann-Piek, GESIS – Leibniz Institute for the Social Sciences,
Quadrat B2 1, 68159 Mannheim, Germany
E-mail: daniela.ackermann-piek@gesis.org

Interview Survey (NHIS). More specifically, in line with the general guidelines issued by leading professional associations in the area of survey research, the interviewer training concepts of national and multinational survey programs include recommendations for study-specific training for inexperienced and experienced interviewers, as well as brief sections on general interviewer training for inexperienced interviewers.

However, what is lacking in the literature is a structured comparison of the content of the various interviewer training guidelines of survey programs or training concepts of large surveys – using a theoretical framework. The present article aims to fill this gap by providing a comparative overview of the extent to which the content of training guidelines of the ESS, PIAAC, and SHARE (Börsch-Supan & Jürgens, 2005; ESS, 2016a, 2016b, Malter & Börsch-Supan, 2017; Organziation for Economic Co-operation and Development [OECD], 2011, 2013) are integrated into the conceptual *total survey error* framework (TSE; Biemer, 2010; Groves et al., 2009; Pennell, Hibben, Lyberg, Mohler, & Worku, 2017; T. W. Smith, 2019). Additionally, we investigate the extent to which the individual components of the TSE are addressed in these guidelines, and we make suggestions for improvements. Specifically, we focus on the topics of interviewer training specified in training guidelines on an international level rather than on interviewer training in specific countries with detailed training content.

The TSE Framework and the Literature on Interviewer Training

The TSE is a theoretical concept that describes statistical error properties of survey estimates, systematically structured along different error sources (Biemer, 2010; Groves & Lyberg, 2010). Error sources are assigned to each step in the survey life cycle, typically along two dimensions each, either sampling error and nonsampling error (Biemer, 2010) or measurement and representativeness (Groves et al., 2009). As our aim is to compare interviewer training of multinational survey programs along the TSE, we follow the approach of Pennell et al. (2017), who adopted the TSE typology for multinational, multiregional, and multicultural surveys (3MC). Pennell et al. (2017) provide a TSE model which combines the complexity in designing and implementing 3MC surveys with the overall aim to minimize comparison error (T. W. Smith, 2011). Following the approach of Groves et al. (2009), the authors structure their model along the two dimensions measurement and representativeness.

Following Groves et al. (2009) and Pennell et al. (2017), the representation dimension of the TSE includes *coverage error*, *sampling error*, *nonresponse error*, and *adjustment error* and the measurement dimension includes *validity*,

measurement error, and *processing error*. *Coverage error* refers to problems of a not perfectly covered target population in the sampling frame. *Sampling error* occurs because only a sample is observed instead of the entire target population. With regard to sampling error, either bias (members of the sampling frame are systematically excluded from selection) or variance (different sets of sample frame elements are selected and each set can have different values in the survey statistic) can occur. *Nonresponse error* occurs when selected sample members that respond the survey request systematically differ from those who do not respond the survey request. After data collection, post-survey adjustments are typically used to correct for representation errors occurred earlier in the process. However, when post-survey adjustments fail to capture each case of misrepresentation in the sample, *adjustment error* occurs. The error components of the measurement dimension of the TSE are associated with errors in the survey instruments and the question-response process. The first error component of the measurement dimension of the TSE, *validity*, reflects an error that describes that the theoretical construct is not optimally reflected in the measure. The next error component, *measurement error,* occurs when the response given by a respondent differs from the true response. Finally, *processing error* reflects the incorrect transfer of responses to data storage during capturing, coding, or editing of data. All errors described in the TSE may result in biased survey estimates of substantive survey variables. Thus, the aim of survey operations is to minimize the errors under the given time and cost constraints to maximize the survey quality (Schouten, Peytchev, & Wagner, 2017).

The TSE framework is regularly used to describe and structure interviewer errors in interviewer-administered surveys (for an overview, see West & Blom, 2017). As interviewers have many tasks when administering a survey, such as contacting sample units, gaining their cooperation, asking questions, and recording answers (e.g., Groves et al., 1992; Loosveldt, 2008; Schaeffer, Dykema, & Maynard, 2010), they can – intentionally or unintentionally – affect a large number of steps in the survey process. In other words, they can be the sources of multiple survey errors. Because interviewer training is organized along interviewer tasks, the TSE framework can be used to structure training content when reviewing general and study-specific interviewer training concepts of multinational survey programs.

A review of the literature on the effects of interviewer training on survey data quality from a TSE perspective reveals that most studies focus either on nonresponse error or measurement error. Studies on the effects of interviewer training on nonresponse error, for example, show a positive effect of refusal avoidance training on reducing nonresponse (Cantor et al., 2004; Daikeler & Bosnjak, forthcoming; Durand, Gagnon, Doucet, & Lacourse, 2006; Hubal & Day, 2006; Mayer & O'Brien, 2001; O'Brien, Mayer, Groves, & O'Neill, 2002). However, not all studies have found overall positive effects of interviewer training on nonresponse (Schnell & Trappmann, 2006). Studies with a focus on the effects of interviewer training

on several quality indicators related to measurement error, for example, have identified a positive effect of interviewer training on the application of standardized interviewing techniques (Billiet & Loosveldt, 1988; Dahlhamer, Cynamon, Gentleman, Piani, & Weiler, 2010; Fowler Jr., 1991), the reduction of item nonresponse (Billiet & Loosveldt, 1988; Daikeler & Bosnjak, forthcoming), and the application of appropriate probing techniques (Billiet & Loosveldt, 1988; Daikeler & Bosnjak, forthcoming). However, Groves (2005) noted that the literature left open the question whether interviewer training effectively reduces measurement error. And finally, the effects of interviewer training on coverage error, sampling error, and processing error have been addressed only occasionally in the literature (Eckman & Kreuter, 2011; Guest, 1954).

