

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Sociology Department, Faculty Publications

Sociology, Department of

2020

The Past, Present, and Future of Research on Interviewer Effects

Kristen M. Olson

Jolene Smyth

Jennifer Dykema

Allyson L. Holbrook

Frauke Kreuter

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/sociologyfacpub>



Part of the [Family, Life Course, and Society Commons](#), and the [Social Psychology and Interaction Commons](#)

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Sociology Department, Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Kristen M. Olson, Jolene Smyth, Jennifer Dykema, Allyson L. Holbrook, Frauke Kreuter, and Brady T. West

The Past, Present, and Future of Research on Interviewer Effects

Kristen Olson,¹ Jolene D. Smyth,¹ Jennifer Dykema,²
Allyson L. Holbrook,³ Frauke Kreuter,⁴ & Brady T. West ⁵

1 University of Nebraska-Lincoln

2 University of Wisconsin-Madison

3 University of Illinois at Chicago

4 University of Maryland, University of Mannheim, & Institute for
Employment Research

5 University of Michigan

Contents

1 Introduction 1

2 Training, Managing, and Monitoring Interviewers 3

3 Interviewer Effects Across Contexts and Modes 6

4 Interviewers and Nonresponse 7

5 Interviewer Pace and Behaviors 8

6 Estimating Interviewer Effects 11

7 Closing Thoughts 12

Acknowledgments 15

References 16

1.1 Introduction

Interviewer-administered surveys are a primary method of collecting information from populations across the United States and the world. Various types of interviewer-administered surveys exist,

Published in *Interviewer Effects from a Total Survey Error Perspective*, ed. Kristen Olson, Jolene D. Smyth, Jennifer Dykema, Allyson L. Holbrook, Frauke Kreuter, and Brady T. West (2020), Boca Raton: CRC Press, pp. 3-15.

Copyright © 2020 Taylor & Francis Group, LLC. Used by permission.

including large-scale government surveys that monitor populations (e.g., the Current Population Survey), surveys used by the academic community to understand what people think and do (e.g., the General Social Survey), and surveys designed to gauge public opinion at a particular time point (e.g., the Gallup Daily Tracking Poll). Interviewers participate in these data collection efforts in a multitude of ways, including creating lists of housing units for sampling, persuading sampled units to participate, and administering survey questions (Morton-Williams 1993). In an increasing number of surveys, interviewers are also tasked with collecting blood, saliva, and other biomeasures, and asking survey respondents for consent to link survey data to administrative records (Sakshaug 2013). Interviewers are also used in mixed mode surveys to recruit and interview non respondents after less expensive modes like mail and web have failed (e.g., the American Community Survey and the Agricultural Resource Management Survey; de Leeuw 2005; Dillman, Smyth and Christian 2014; Olson et al. 2019). In completing these varied tasks, interviewers affect survey costs and coverage, nonresponse, measurement, and processing errors (Schaeffer, Dykema and Maynard 2010; West and Blom 2017).

Errors introduced by interviewers can take the form of bias or variance. Early research found that interviewers vary in how they administer survey questions and that their effects were similar to sample clusters in both face-to-face (Hansen, Hurwitz and Bershad 1961; Kish 1962) and telephone surveys (Groves and Magilavy 1986; Mathiowetz and Cannell 1980). In particular, similar to a design effect for cluster samples, interviewers increase the variance of an estimated mean as a function of their average workload (b) and the intra-interviewer correlation (IIC) (the degree of within-interviewer correlation in measurements): $1 + (\bar{b} - 1) IIC$. IIC values range from very small (0.001) to large (0.20) and larger (Elliott and West 2015; Groves and Magilavy 1986; O’Muircheartaigh and Campanelli 1998), with median values typically below 0.02. Given an IIC of 0.02 and a workload of 50 interviews per interviewer, the variance of an estimated mean is almost doubled and standard errors are increased by 40%. Thus, a fundamental goal of research on interviewers is understanding what contributes to (and how to minimize) the IIC.

Even if the IIC is small, interviewer characteristics and behaviors

can still bias responses. Interviewer sociodemographic characteristics (e.g., sex, race, age, education), personality traits, experience, attitudes and expectations, or even the paralinguistic qualities of their voices may be associated with responses to survey questions or indicators of survey error (e.g., an indicator of a sampled unit responding to a survey request; Schaeffer et al. 2018). Inasmuch as these are fixed characteristics of interviewers, they typically bias estimates. In addition, research based on coding information about the interviewer-respondent interaction shows that interviewers' behaviors such as misreading questions, probing directly, and acting non-neutrally may affect respondents' behaviors and survey error (e.g., Fowler 2011; Ongena and Dijkstra 2006).

