Interviewers and Interviewing

Chapter · January 2010

CITATIONS
43

READS
469

3 authors, including:

Nora Cate Schaeffer
University of Wisconsin–Madison
76 PUBLICATIONS **2,214** CITATIONS
SEE PROFILE

Jennifer Dykema
University of Wisconsin–Madison
58 PUBLICATIONS **653** CITATIONS
SEE PROFILE

Chapter 13

Interviewers and Interviewing

Nora Cate Schaeffer, Jennifer Dykema and Douglas W. Maynard

13.1. Introduction

Two components of surveys are essential to the goal of describing a population: probability sampling and standardized measurement. Probability sampling allows sample estimates to be generalized to a population, and standardized measurement permits aggregation and comparison of answers provided by different respondents.

Interviewers have been integral to implementing both design elements since the first probability samples (Converse, 1987). Interviewers listed addresses in sampled areas, applied rules to select households with known inclusion probabilities, screened households to select a specific member, explained a study's purpose and details, administered questions written on paper instruments, and edited and field-coded responses. Contemporary survey designs have become more complex, but survey interviewers have essentially the same key roles now: they implement aspects of complex sample designs, locate sample members and persuade them to participate, administer long and complicated instruments, and maintain respondents' motivation through their presence. Interviewers complete these tasks so competently that researchers rely on survey modes involving interviewers (rather than self-administered instruments), notwithstanding their additional cost and potential for interviewer-related error.

Interviews are only one of the several modes of data collection and are often used together with other modes such as mail or web. Studies based on area probability sampling frames typically use face-to-face interviews, while those built around telephone frames such as random digit dialing (RDD) usually rely on telephone interviews. List frames can use either type of interview depending on whether addresses, telephone numbers, or both are available. The interviewer's role in sampling and recruitment is crucial for face-to-face and telephone studies and is the primary reason for the high response rates historically yielded by these modes. Interviewers remain an

Handbook of Survey Research, Second Edition

ISBN: 978-1-84855-224-1

Table 13.1: Typology of interviewer characteristics.

	Type of characteristic	
	Personal	**Role**
Directly observable	Race and ethnicity, gender, age, voice	Task-related verbal behavior (prescribed by training or improvised), task-related nonverbal behavior (prescribed by training or improvised), technical competence
Unobservable	Personality	Experience Knowledge or beliefs Attitudes Expectations Stereotypes

important part of many survey designs despite the increasing expense of face-to-face interviews and problems due to coverage and nonresponse in telephone surveys.

This chapter begins by introducing relevant characteristics of survey interviewers and their role. We then briefly discuss study designs and methods for studies assessing interviewer effects. Next, we summarize findings about how interviewers and interviewer-respondent interaction affect nonresponse and measurement error.[1] We briefly review practical issues of managing and supervising survey interviewers. In closing, we comment on the current and future role of interviewers in standardized surveys.

13.2. Interviewer Characteristics Influencing Survey Error

When interviewers administer surveys, their utterances and behaviors are influenced by the complex interplay among their personal characteristics and attributes, the role of interviewer, and their interaction with sample members and respondents. The typology in Table 13.1 provides a useful categorization of the attributes and behaviors of interviewers.

Interviewer characteristics vary along two important dimensions: whether they are directly observable by the respondent and whether they are components of

1. We do not discuss the impact of the interviewer on sample implementation here, but see Manheimer and Hyman (1949) and O'Muircheartaigh, Eckman, and Weiss (2002) for a description of errors that interviewers introduce into sampling procedures. We found no studies of the impact of the interviewer on screening for special populations as the last stage of sampling, although studies of screening filters within the interview suggest that some effect is likely.

the interviewer's role. Respondents can react directly to characteristics they observe. Characteristics that are not directly observable nonetheless may influence the behavior of the interviewer, which the respondent may in turn perceive. Personal attributes are relatively fixed and rarely vary across contexts. Role characteristics and behaviors develop through training and emerge during the conduct of interviews.

Observable personal characteristics include the interviewer's race, gender, age, and voice. A respondent can perceive or infer each of these, though perhaps not correctly. The interviewer's personality is personal but not directly observable. Observable role characteristics include the words interviewers say, which may be either dictated by standardization (e.g., the wording of survey questions), influenced by other aspects of training (such as instructions for probing or providing feedback), or improvised (e.g., in the course of helping respondents complete instruments such as calendars). Some nonverbal behaviors are due to training (e.g., when face-to-face interviewers wear badges to increase their legitimacy) while others are not (e.g., nodding or gesturing as the respondent talks). Role characteristics that are not directly observable include the interviewer's experience, knowledge or beliefs about the target population, attitudes toward the survey topic and questions, expectations about the difficulty or sensitivity of questions, and stereotypes about the respondent. Features of the survey design — such as the survey topic, the data collection mode, the extent and type of training, or the inclusion of incentives — also affect the interviewer's role characteristics (Campanelli, Sturgis, & Purdon, 1997; Groves & Couper, 1998).