Methodology and Resources

The present study compares interviewer training concepts of three large-scale multinational survey programs in the social sciences, namely, the ESS (ESS, 2016a, 2016b, 2018), PIAAC (OECD, 2011, 2013), and SHARE (Börsch-Supan & Jürges, 2005; Malter & Börsch-Supan, 2017). In most cases, only multinational survey programs have the funds to develop detailed interviewer training guidelines and implement interviewer training accordingly. These programs need predefined detailed specifications for participating countries, because such programs are imperative to ensure a harmonized data collection process across countries, which is a prerequisite for obtaining high-quality data (Pennell, Harkness, Levenstein, & Quaglia, 2010; Survey Research Center at the University of Michigan, 2016). Ensuring harmonization across countries also applies to the training of interviewers.

Overall, the ESS, PIAAC and SHARE fulfilled the following selection criteria: First, all three are administered by interviewers face-to-face. Second, they have participants from many European countries. Third, in all three cases detailed documents were publicly available that contained information on the survey programs' interviewer training guidelines. Very often this information is confidential and not accessible.

The ESS is a cross-sectional survey of attitudes, beliefs, and behaviors that is conducted every two years. SHARE is a longitudinal survey on health, ageing and retirement. SHARE is also conducted every two years. For these two survey programs, we selected the specifications and characteristics from the last round for which interviewer training guidelines are available (ESS Round 8, 2016; SHARE Wave 6, 2015). PIAAC is a multi-cycle program for the assessment of basic adult

competencies; a cross-sectional "Survey of Adult Skills" is carried out every 10 years. We used the training material from PIAAC Cycle 1, Round 1, 2011/2012[1].

Training concepts relate to training guidelines on an international level rather than to interviewer training in specific countries. Comparing the implementation of interviewer training in the various participating countries in detail would be another important research question. Also, we focus our research on general as well as survey specific interviewer training content provided by the three multinational survey programs.

Interviewer Tasks within ESS, PIAAC, and SHARE

When comparing interviewer training guidelines of different survey programs, it is important to take interviewers' tasks in the surveys and the resulting complexity of their roles into account. First of all, this refers to the target population, as these are the persons with whom interviewers interact. The target population of the ESS and PIAAC is quite similar and refer to the general population aged either 15 years or older (ESS) or between 16 and 65 years (PIAAC). The target population in SHARE also refers to the general population, however, only to persons who are 50 years or older at the time of sampling. In addition, in SHARE, spouses or partners of the sampled person are interviewed as well, if applicable.

Another interviewers' task for all three survey programs was the administration of the core questionnaire face-to-face using computer-assisted personal interviewing (CAPI). In addition, PIAAC and SHARE interviewers had to perform additional tasks. In PIAAC, interviewers had to administer a cognitive assessment where respondents worked independently on a number of tasks on the interviewer's laptop or in a paper booklet under the supervision of the interviewer. For this purpose, interviewers switched from their traditional role of asking questions and took on a passive, monitoring role, adapting their behavior accordingly. If the respondent opted for the paper-based cognitive assessment, the interviewer additionally had to score some items for routing purposes.[2] In SHARE, interviewers administered a self-completion paper questionnaire to the respondents in some countries. As the target population in SHARE consisted of elderly persons, interviewers had to be able to interact with this special population. A special and new task for some SHARE interviewers was to collect biomarkers from respondents and conduct physical tests (e.g., measuring blood pressure). The average interview duration was

1 For a more detailed overview of the specifications and characteristics of the ESS, PIAAC, and SHARE across all participating countries, see the Appendix.

2 Scoring means that the interviewer has to determine a value (correct or incorrect) for each response to a number of selected items based on scoring guidelines (Zabal et al., 2014, p. 104).

60 minutes for the ESS (Round 8), whereas SHARE and PIAAC had longer average durations (80 minutes and 90 minutes, respectively).

Interviewer Training Concepts of ESS, PIAAC, and SHARE

The ESS specifications for the countries distinguish between two types of interviewer preparation: training and briefing. ESS interviewers are expected to have previous face-to-face interviewing experience and to be trained in effective doorstep interaction, standardized interviewing techniques, and general interviewer behavior before administering the survey instrument. In each round of the ESS, experienced interviewers receive a briefing, whereas inexperienced interviewers should additionally undergo general interviewer training prior to the briefing.

With respect to interviewer training in PIAAC, several features can be highlighted: (1) the extensive interviewer training package (including, e.g., fully scripted training sessions); (2) the train-the-trainer session prior to national interviewer training in which the training staff is introduced to the scripts and interview materials; (3) the close monitoring by the international consortium of the implementation of the country-based interviewer training. As PIAAC Round 1, Cycle 1 also included a field trial in which all aspects of the survey – including interviewer-related topics – were tested, the interviewer training sessions for the main study were shortened depending on the performance of the interviewers in the field trial.

For SHARE, the survey programs' multiplier approach to interviewer training can be highlighted: a centralized train-the-trainer program is conducted to facilitate decentralized interviewer training in the participating countries. Moreover, all interviewers are expected to have extensive general face-to-face interviewing experience and to have received in-person general interviewer training prior to undergoing study-specific training.