The goal of the *Interviewers and Their Effects from a Total Survey Error Perspective Workshop* in February 2019 was to convene an international group of leading academic, government, and industry researchers and practitioners to discuss methods, practical insights, and research findings on interviewers. Specifically, the workshop aimed to (1) synthesize and expand knowledge about the impacts of interviewers on multiple error sources, (2) evaluate study design and estimation approaches for studying interviewer effects, and (3) produce an agenda for future studies of interviewers. After two days of presentations and posters, workshop participants spent the third day discussing and identifying areas where more work is needed. This chapter introduces an edited volume consisting of chapters written by workshop participants. In order to do so, we first provide an overview of research and practice related to interviewers at different stages of the survey process. Second, we situate the chapters from this volume within the existing literature. Finally, we identify areas for future research that arose from the third day of focused discussion.

2 Training, Managing, and Monitoring Interviewers

Standardized interviewing aims to limit the effect of the interviewer on the resultant survey data. Although standardized interviews are the gold standard, a strict implementation of standardization may not exist in practice because “standardized” interviewer training, monitoring, and feedback systems vary widely across survey organizations

(Vitema and Maynard 2002). Standardized interviewing commonly includes asking questions exactly as written, recording answers as provided, following up inadequate answers nondirectively, and acknowledging adequate answers (Fowler and Mangione 1990). Yet some of the central tenants of standardized interviewing- including reading questions verbatim - are inadequately operationalized in actual training materials. Additionally, survey practitioners often make decisions about interviewer training with little to no empirical evidence regarding effectiveness. Furthermore, there has been almost no update to standardized training philosophy or materials since Fowler and Mangione's (1990) canonical volume. In Chapter 3, Schaeffer et al. tackle this important issue, updating basic training of General Interviewing Techniques for question administration based on decades of research on interviewer-respondent interaction and survey practice. Among other things, their update tackles thorny issues of how to recognize a codable answer for different question types (thus acknowledging the critical role of characteristics of survey questions in determining what makes an answer codable), how to maintain respondent engagement, and what common conversational practices can be allowed in the interview.

While studies of interviewers and their effects often have implications for survey operations, it was clear during the workshop that this research has not been fully adopted by survey organizations. We have few recent descriptions of how survey organizations select, train, and monitor survey interviewers, especially for smaller survey organizations. Yet many operational concerns are researchable, and such research could yield both theoretical and practical insights. Miller and Mathiowetz (Chapter 2) describe how Charlie Cannell's seminal work expanded theoretical insights into interviewer-respondent interaction, how this interaction can positively or negatively affect data quality, and what kind of practical training methods may address these concerns. Unfortunately, few studies have systematically examined properties of interviewer training protocols. While Daikeler and Bosnjak's (Chapter 4) meta-analysis of interviewer training includes many studies on training efficacy for avoiding refusals, few studies evaluate training efficacy for administering survey questions. The workshop identified many areas for research on recruiting, training, and monitoring interviewers, including how to identify successful interviewers

from a pool of applicants, how the pool of interviewers has changed over time and across modes of data collection, and the influence of the supervisor on the interviewer. Future research is also needed to address the mix of training that should be devoted to skill development, to shaping interviewer perceptions about their tasks, what combination of in-person and online methods are best for delivering content, the optimal length of training for different aspects of the survey process, and interviewer training when standardization is difficult to maintain.

Many have questioned the utility of standardization, particularly for sensitive, complicated, or difficult questions or questionnaires (e.g., Conrad and Schober 2000; Maynard et al. 2002; Schaeffer 1991; Schober and Conrad 1997; Suchman and Jordan 1990). Interviewers often break from standardization on these types of items and questionnaires. Based on qualitative in-depth interviews with survey interviewers and their reactions to vignettes, Kaplan and Yu (Chapter 5) describe interviewers' approaches to administering sensitive and difficult questions, providing unique insights into why deviations from standardization occur. Innovations in technology, increased computing resources in the field, and hands-on monitoring of interviewers using paradata have dramatically changed how we monitor and train interviewers, particularly for in-person interviewing (e.g., Edwards, Maitland and Connor 2017; Olson and Wagner 2015). Edwards, Sun, and Hubbard (Chapter 6) report on an attempt to intervene in real time when interviewers break from standardization for difficult questions as detected via monitoring using computer-assisted recorded interviewing (CARI). These two chapters highlight the need to train interviewers on follow-up methods such as probing and clarification, and the value of reinforcing this training through real-time monitoring. Throughout the workshop, participants echoed the need for more information on how survey organizations monitor and provide feedback to interviewers and what methods are most efficacious for in-person and telephone interviews.

Although in-depth interviews and CARI recordings provide unique insights into interviewers' behavior during the survey data collection process, many organizations lack resources to conduct such studies and instead use paradata and interviewer observations to evaluate the survey process after data collection is finished. Schwanhauser

et al. (Chapter 7) use paradata on interview duration and indicators from the survey data to detect potential falsification by interviewers; this study is unique in that the interviewer-level falsification indicators are compared to data from the field that identified three *actual* falsifiers, revealing where the statistical analysis-based falsification indicators help to identify problems and where they fail. West et al. (Chapter 8) use data from two large surveys to assess whether post-survey interviewer observations about the survey process are associated with indicators of measurement error. The sets of observations that are most important to collect for monitoring and evaluation purposes were difficult to discern given the tremendous heterogeneity in the types of survey questions asked and observations collected across studies. More work is needed to align these observations across survey organizations, to facilitate replication of effects and potentially benefit secondary data users.

3 Interviewer Effects Across Contexts and Modes

The survey context also matters. Although interviewers are instructed to conduct interviews in private, many interviews are conducted in the presence of known others, violating respondents' need for privacy when reporting about sensitive topics (Mneimneh et al. 2015). In community-based studies where in-community interviewers are recruited who may know the respondent personally, these effects may be amplified. Alternatively, third-party presence may facilitate recall when questions are particularly difficult. Mneimneh, de Jong, and Altwaijri (Chapter 9) examine whether conducting an interview with various family or non-family members present affects reports of health behaviors and attitudes in the in-person Saudi National Mental Health Survey. Habecker and Ivanich (Chapter 10) examine how youth's reports of internalizing and externalizing behaviors and other sensitive topics are affected when the youth are interviewed by a member of the community who is known to them in a community-based participatory research study in an American Indian community. Combined, these studies reveal the need for more research into interview privacy and the importance of interviewer training on how to maximize privacy.

Additionally, the mode and/or device for the interview - in-person,

landline phone, cellular phone, or audio computer-assisted self-interviewing; and interviewer input into a desktop or laptop, tablet, or smartphone - provides important context and potential for variation in interviewer-related errors (e.g., Childs and Landreth 2006; Timbrook, Smyth and Olson 2018). Notably, the mode or device for the interaction changes the nature of the interaction between interviewers and respondents. In Chapter 12, Ongena and Haan examine differences in the interviewer-respondent interaction across telephone and in-person interviews; in Chapter 13, Schober et al. examine differences in interviewer-respondent interaction across voice and text message-based interviews. Conrad et al. (Chapter 11) replace human interviewers with avatars, examining race-of-interviewer effects when the interviewer is virtual, but their voice is that of an audio-recorded human. These chapters raise important questions about interviewing: Are there contexts in which the presence of the interviewer (live or otherwise) is an important feature of the surveyor where tasks are critical for the interviewer to consider? Can virtual interviewers provide some of the benefits of live interviewers while minimizing interviewer error? Are text message-based interviews or avatars particularly beneficial for specific populations?

4 Interviewers and Nonresponse

While most of the research identified above focuses on measurement error, interviewers affect other error sources. For example, interviewers' nonresponse rates vary extensively (e.g., Campanelli, Sturgis and Purdon 1997; Groves and Couper 1998) - due to both heuristic cues from their voices (e.g., Groves et al. 2008; Schaeffer et al. 2018) and their behaviors during the recruitment interaction (e.g., Couper and Groves 2002; Schaeffer et al. 2013) - and substantially contribute to nonresponse error variance (e.g., West, Kreuter and Jaenichen 2013; West and Olson 2010). Interviewer flexibility and tailoring have been linked to successful recruitment, although there are only limited examples of how tailoring is operationalized (e.g., Groves and Couper 1998). Research about tailoring generally relies on interviewer reports with measures that vary across studies. To address this, Ackermann-Piek, Korbmacher, and Krieger (Chapter 14) predict survey contact

and cooperation with the same set of covariates across four different studies conducted by the same survey organization. Because they find little replication in associations between the covariates and the non-response outcomes across studies, they emphasize the importance of real-time monitoring of interviewers.

Interviewer flexibility may not always be a good thing. In Chapter 15, Wescott discusses a case management model for a telephone survey in which interviewers “own” cases and make their own decisions about when to call cases. While interviewers report being more satisfied with the autonomy afforded by the case management model, the model yields lower productivity than using a call scheduler. More work is needed to understand how interviewer autonomy and insights from a field data collection approach may be integrated into a telephone survey organization to increase interviewer engagement and ultimately retain interviewers.