Standards for the Implementation of Interviewer Training within ESS, PIAAC, and SHARE

The extent to which the implementation of interviewer training is specified differs considerably across the three multinational survey programs (for an overview, see Table 1). As a first impression, when counting the number of pages in the overall survey specifications which are provided to the participating countries for the respective survey[3], it becomes obvious that the specifications for the ESS (65 pages)

3 The survey specifications are provided to the country contact and are intended to be used as orientation for the implementations of the survey. Typically, these specifications are not handed out to the interviewers.

Table 1 ESS, PIAAC, and SHARE Standards for the Implementation of Interviewer Training and Desired Interviewer Characteristics

	ESS 2016 (Round 8)[1]	PIAAC (Cycle 1, Round 1)[2]	SHARE (Wave 6)[3]
General concept	Interviewer experience important: GIT for inexperienced interviewers and reduced training for experienced interviewers	Extensive material provided for standardized training sessions for all interviewers Training for pretest and main study (the latter can be reduced depending on experience and performance in pretest)	Multiplier approach: Centralized TTT program to facilitate decentralized national training Interviewers expected to be experienced
Length of survey specifications (number of pages)	65	199	542
Material for trainers	NC manual Series of pre-structured slides with movie clips and related material (incl. guidelines for training, scripted practice interview)	Technical Standards and Guidelines manual for NPM teams (plus comprehensive interviewer training material on planning and implementation, incl. PowerPoint slides, training scripts)	Interviewer project manual Facilitator guide (incl. PowerPoint slides, training scripts) CD-based training on gaining respondent cooperation (incl. training videos) Computer-based tutorial (using SHARE CMS) Training evaluation protocols Sample management monitoring Specific guidelines (to be followed by interviewers to ensure cross-national comparability)

	ESS 2016 (Round 8)[1]	PIAAC (Cycle 1, Round 1)[2]	SHARE (Wave 6)[3]
Material for interviewers	Interviewer manual	Home-study material, incl. written exercises Interviewer manual	Interviewer manual (incl. instructions for blood collection with a short video)
Training facilities and equipment	U-shaped chair setup, technical equipment	Adequate space (separate rooms, each with a lead trainer, technical equipment)	---
Training staff	---	Competent and experienced lead trainers Assistant trainers Technical support (attended TTT sessions)	International trainer trains the national trainers
Supervisory staff	---	Regional supervisors and field managers who attended supervisor training session	---
Training delivered to:	Interviewers	Interviewers Trainers Supervisors Field managers	National trainer (from the survey agency) Country team leader Operator Interviewers
Interviewer experience	Experienced, trained in general interviewer tasks F2F Reduced training (“briefing”) for all interviewers, general training for new interviewers	New interviewers need GIT/training in CAI prior to/during study-specific training	New interviewers need GIT Extensive general face-to-face interviewing experience

	ESS 2016 (Round 8)[1]	PIAAC (Cycle 1, Round 1)[2]	SHARE (Wave 6)[3]
Mode of training	In-person	In-person	In-person
Trainee group size	---	15–20	---
Training length	6 hrs	Pretest: 36 hrs all interviewers Main study: 15 hrs (pretest interviewers with good reviews); 30 hrs (pretest interviewers with less than favorable reviews or interviewers with experience on other surveys), 36 hrs inexperienced interviewers	TTT pilot: 2 days TTT pretest 1.5 days TTT main: 1 days national training: 2 days
Training scheduling	One month before fieldwork starts	No earlier than two weeks before, but preferably the week immediately before the scheduled start of data collection	TTT pilot: --- TTT pretest: --- TTT main: --- National training: Dec. 2014 – Feb. 2015
Training methods	Presentation Mock interviews Video clips (scenarios) Role play Interactive discussion	Lectures Scripted mock interviews Roundtable exercises Role play Practice interviews Training in multiple languages (if applicable)	Mock interviews Question cards Question-answer session Practice for blood collection
Training evaluation	---	Interviewers' performance (during training) Supervisors' ability to perform responsibilities	Local survey agency trainers evaluate TTT in writing, preparation for local training sessions, materials, and possible improvements

	ESS 2016 (Round 8)[1]	PIAAC (Cycle 1, Round 1)[2]	SHARE (Wave 6)[3]
Quality control	NC must oversee interviewer selection and training and make sure that in-person training takes place	Documentation of plan, scheduling, material, implementation, and evaluation of national training Training report Retraining and attrition training during data collection (if applicable)	Trainer certification Interviewer certification

Note. Survey programs in alphabetical order; information retrieved from general project guidelines or project specifications (see references below); CAI = computer-assisted interviewing; CMS = case management system; ESS = European Social Survey; F2F = face-to-face; GIT = general interviewer training; NC = National Coordinator; NPM = National Project Management; PIAAC = Programme for the International Assessment of Adult Competencies; SHARE = Survey of Health, Ageing and Retirement in Europe; TTT = train-the-trainers session; --- = no information included in the specifications.

[1] ESS (2016b); ESS (2016a); ESS (2018).
[2] OECD (2011); OECD (2013, Chapter 10.4).
[3] Börsch-Supan and Jürges (2005); Malter, Schuller, and Börsch-Supan (2016).

are not as extensive as those for PIAAC (199 pages) and SHARE (542 pages). This is not surprising when one considers the additional, non-standard, tasks that interviewers in PIAAC and SHARE must perform. In addition, experienced ESS interviewers typically only receive a reduced version of the interviewer training (referred to as "interviewer briefing").