Survey interviews increasingly ask respondents to provide blood, saliva, urine, or other biomeasures or for permission to link their survey data to administrative data (e.g., Jaszczak, Lundeen and Smith 2009; Sakshaug 2013; Sakshaug et al. 2012). In Chapter 16, Pashazadeh, Cernat, and Sakshaug use nurse characteristics and paradata to predict different stages of nonresponse for nurses’ attempts to collect biomeasures for a general population survey. These stages—participating in the nurse visit, consenting to a blood sample, and obtaining a blood sample—reveal substantial variation in the nonresponse outcomes related to the nurses, and that the predictors of nonresponse vary across the stages. This work and additional workshop discussion suggest that we need more research on the antecedents and consequences of interviewer variation in the ability to successfully collect auxiliary measures.

5 Interviewer Pace and Behaviors

Four chapters examine interview pace and behaviors. Holbrook et al. (Chapter 17) test what interviewer and question characteristics predict interviewer reading speed (IRS) and what effect IRS has on response latencies and indicators of respondent comprehension and mapping difficulties. Garbarski et al. (Chapter 18) examine how the

time taken to administer and answer questions is associated with interviewer, respondent, and question characteristics. Kelley (Chapter 19) examines how well question administration time thresholds generated from timing paradata can be used to identify question misreadings, testing three methods of setting thresholds. Olson and Smyth (Chapter 20) address the longstanding finding that interviews get faster later in the field period by examining changes in interviewers' behaviors over time.

Although the studies ostensibly examine the same phenomenon—pace—their conceptualizations and operationalizations vary (e.g., interview duration, question duration, interviewer speaking time or speed, response latencies, words per second, and questions per minute; see also Chapter 21, in which Dahlhamer et al. use average number of seconds per question across a questionnaire). Other studies examining pace use behavior or interaction coding to examine interviewer and respondent behaviors (e.g., Fowler 2011; Ongena and Dijkstra 2006), producing evidence of similar variation in how behavior coding is implemented (from live interviews, recordings, or transcripts, and focused on respondents, interviewers, or both) and the variety of operational and analytical decisions often made in such research. Such decisions include assigning codes at the question level or the conversational turn level (Olson and Parkhurst 2013); coding the entire questionnaire or a subset of items (e.g., see review in Ongena and Dijkstra 2006); using codes individually (e.g., Fowler 2011; Mathiowetz and Cannell 1980) or in combination (e.g., to measure rapport as in Garbarski, Schaeffer, and Dykema 2016); examining behaviors when they initially occur or at any point during the question-answer interaction; examining behaviors individually or sequentially; and dealing with overlapping speech, interruptions, and other normal conversational events. Even simple issues of how many interviews to code and (acceptable) reliability of the codes vary across studies.

Despite this heterogeneity, there are common patterns observed across these studies. First, interviewers speed up with more experience (Holbrook et al., Chapter 17; Garbarski et al., Chapter 18; Olson and Smyth, Chapter 20; Olson and Bilgen 2011; Olson and Peytchev 2007). Second, question characteristics drive this phenomenon. **Figure 1** shows the percent of variance in question duration attributable

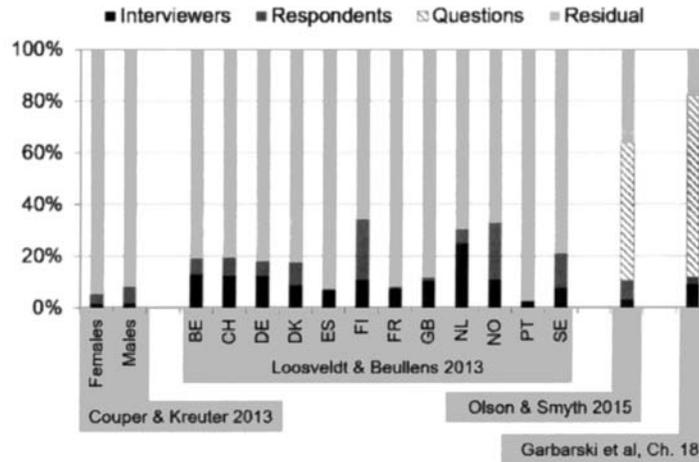


Figure 1 Variance in question duration due to interviewers, respondents, and questions.

to interviewers, respondents, and questions across multiple studies with question-level duration as the outcome. Some models are estimated as three-level multilevel models, so the residual variance accounts for question-level variation. Other models are estimated as cross-classified models, where question-level variation is explicitly estimated as part of the model. Across these four studies, interviewer- and respondent-level variation is small and the question-level (or residual that incorporates questions) variation is large.