The interviewer training guidelines of all three survey programs specify that the training should be conducted in-person. All guidelines require measures for controlling the quality of the training (e.g., review of interviewer selection and training report by country). In contrast, other topics are not covered by the survey specifications of all three survey programs. For example, the ESS and PIAAC specify that training should be scheduled to take place shortly before the start of data collection, whereas this is not addressed in the SHARE survey specifications. Moreover, the ESS and SHARE specifications related to interviewer training do not address training of supervisory staff, training-group size, or the structure of training groups, whereas PIAAC defines these aspects clearly. In addition, the SHARE specifications do not address training facilities and equipment, and the ESS specifications do not include information on the evaluation of the interviewer training. Other examples are that in the ESS, for example, the national coordinators, who are responsible for organizing the national interviewer training, are given research-based information on interviewer effects to demonstrate the positive effects of interviewer training. The SHARE specifications contain information about the national interviewer training and the train-the-trainer sessions and the PIAAC specifications emphasize the importance of quality controls and provide very detailed guidelines on the implementation of interviewer training.

Mapping of Interviewer Training Content to the TSE

In this section, we map the specifications for general interviewer training and the program-specific training content of the three multinational survey programs along the TSE framework.

General Interviewer Training for Inexperienced Interviewers

Table 2 provides an overview of training content of the interviewer training guidelines of the three multinational survey programs from a TSE perspective for general interviewer training, intended for inexperienced interviewers. Of the three survey programs, only PIAAC provides comprehensive guidelines on general interviewer

training for inexperienced interviewers, which go beyond the coverage of nonresponse and measurement error. In contrast, the ESS and SHARE interviewer training guidelines include only some examples of topics covering interviewer training for inexperienced interviewers.

Table 2 ESS, PIAAC, and SHARE General Interviewer Training Content from a TSE Perspective for Inexperienced Interviewers

	ESS[1]	PIAAC[2]	SHARE[3]
Representation			
Coverage error	---	---	---
Sampling error	---	---	---
Nonresponse error	Doorstep interaction	Gaining cooperation (incl. detailed contact and refusal conversion strategies)	Collecting process data information (incl. contact attempt and result of contact attempt)
Measurement			
Measurement error	Standardized interviewing (incl. detailed rules)	Asking questions (incl. exercises) Probing techniques	Standardized question-asking Probing
Processing error	---	Recording answers (incl. exercises)	---
Content of relevance for multiple TSE components			
	---	Introduction to survey research (incl. types of survey questions, interviewing terminology) Standards and ethics in survey research (incl. informed consent, data confidentiality, data security, exercises) Remuneration and administrative aspects Basics of computer-assisted interviewing (CAI)	---

Note. Survey programs in alphabetical order; ESS = European Social Survey; PIAAC = Programme for the International Assessment of Adult Competencies; SHARE = Survey of Health, Ageing and Retirement in Europe; --- = no information included in the guidelines.

[1] ESS (2016c); Beullens, Loosveldt, Denies, and Vandenplas (2016).

[2] OECD (2013).

[3] Börsch-Supan and Jürges (2005).

The PIAAC specifications for general interviewer training, that affect all components of the TSE, comprise an introduction, some standards and ethics in survey research, administrative aspects, and instructions on the basics of computer-assisted interviewing (CAI). With regard to nonresponse, these specifications include strategies for gaining cooperation, and with regard to measurement error, they relate to question-answering and probing techniques. In addition, processing error is covered by techniques for recording answers. Nevertheless, looking at the proposed length of the training sessions for each topic, it becomes apparent that the focus is clearly on measurement error and on related practice sessions (OECD, 2013, Chapter 10.4).

In contrast, the ESS specifications for general interviewer training cover only nonresponse error (doorstep interaction training) and measurement error (training of standardized interviewing). Similarly, the SHARE specifications for general interviewer training are quite brief and cover only nonresponse error (training in process data collection) and measurement error (standardized interviewing and probing techniques). Neither the ESS nor the SHARE specifications include information about the length of the general interviewer training.

Program-Specific Interviewer Training for Inexperienced and Experienced Interviewers

In Table 3, we map the program-specific training content of the three survey programs to the components of the TSE. Training content that is relevant for all error sources is presented at the bottom of the table. Program-specific training is intended for interviewers who have general interviewing experience or who have attended general interviewer training but are not familiar with the program-specific interview tasks.

Table 3 Program-Specific Interviewer Training Content of the ESS, PIAAC, and SHARE from a TSE Perspective for all Interviewers

	ESS[1]	PIAAC[2]	SHARE[3]
Representation			
Coverage error	Selecting respondents	Screener administration	
Sampling error	---	---	---
Nonresponse error		Screener administration	
		Locating strategy	Locating strategy
	Contact strategy (detailed)	Contact strategy	Contact strategy (detailed)
	Gaining cooperation (incl. realistic examples)	Gaining cooperation	Gaining cooperation (detailed, focus: representativeness)
	Refusal conversion	Refusal avoidance and conversion	Refusal conversion
Measurement			
Measurement error		Screener administration	
	Instrument overview (incl. practice, focus: specific questions)	Instrument administration (incl. practice)	Instrument overview (incl. practice)
			Probing
	Standardized interviewing		
			Collection of a dried blood sample (detailed)
Processing error	---	---	Coding conventions (Mental Health section)
			Recording responses

	ESS[1]	PIAAC[2]	SHARE[3]
Content of relevance for multiple TSE components			
	Introduction (incl. goal, findings previous rounds, data quality issues, data usage)	Introduction	Introduction (incl. goal, questionnaire overview)
	Logistics (incl. target response rate, *supplementary material*, *remuneration*, fieldwork procedures, etc.)		Logistics
		Review of advanced materials	
		Case management system (incl. data transmission)	Laptop overview (incl. installation check)
	Administrative tasks (contact forms)	Administrative tasks (incl. disposition codes, case folder, record of contact)	Case management system (incl. practice)
	Information for respondents prior interview (e.g., data protection; general information ESS; data confidentiality; data storage, etc.)		
	Experienced interviewers: changes since last round, comparison with other surveys	*Demo* and practice interviews (role play and/or live respondent practice)	
		Question-and-answer session (after practice)	Question-and-answer session
		Additional practice sessions	
	Quality control	Quality control and monitoring	

Note. Survey programs in alphabetical order; ESS = European Social Survey; PIAAC = Programme for the International Assessment of Adult Competencies; SHARE = Survey of Health, Ageing and Retirement in Europe; --- = no information included in the specifications; optional training content in italics.

[1] ESS (2016c); Beullens et al. (2016).

[2] OECD (2013).

[3] Börsch-Supan and Jürges (2005); Malter and Börsch-Supan (2017).

A comparison of the training content of the ESS, PIAAC, and SHARE reveals that the survey programs focus on different components of the TSE framework. However, some training content is similar for all three survey programs, for example, an overall introduction to the survey, which is relevant for multiple components of the TSE. In addition, all three survey programs offer training content on contacting, gaining cooperation, and refusal avoidance strategies, with the goal of reducing nonresponse error. To address measurement error, the interviewer training specifications of all three survey programs include sessions providing an overview of the survey instruments.

The differences between the training guidelines of the three survey programs reveal that the ESS specifications include very precise information on survey logistics (e.g., target response rate, fieldwork procedures), administrative tasks, and information that must be provided to respondents before the interview starts (e.g., data confidentiality, data storage). These topics included in the ESS specifications are relevant for all TSE components as it might affect more than one error source. Similarly, the specifications for interviewer training in PIAAC are quite detailed with respect to administrative tasks, and they additionally include a large section on practical sessions (e.g., question-and-answer sessions, demo interviews), which are also relevant for multiple error sources. In comparison, SHARE does not include detailed specifications for administrative tasks. However, in the training session on mental health, there is a large sub-section on coding conventions. Training in coding conventions aims to reduce processing error; this is not covered by the specifications of the other two survey programs. Yet, the SHARE specifications for interviewer training do not include any information on quality control and monitoring, which is covered by the ESS and PIAAC training content specifications.

Discussion

In the present paper, we aimed to review program-specific interviewer training guidelines of three multinational large-scale survey programs (ESS, PIAAC, and SHARE) using the TSE framework. Our results show that there is a clear focus on measurement error, nonresponse error, and introductory and administrative topics in the training materials. Other error sources are either covered by more general parts of the interviewer training guidelines (e.g., logistics, technical issues), which address multiple components of the TSE framework, or are rarely (e.g., processing error is covered only by the SHARE interviewer training guidelines) or not covered at all. There are several possible explanations for this. First, it is reasonable that the focus of the training reflects the actual tasks of interviewers: gaining cooperation and the administration of the question-and-answer process are among an interviewer's main tasks in almost every survey program. The tasks assigned to

interviewers vary largely across different survey projects. Thus, the involvement of interviewers related to tasks affecting coverage, sampling, and processing is not part of every survey project, as in the three cases examined here. Second, the measurement and detection of coverage, sampling, and processing errors requires a higher control effort compared to the two other error sources. And third, most surveys have experienced a dramatic decrease in response rates in recent years (Beullens, Loosveldt, Vandenplas, & Stoop, 2018), which might make the skill of gaining cooperation more salient.

When looking at the interviewer training guidelines of the three survey programs in more detail, we identified several differences between the training content of the three multinational training programs. Training content aimed at reducing nonresponse error was identified in the interviewer training guidelines of all three multinational survey programs. However, training in locating sample units is not mentioned in the ESS guidelines. Training content aimed at reducing measurement error is covered in all three guidelines: the ESS guidelines ar the only guidelines including standardized interviewing, whereas probing techniques and the collection of biomarkers or the administration of a cognitive test are included only in the SHARE or PIAAC guidelines. These differences are due, in part, to the scope of the respective studies, which obviously differs across the three surveys we have compared in the present paper. Besides, training content relating to processing error is covered only in the SHARE training guidelines. Moreover, the PIAAC interviewer training guidelines include an extensive general interviewer training agenda for inexperienced interviewers, which is only sparsely addressed by the ESS and SHARE. As PIAAC was conducted for the first time in the participating countries and – as general interviewer training forms the basis for additional project-specific training – we suspect that the interviewer training guidelines aimed to ensure that all interviewers working for PIAAC were at a similar level of knowledge. However, all three multinational survey programs require that only interviewers who are trained in general interviewer tasks are employed for the survey. This is in line with Pennell et al. (2017), as interviewers with more interviewing experience are likely to minimize comparison error.