There is a consistent tendency for longer questions and questions written at higher grade levels to have longer durations (e.g., Garbarski et al. Chapter 18; Couper and Kreuter 2013; Olson and Smyth 2015). Yet other question characteristics, including question placement, question sensitivity, type of question (attitude, factual knowledge), response option format, presence of definitions, emphasis, instructions, parentheticals, battery items, and measures from survey evaluation tools such as QUAID (Graesser et al. 2006) or SQP (Saris and Gallhofer 2007), are inconsistently parameterized or have inconsistent associations with duration. Pace reflects interviewer behaviors and influences respondent behaviors, although understanding the mechanisms for these connections still requires more work. One clear direction from the workshop for future research was to identify a common set of dependent variables related to pace and behaviors, a

common set of question characteristics, and a common set of respondent and interviewer characteristics, parameterize these identically, and evaluate whether this standardization “solves” some of the inconsistencies across these different observational studies. In addition, with advances in text analysis and searches, the relative cost and error trade-offs of human coders versus computer coding for this type of research is also of interest.

6 Estimating Interviewer Effects

Given the nesting of respondents within interviewers, following Kish’s ANOVA-based model (Kish 1962), hierarchical or random effects models have long been used for the study of interviewer effects (e.g., Dijkstra 1983; Hox 1994; O’Muircheartaigh and Campanelli 1998). Recent applications of hierarchical models expand the complexity of these models, including hierarchical cross-classified random effects models to study how properties of survey questions themselves may affect and be affected by interviewers (e.g., Couper and Kreuter 2013; Holbrook et al. 2016; Olson and Smyth 2015) or to separately estimate area effects from interviewer effects (e.g., O’Muircheartaigh and Campanelli 1998, 1999).

Multilevel models are flexible and can be used to infer whether interviewer effects differ across subgroups of items, respondents, and interviewers. In Chapter 21, Dahlhamer et al. use cross-classified multilevel models to disentangle area effects from interviewer effects on over 100 outcomes in the National Health Interview Survey. They then meta-analyze these interviewer effects across question and interviewer characteristics, finding that longer questions and those with higher reading levels have larger interviewer effects, as do questions on more complex and difficult topics. Interviewers who administer questions at a faster pace also have larger interviewer effects than those who administer questions at a slower pace. Similarly, Loosveldt and Wuyts (Chapter 22) use two-level random effects models to examine interviewer effects on 14 questions from the European Social Survey across subgroups of respondents defined by education in multiple countries, comparing two analytic approaches to estimating these effects. Similar to Dahlhamer et al., Loosveldt and Wuyts meta-analyze

interviewer effects across education subgroups, finding consistently larger interviewer effects for the low education group.

Other methods exist for testing systematic differences in interviewer variance across independent groups of interviewers. In Chapter 23, West reviews design decisions needed to compare interviewer variance components across two groups of interviewers, using a unique study in Germany comparing standardized and conversational approaches to interviewing as an example. From allocation of interviewers to conditions to power analyses to analytic decisions, this comprehensive review of the decisions needed to effectively design and analyze an experiment on interviewers yields easy-to-follow and practical insights into these complicated designs.

7 Closing Thoughts

Survey research is adaptive, as reflected by the content and range of chapters in this volume. Even though interviewers have been central to data gathered to understand society since the beginning of survey research, we know surprisingly little about them. Namely, what interviewer characteristics exist across studies and across organizations, how interviewers perceive their job, and how we can recruit and retain high-quality interviewing staff. Yet these are important issues - interviewers can do a lot of harm to survey data if they try to do so (Chapter 7) but also inadvertently introduce error into data even when they are attempting to follow their training. Understanding the challenges and constraints interviewers face will facilitate understanding the mechanisms underlying interviewer-related survey errors.

Most of the chapters in this volume use observational data. Observational research is constrained by the data that a research team has available. Through the workshop and studies featured in this volume, we have the opportunity to refine our conceptualizations and operationalizations and create more consistency across operational implementations and research studies. Examples of key concepts/topics that would benefit from a shared set of definitions include “conversational interviewing,” “standardized interviewing,” “interviewer effects,” “interviewer experience,” “interview(er) pace,” and “question

characteristics” (and the types therein).

More work is needed to examine interviewers cross-nationally. In some countries, a more flexible form of interviewing is standard practice. How this relates to “conversational interviewing” in United States nomenclature (e.g., Conrad and Schober 2000), as well as the concomitant effects on survey data, are unknown. The heterogeneity in interviewer-related variance across countries (Loosveldt and Wuyts, Chapter 22) reveals the need for understanding interviewing practice, monitoring, and supervision in a multinational context. This volume contains many studies conducted outside of the United States (e.g., Chapters 4, 7, 8, 9, 12, 14, 16, 19, 22, and 23). Yet a deep understanding of how interviewers operate in different contexts is an area for future research.