An important limitation of our study is that we focus on general guidelines of multinational survey programs only, but do not include a country-level comparison. Although specific training content or formal aspects (e.g., the number of interviewers, interviewer payment) are defined in the guidelines of the survey programs, compliance with and implementation of these guidelines can vary considerably in the participating countries. In addition, in many countries survey agencies are responsible for organizing and conducting interviewer training. Arrangements between the survey agency and the national coordinator of the survey can also determine the content of the training. Therefore, examining the technical reports to compare the actual implementation of interviewer training at the country level

would be a promising avenue for future research. In addition, we use the TSE as theoretical framework for comparing training content of the three multinational survey programs. However, the TSE itself also has limitations. While using the TSE framework does allow us to map the recommended content, it does not allow us to map, for example, the recommended didactic methods.

Moreover, different learning strategies (e.g., class instructions and practical sessions) are implemented mainly in the case of training content related to measurement error. Specifically, all three guidelines highlight the importance of practical training sessions. For the development of interviewer training concepts, the application of different learning modes and methods appears to be an important aspect which has only be sparsely taken into account so far (e.g., Daikeler & Bosnjak, forthcoming; Rutgers Online Degrees, n.d.; M. K. Smith, 2002): the field of andragogy addresses this topic and offers principles that are useful to consider when designing interviewer training programs (Tusting & Barton, 2003). For example, following Malcolm Knowles, adults prefer a self-directed approach and learning that is centered around common tasks (Meuler, 2010; Rutgers Online Degrees, n.d.; M. K. Smith, 2002). Moreover, interviewer training materials should take account of the fact that levels of educational attainment and experience vary greatly among adults.

Looking forward, in order to empirically investigate the different effects of general and study-specific interviewer training on the components of the TSE and, thus, on data quality, more experimental studies are needed. These studies should explore the effects of the various interviewer training contents on the different error sources as well as the interaction of different error sources. For example, experimental evidence is needed to ensure that the focus on gaining cooperation, which is typical of many interviewer training concepts, contributes to effectively reducing nonresponse. A theoretical foundation could be the organizing model for future research investigating explanations for interviewer effects on multiple error sources, which West and Blom (2017) proposed in their research synthesis on interviewer effects. Their proposed model includes interviewer training as a background characteristic that can be modified depending on the sources of interviewer effects identified. Future studies could structure research topics of interviewer training along this model in order to evaluate their impact on the respective survey errors.

Conclusion

The study showed that the interviewer training guidelines of all three multinational survey programs provided an extensive training content that addresses multiple error sources of the TSE framework. While the coverage of some error sources such as sampling and processing error could be improved when interviewers are

involved in these processes, the most important error sources nonresponse and measurement error are broadly covered by all three training guidelines. Altogether, those guidelines could serve, together with the survey guidelines formulated outside of the context of specific survey programs (e.g., Alcser, et al., 2016; Daikeler, Silber, Bosnjak, Zabal, & Martin, 2017; ISO, 2012; Lessler et al., 2008), as an excellent starting point for new – small as well as large-scale – surveys to design their interviewer training. Even the interviewer training of existing multinational survey programs could benefit from learning how other surveys plan and implement interviewer training to ensure interviewers are well prepared for all tasks they have to fulfil during the implementation of the respective survey. It might be helpful to define the focus of interviewer training through determining the time devoted to a specific topic dependent on the magnitude of the survey error related to that training topic (West & Blom, 2017). For example, a focus on training in contact and cooperation strategies is undoubtedly a good strategy in times of lower response rates or higher nonresponse bias. However, other components of the TSE should be likewise addressed in the respective interviewer training guidelines.

All this is only possible when interviewer training guidelines and materials are publicly available. Consequently, all survey programs would benefit from a high level of transparency (e.g., published interviewer training material). And since not all survey programs can afford a cost intensive high quality interviewer training, it would be imperative to have a standardized, pre-established training manual from which even smaller surveys can use relevant training modules.

References

American Association for Public Opinion Research (AAPOR, n.d.). best practices for survey research. Retrieved from https://www.aapor.org/Standards-Ethics/Best-Practices.aspx (03/13/2018).

Alcser, K., Clemens, J., Holland, L., Guyer, H., & Hu, M. (2016). Interviewer recruitment, selection, and training. Guidelines for Best Practice in Cross-Cultural Surveys. Survey Research Center, Institute for Social Research, University of Michigan.

Benson, M. S., & Powell, M. B. (2015). Evaluation of a comprehensive interactive training system for investigative interviewers of children. Psychology, Public Policy, and Law, 21(3), 309-322. https://doi.org/10.1037/law0000052

Beullens, K., Loosveldt, G., Denies, K., & Vandenplas, C. (2016). Quality matrix for the European Social Survey, round 7. Retrieved from http://www.europeansocialsurvey.org/docs/round7/methods/ESS7_quality_matrix.pdf (2018/05/02)

Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018). Response rates in the European Social Survey: Increasing, decreasing, or a matter of fieldwork efforts? Survey Methods Insights from the Field. Retrieved from https://surveyinsights.org/?p=9673

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. Public Opinion Quarterly, 74(5), 817-884. https://doi.org/10.1093/poq/nfq058

Billiet, J., & Loosveldt, G. (1988). Improvement of the quality of responses to factual survey questions by interviewer training. Public Opinion Quarterly, 52(2), 190–211. Retrieved from http://www.jstor.org/stable/2749273

Börsch-Supan, A., & Jürges, H. (2005). The Survey of Health, Aging, and Retirement in Europe – Methodology. Mannheim.