As researchers and practitioners, we call on survey organizations to make information on interviewers available in public-use analytic data files. **Table 1** contains a list of concepts that workshop participants considered important to examine and include in future research studies. At the bare minimum, an anonymized interviewer ID variable on data files would allow analysts to estimate interviewer variance components. Additional data on interviewers, extending beyond simply demographics and experience, would facilitate understanding the mechanisms by which interviewers affect survey data. At the interviewer level, these include measures of work productivity and quality, measures of how thinly spread the interviewer is across multiple projects and/or organizations, attitudinal and expectation measures from the interviewers themselves about their job, as well as measures from field performance. At the study or organizational level, information about the amount, type, and content of interviewer training as well as details about the supervision and monitoring practices and feedback provided to interviewers from supervisors or monitors would greatly enhance studies of interviewer effects. Detailed information about the content of general interviewing techniques training versus study-specific training also would provide insights into how these individual decisions cumulate to affect survey errors related to interviewers. Some of this information could be included in methodology reports. Although many organizations may consider information about training, supervision, and monitoring to be proprietary, more complete disclosure is certainly needed to understand these rarely

Table 1 A List of Recommended Information About Interviewers to Include in Public-use Data Files and Methodology Reports

Interviewer ID**Interviewer characteristics***Demographics (gender, race, age, educational status)**Personality assessments**Experience (within survey, within organization, across organizations)**Certification test scores**Number of other jobs they are currently working**Performance metrics and problems on other studies**Ever been fired for other performance issues on other studies**Number of hours work on other studies***Interviewer expectation and attitudinal measures***Ratings of the importance of a completed interview**Ratings of the importance of obtaining high-quality data**Description of how sensitive and difficult questions are approached**Ratings of other variables related to job satisfaction and engagement***Interviewing training variables***Content of training and training methods (e.g., round-robin, type and content of at-home study, hours/days of training by topic)**Participation in any specialized training (e.g., for refusal avoidance)**Number of trainings attended***Interview process variables***Sanitized post-survey observations**Number of other projects interviewers involved in current project worked on during current project**Whether and how interviewers were matched to respondents**Measures of field performance, including ICC information for variables**Adaptive design features and implementation**Interviewers' notes about survey questions***Organization and study-specific characteristics***Information on recruiting, hiring, training, and attrition**Monitoring, feedback, and falsification detection activities**Description of the supervisory structure of the interviewer corps, e.g., number of interviewers per supervisor*

studied, yet critically important, survey practices.

Finally, dissemination and integration of research on interviewers into survey practice is hard. Many new practices may face cultural opposition at organizations simply because it is not the way that work has been done in the past. Clients, survey project managers, and supervisors are often risk averse to trying something new, even if integrating recommendations based on research can improve survey practice. Furthermore, most survey field professionals do not have time to read the latest research, and survey methodologists are often disconnected from field operations. These factors conspire to make translation of research into practice difficult.

We suggest a few ways forward. First, many professional association meetings contain both survey field professionals and methodological researchers. Carving out space for these two disparate groups to discuss mutually interesting problems can facilitate research translating into practice and practice informing research questions. For several years the Midwest Association for Public Opinion Research has sponsored a workgroup in which researchers and practitioners talk about important topics on interviewers. In 2019, the discussion focused on interviewer training, for instance. Second, methodologists who work at organizations with survey shops or who contract for research with survey shops are well positioned for translation of research. Seminars or brown bag discussions with senior field managers at the organizations that collected the survey data about findings related to practice could inspire some changes in survey practice that would improve data quality. Furthermore, recognizing contributions other than simply research articles - for instance, data being available in the public domain, availability of code, availability of interviewer debriefing reports - or short articles that provide case studies of survey practices that worked or did not work would ease translation of work from one research team to another. Although these are hard, we think they are worthwhile future pursuits.

Acknowledgments The workshop and this edited volume were supported by the National Science Foundation (SES-1758834), the Charles Cannell Fund in Survey Methodology of the Survey Research Center at the University of Michigan, and the Rensis Likert Fund for Research