Cantor, D., Allen, B., Schneider, S. J., Hagerty-Heller, T., & Yuan, A. (2004). Testing an automated refusal avoidance training methodology. Paper presented at the 60th Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ.

Dahlhamer, J. M., Cynamon, M. L., Gentleman, J. F., Piani, A. L., & Weiler, M. J. (2010). Minimizing survey error through interviewer training: New procedures applied to the National Health Interview Survey (NHIS).

Daikeler, J., & Bosnjak, M. (forthcoming). How to conduct effective interviewer training: A meta-analysis. In K. Olson, J. D. Smyth, J. Dykema, A. Holbrook, F. Kreuter, & B. T. West (Eds.), Interviewer Effects from a Total Survey Error Perspective: CRC Press.

Daikeler, J., Silber, H., Bosnjak, M., Zabal, A., & Martin, S. (2017). General interviewer training curriculum for computer-assisted personal interviews (GIT-CAPI; Version 1, 2017). GESIS Survey Guidelines. Mannheim, Germany: GESIS – Leibniz-Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_022

Durand, C., Gagnon, M.-E., Doucet, C., & Lacourse, E. (2006). An inquiry into the efficacy of a complementary training session for telephone survey interviewers. Bulletin de Méthodologie Sociologique, 92(1), 5-27. https://doi.org/10.1177/075910630609200103

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

ESS. (2016a). ESS interviewer briefing: NC manual. London: ESS ERIC headquarters. Retrieved from Bergen, European Social Survey Data Archive: http://www.europeansocialsurvey.org/methodology/ess_methodology/data_collection.html (2018/04/30)

ESS. (2016b). ESS round 8 interviewer briefing: Interviewer manual. London: ESS ERIC headquarters. Retrieved from Bergen, European Social Survey Data Archive: http://www.europeansocialsurvey.org/methodology/ess_methodology/data_collection.html (2018/04/30)

ESS. (2016c). Survey specification for ESS ERIC member, observer and guest countries. Retrieved from http://www.europeansocialsurvey.org/methodology/ess_methodology/survey_specifications.html (04/10/2018)

ESS. (2018). ESS-8 2016 documentation report. Edition 2.1. Retrieved from Bergen, European Social Survey Data Archive: https://www.europeansocialsurvey.org/docs/round8/survey/ESS8_data_documentation_report_e02_1.pdf (2018/05/29)

Fowler Jr, F. J., & Mangione, T. W. (1986). Reducing interviewer effects on health survey data (Report No. 141). Retrieved from Rockville, MD: National Center for Health Services Research and Health Care Technology.

Fowler Jr., F. J. (1991). Reducing interviewer-related error through interviewer training, supervision, and other means. In P. P. Biemer, R. M. Groves, L. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), Measurement errors in surveys (pp. 259-278). Hoboken, NJ: John Wiley & Sons.

Groves, R. M. (2005). The interviewer as a source of survey measurement error. In R. M. Groves (Ed.), Survey errors and survey costs (pp. 357-406): John Wiley & Sons.

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. Lep-

kowski, 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. (2009). 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. https://doi.org/10.1093/poq/nfq065

Guest, L. (1954). A new training method for opinion interviewers. Public Opinion Quarterly, 18(3), 287-299.

Hubal, R. C., & Day, R. S. (2006). Informed consent procedures: An experimental test using a virtual character in a dialog systems training application. Journal of Biomedical Informatics, 39(5), 532-540.

International Organization for Standardization (ISO). (2012). Market, opinion and social research - Vocabulary and service requirements (ISO Standard No. 20252: 2012 [en]. Geneva, Switzerland: International Organization for Standardization.

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 (pp. 442-478). New York, NY: Taylor & Francis Group.

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.

Malter, F., & Börsch-Supan, A. (2017). SHARE Wave 6: Panel innovations and collecting dried blood spots. Munich: Munich Center for the Economics of Aging (MEA) at the Max Planck Institute for Social Law and Social Policy (MPISOC).

Malter, F., Schuller, K., & Börsch-Supan, A. (2016). SHARE compliance profiles – Wave 6. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Mayer, T. S., & O'Brien, E. (2001). Interviewer refusal aversion training to increase survey participation. Proceedings of the Annual Meeting of the American Statistical Association, 2001.

McConaghy, M., & Carey, S. (2005). What happens after ART? Results of two experiments designed to improve response rates with interviewers at the Office of National Statistics, UK. Paper presented at the 60th Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ.

Meuler, E. (2010). Didaktik der Erwachsenenbildung - Weiterbildung als offenes Projekt. In R. Tippelt & A. Von Hippel (Eds.), Handbuch Erwachsenenbildung/Weiterbildung (Vol. 4, pp. 973-988). Wiesbaden: VS Verlag für Sozialwissenschaften.

O'Brien, E. M., Mayer, T. S., Groves, R. M., & O'Neill, G. E. (2002). Interviewer training to increase survey participation. Proceedings of the 2002 Joint Statistical Meetings of the American Statistical Association.

OECD. (2011). PIAAC technical standards and guidelines. Retrieved from http://www.oecd.org/skills/piaac/documentation.htm (2014/06/12)

OECD. (2013). The survey of adult skills: Reader's companion, second edition, OECD skills studies, OECD Publishing, Paris. Retrieved from https://doi.org/10.1787/9789264258075-en (2016/12/05)

Pennell, B.-E., Harkness, J. A., Levenstein, R., & Quaglia, M. (2010). Challenges in cross-national data collection. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), Survey methods in multi-

national, multiregional, and multicultural contexts (pp. 269-298). Hoboken, NJ: John Wiley & Sons.