in Survey Methodology at the University of Michigan. Any opinions, findings, and conclusions or recommendations expressed in this text are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Campanelli, P., P. Sturgis, and S. Purdon. 1997. *Can you hear me knocking: An investigation into the impact of interviewers on survey response rates*. London: Survey Methods Centre at SCPR.
- Childs, J. H. and A. Landreth. 2006. Analyzing interviewer-respondent interactions while using a mobile computer-assisted personal interview device. *Field methods* 18:335-351.
- Conrad, F. G. and M. F. Schober. 2000. Clarifying question meaning in a household telephone survey. *Public opinion quarterly* 64:1-28.
- Couper, M. P. and R. M. Groves. 2002. Introductory interactions in telephone surveys and nonresponse. In *Standardization and tacit knowledge: Interaction and practice in the survey interview*, ed. D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, and J. van der Zouwen, 161-177. New York: John Wiley & Sons.
- Couper, M. P. and F. Kreuter. 2013. Using item-level paradata to explore response time and data quality. *Journal of the royal statistical society, A* 176:271-286.
- de Leeuw, E. 2005. To mix or not to mix data collection modes in surveys. *Journal of official statistics* 21:233-255.
- Dijkstra, W. 1983. How interviewer variance can bias the results of research on interviewer effects. *Quality and quantity* 17:179-187.
- Dillman, D. A., J. D. Smyth, and L. M. Christian. 2014. *Internet, phone, mail, and mixed mode surveys: The tailored design method*. Hoboken, NJ: John Wiley & Sons.
- Edwards, B., A. Maitland, and S. Connor. 2017. Measurement error in survey operations management. In *Total survey error in practice*, ed. P. P. Biemer, E. De Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, and B. T. West, 253-277. Hoboken, NJ: John Wiley & Sons.
- Elliott, M. R. and B. T. West. 2015. "Clustering by interviewer": A source of variance that is unaccounted for in single-stage health surveys. *American journal of epidemiology* 182:118-126.
- Fowler, F. J. 2011. Coding the behavior of interviewers and respondents to evaluate survey questions. In *Question evaluation methods: Contributing to the science of data quality*, ed. J. Madans, K. Miller, A. Maitland, and G. Willis, 7-21. Hoboken, NJ: John Wiley & Sons.
- Fowler, F. J. and T. W. Mangione. 1990. *Standardized survey interviewing: Minimizing interviewer-related error*. Newbury Park, CA: Sage Publications.
- Garbarski, D., N. C. Schaeffer, and J. Dykema. 2016. Interviewing practices,

- conversational practices, and rapport: Responsiveness and engagement in the standardized survey interview. *Sociological methodology* 46:1-38.
- Graesser, A. c., Z. Cai, M. M. Louwerse, and F. Daniel 2006. Question understanding aid (QUAID): A web facility that helps survey methodologists improve the comprehensibility of questions. *Public opinion quarterly* 70:3-22.
- Groves, R. M. and M. Couper. 1998. *Nonresponse in household interview surveys*. New York: John Wiley & Sons, Inc.
- Groves, R. M. and L. J. Magilavy. 1986. Measuring and explaining interviewer effects in centralized telephone facilities. *Public opinion quarterly* 50:251-266.
- Groves, R. M., B. C. O'Hare, D. Gould-Smith, J. Benki, and P. Maher. 2008. Telephone interviewer voice characteristics and the survey participation decision. In *Advances in telephone survey methodology*, ed. J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japac, P. J. Lavrakas, M. W. Link, and R. L. Sangster, 385-400. Hoboken, NJ: John Wiley & Sons.
- Hansen, M. H., W. N. Hurwitz, and M. A. Bershad. 1961. Measurement errors in censuses and surveys. *Bulletin of the international statistical institute* 38:351-374.
- Holbrook, A. L., T. P. Johnson, Y. I. Cho, S. Shavitt, N. Chavez, and S. Weiner. 2016. Do interviewer errors help explain the impact of question characteristics on respondent difficulties? *Survey practice* 9:no pp.
- Hox, J. J. 1994. Hierarchical regression models for interviewer and respondent effects. *Sociological methods and research* 22:300-318.
- Jaszczak, A., K. Lundeen, and S. Smith. 2009. Using nonmedically trained interviewers to collect biomeasures in a national in-home survey. *Field methods* 21:26-48.
- Kish, L. 1962. Studies of interviewer variance for attitudinal variables. *Journal of the American statistical association* 57:91-115.
- Mathiowetz, N. A. and C. F. Cannell. 1980. Coding interviewer behavior as a method of evaluating performance. *ASA Proceedings of the section on survey research methods*.
- Maynard, D. W., H. Houtkoop-Steenstra, N. C. Schaeffer, and J. van der Zouwen. 2002. *Standardization and tacit knowledge: Interaction and practice in the survey interview*. New York: John Wiley & Sons, Inc.
- Mneirneh, Z. M., R. Tourangeau, B.-E. Pennell, S. G. Heeringa, and M. R. Elliott. 2015. Cultural variations in the effect of interview privacy and the need for social conformity on reporting sensitive information. *Journal of official statistics* 31:673--697.
- Morton-Williams, J. 1993. *Interviewer approaches*. Cambridge: University Press.
- O'Muircheartaigh, C. and P. Campanelli. 1998. The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the royal statistical society, A* 161:63--77.
- O'Muircheartaigh, C. and P. Campanelli. 1999. A multilevel exploration of the role of interviewers in survey non-response. *Journal of the royal statistical society, A* 162:437-446.
- Olson, K. and 1. Bilgen. 2011. The role of interviewer experience on acquiescence.