Pennell, B.-E., Hibben, K. C., Lyberg, L. E., Mohler, P. P., & Worku, G. (2017). A total survey error perspective on surveys in multinational, multiregional, and multicultural contexts. In P. P. Biemer, E. D. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, C. N. Tucker, & B. T. West (Eds.), Total Survey Error in practice (pp. 179-201). Hoboken, NJ: John Wiley & Sons.

Rutgers Online Degrees. (n.d.). The principles of adult learning theory. Retrieved from https://online.rutgers.edu/blog/principles-of-adult-learning-theory/ (2018/05/04)

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.

Schnell, R., & Trappmann, M. (2006). The effect of the refusal avoidance training experiment on final disposition codes in the German ESS-2 (Working Paper 3/2006: Center for Quantitative Methods and Survey Research, University of Konstanz). Retrieved from http://nbn-resolving.de/urn:nbn:de:0168-ssoar-114189

Schouten, B., Peytchev, A., & Wagner, J. (2017). Adaptive survey design. Boca Raton, FL: CRC Press.

Smith, M. K. (2002). Malcolm Knowles, informal adult education, self-direction and andragogy. Retrieved from http://infed.org/mobi/malcolm-knowles-informal-adult-education-self-direction-and-andragogy/ (2018/05/04)

Smith, T. W. (2011). Refining the total survey error perspective. International Journal of Public Opinion Research, 23(4), 464–484. https://doi.org/10.1093/ijpor/edq052

Smith, T. W. (2019). Improving multinational, multiregional and multicultural comparability (3MC) using the total survey error (TSE) paradigm In T. P. Johnson, B.-E. Pennell, I. Stoop, & B. Dorer (Eds.), Advances in comparative surveymethods: Multinational, multiregional, and multicultural contexts (3MC) (pp. 1-21). Hoboken, NJ: John Wiley & Sons.

Survey Research Center at the University of Michigan. (2016). Guidelines for best practice in cross-cultural surveys. Survey Research Center, Institute for Social Research, University of Michigan.

Tusting, K., & Barton, D. (2003). Models of adult learning: a literature review. Retrieved from London, UK: National Research and Development Centre for Adult Literacy and Numeracy (www.nrdc.org.uk (2018/05/09)).

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

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

Appendix

Survey Specifications and Characteristics of the ESS, PIAAC, and SHARE Across all Participating Countries

	ESS 2016 (Round 8)[1]	PIAAC (Cycle 1, Round 1)[2]	SHARE (Wave 6)[3]
Scope	Attitudes, beliefs, and behavior	Adult competencies	Health, ageing, and retirement
Survey type	Cross-sectional (first round in 2002, Round 8 in 2016) Every 2 years	Cross-sectional (first cycle in 2011/12) Multi-cycle (every 10 years)	Longitudinal (start in 2004/05, Wave 6 in 2014/15) Every 2 years
Survey mode	F2F: CAPI	F2F core questionnaire: CAPI Self-administered cognitive assessment: CASI, PAPI	F2F: CAPI Self-administered PAPI for sensitive questions in some countries[4] Physical measurements
Sampling method	Probability-based	Probability-based	Probability-based
# of participating countries	23	24	18
Target population	General, 15 years +	General, 16 - 65 years	General, 50 years +
Field period	September–December 2016	August 2011–March 2012	February–November 2015
Realized sample size per country (range)	880–2,852	3,761–27,285	1,169–6,100[5]
Interview duration (average)	60 minutes	90 minutes	80 minutes
# of trained interviewers per country (range)	41–281	70–810	---[6]

	ESS 2016 (Round 8)[1]	PIAAC (Cycle 1, Round 1)[2]	SHARE (Wave 6)[3]
Non-standard interviewer tasks[7]	No	Yes	Yes
Interviewer remuneration scheme	General approach: freelance contract, paid per completed interview	Independent of the number of completed interviews	---
Interviewer workload	Max. of 48 cases (respondents and non-respondents)	Max. of 40 completed assessments per month	On average 15 interviews; must not exceed 50
Interviewer recruitment & hiring	Responsibility of survey institute Interviewers with general training	Responsibility of survey institute Aim is to hire interviewers with at least 2 years working experience Selected from a variety of job-offering sources Applicants with various qualifications Process should start at least 8 weeks prior to the start of data collection	The appropriate number of interviewers have to be available in a sufficient regional spread

Note. Survey programs in alphabetical order; CAPI = computer-assisted personal interviewing; CASI = computer-assisted self-interviewing; ESS = European Social Survey; F2F = face to face; PAPI = paper and pencil interviewing; PIAAC = Programme for the International Assessment of Adult Competencies; SHARE = Survey of Health, Ageing and Retirement in Europe; --- = no information available in the survey documentation.

[1] ESS (2016c); Beullens, Loosveldt, Denies, and Vandenplas (2016); ESS (2018).
[2] OECD (2013).
[3] Börsch-Supan and Jürges (2005); Malter, Schuller, and Börsch-Supan (2016); SHARE (2018).
[4] Austria, Czech Republic, Greece, Israel, Slovenia, Switzerland.
[5] Girona added to Spain.
[6] 44–170 interviewers were working in the field. Information about the number of interviewers trained was not published.
[7] Refers to tasks that go beyond the standard tasks (e.g., contacting, gaining cooperation, conducting face-to-face interviews).