Public opinion quarterly 75:99-114.

- Olson, K. and B. Parkhurst. 2013. Collecting paradata for measurement error evaluations. In *Improving surveys with paradata: Analytic uses of process information*, ed. F. Kreuter, 43--72. Hoboken, NJ: John Wiley & Sons.
- Olson, K. and A. Peytchev. 2007. Effect of interviewer experience on interview pace and interviewer attitudes. *Public opinion quarterly* 71:273--286.
- Olson, K. and J. D. Smyth. 2015. The effect of CATI questions, respondents, and interviewers on response time. *Journal of survey statistics and methodology* 3:361-396.
- Olson, K., J. D. Smyth, R. Horwitz, S. Keeter, V. Lesser, S. Marken, N. Mathiowetz, J. McCarthy, E. O'Brien, J. Opsomer, D. Steiger, D. Sterrett, J. Su, Z. T. Suzer-Gurtekin, C. Turakhia, and J. Wagner. 2019. *Transitions from telephone surveys to self-administered and mixed-mode surveys*. Oakbrook Terrace, IL: American Association for Public Opinion Research.
- Olson, K. and J. Wagner. 2015. A field experiment using GPS devices to monitor interviewer travel behavior. *Survey research methods* 9:1-13.
- Ongena, Y. P. and W. Dijkstra. 2006. Methods of behavior coding of survey interviews. *Journal of official statistics* 22:419-451.
- Sakshaug, J. W. 2013. Using paradata to study response to within-survey requests. In *Improving surveys with paradata: Analytic uses of process information*, ed. F. Kreuter, 171-190. Hoboken, NJ: John Wiley and Sons.
- Sakshaug, J. W., M. P. Couper, M. B. Ofstedal, and D. R. Weir. 2012. Linking survey and administrative records: Mechanisms of consent. *Sociological methods & research* 41:535-569.
- Saris, W. E. and I. N. Gallhofer. 2007. *Design, evaluation, and analysis of questionnaires for survey research*. Hoboken, NJ: John Wiley and Sons.
- Schaeffer, N. C. 1991. Conversation with a purpose - or conversation? Interaction in the standardized interview. In *Measurement errors in surveys*, ed. P. Biemer, R. M. Groves, L. Lyberg, N. A. Mathiowetz, and S. Sudman, 367-391. New York: John Wiley & Sons, Inc.
- Schaeffer, N. C., J. Dykema, and D. W. Maynard. 2010. Interviewers and interviewing. In *Handbook of survey research*, ed. P. V. Marsden and J. D. Wright, 437-470. Bingley, UK: Emerald Group Publishing.
- Schaeffer, N. c., D. Garbarski, J. Freese, and D. W. Maynard. 2013. An interactional model of the call for survey participation: Actions and reactions in the survey recruitment call. *Public opinion quarterly* 77:323--351.
- Schaeffer, N. c., B. H. Min, T. Purnell, D. Garbarski, and J. Dykema. 2018. Greeting and response: Predicting participation from the call opening. *Journal of survey statistics and methodology* 6:122-148.
- Schober, M. F. and F. G. Conrad. 1997. Does conversational interviewing reduce survey measurement error? *Public opinion quarterly* 61:576-602.
- Suchman, L. and B. Jordan. 1990. Interactional troubles in face-to-face survey interviews. *Journal of the American statistical association* 85:232-253.
- Timbrook, J., J. Smyth, and K. Olson. 2018. Why do mobile interviews take longer?

- A behavior coding perspective. *Public opinion quarterly* 82:553-582.
- Viterna, J. and D. W. Maynard. 2002. How uniform is standardization? Variation within and across survey research centers regarding protocols for interviewing. In *Standardization and tacit knowledge: Interaction and practice in the survey interview*, ed. D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, and J. van der Zouwen, 365-397. New York: John Wiley & Sons.
- West, B. T. and A. G. Blom. 2017. Explaining interviewer effects: A research synthesis. *Journal of survey statistics and methodology* 5:175-211.
- West, B. T., F. Kreuter, and U. Jaenichen. 2013. “Interviewer” effects in face-to-face surveys: A function of sampling, measurement error, or nonresponse? *Journal of official statistics* 29:277-297.
- West, B. T. and K. Olson. 2010. How much of interviewer variance is really nonresponse error variance? *Public opinion quarterly* 74:1004-1026.