13.3. Studying Interviewer Error

To provide background for interpreting and understanding the results of empirical studies on interviewer effects, we briefly introduce methods used in such studies. Models of survey error (see overview in Biemer, this volume) partition the total or mean-squared error (MSE) in a survey statistic (e.g., a mean or proportion) into two components:

$$\text{MSE} = \text{variance} + \text{bias squared}$$

The variance and bias squared represent errors that vary over replications of the survey and those that are constant over replications, respectively. Interviewers can contribute to either error component. For example, if interviewers vary in how they ask a question about income, some variation in respondents' answers will be due to the variation in interviewer behavior, rather than to true differences in income, which increases variance. Or if, for example, black adolescents interviewed by white interviewers give fewer correct answers to items on a vocabulary test than those interviewed by black interviewers, part of the difference between the two groups of respondents is a systematic error or bias due to interviewer race.

13.3.1. Designs

Ideally, studies seeking to estimate interviewer-related variance would assign interviewers to equivalent respondents using either a reinterview design, in which two different interviewers interview the same respondent, or an interpenetrated design, which assigns interviewers to random subsets of respondents, to ensure that assignments are equivalent in expectation.

Ideal designs for estimating interviewer variance are difficult to implement. An interpenetrating design is most feasible in a compact geographic area such as a single city or county, although designs that pair neighboring primary sampling units (PSUs) and nest interviewer assignments within the PSUs can be implemented in a national study (see Biemer, this volume). Moreover, partly because complex designs or experiments require special procedures that differ from typical survey practice, the results may not be generalizable (see Nock and Guterbock, this volume). For example, carrying out a single-site experiment with a sufficiently large number of interviewers may require recruiting interviewers who are less experienced than usual hires. Or, by randomly assigning interviewers to respondents and requiring that only the assigned interviewer contact the respondent, a design may depart from common practices such as assigning interviewers to areas close to where they live and authorizing them to request help from a supervisor or refusal converter when a selected sample person is reluctant to participate. In short, the designs and experimental procedures required for unbiased estimation of interviewer effects may reduce the generalizability of the findings.

Further complications arise if a study concerns how some feature of the measurement process affects interviewer variance. For example, in comparing highly standardized and less standardized styles of interviewing, investigators must decide how to distinguish the effect of interviewing style from the effect of the different survey instruments those styles require (e.g., standardized questions vs. less structured instruments such as event history calendars). They must also determine how to conduct equivalent training when one group uses a more demanding style of interviewing, implement manipulation checks to assess the extent to which prescribed practices are followed, devise means of recording answers in the less standardized mode that permit comparisons with those from the more standardized format, and decide on outcomes to evaluate.

Investigations without reinterview designs or interpenetration can use more realistic survey designs, such as national area probability samples and locally recruited interviewers, but must substitute statistical controls for randomization (see Groves & Couper, 1998). Analyses of recent studies of this type use hierarchical models that nest respondents within interviewers, and control characteristics of respondents that might be correlated with those of interviewers (e.g., Dijkstra, 1983; Hox, de Leeuw, & Kreft, 1991). They then attempt to explain interviewer-related variance controlling for interviewer characteristics or examine changes in substantive results when interviewer effects are controlled. Unfortunately, information that could be used to control the various dimensions on which relevant interviewer and respondent characteristics might be associated is often meager, correlated with other variables, or highly aggregated such as the mean income of a census tract (see, e.g., Schaeffer, 1980; Anderson, Silver, & Abramson, 1998a).

13.3.2. Measures and Outcomes

The impact of interviewers has been expressed in several related ways, including the proportion of variance associated with interviewers, the correlated variance due to interviewers, a design effect, or associations between interviewer characteristics (such as race) and outcomes (such as the mean on an attitude scale) controlling for characteristics of respondents or of the interviewer's caseload. Some studies estimate differences in the responses obtained by different groups of interviewers, such as the difference in answers obtained by male and female interviewers, or examine variation in interviewer-level measures, such as cooperation rates in studies of nonresponse. Others use characteristics of interviewers, such as race, or features of the interaction during the interview, such as probing, to predict outcomes presumably associated with data quality, such as agreement between respondents' self-reports about behaviors and that recorded in external records. Although "interviewer effect" in a strict sense might refer to formal expressions of interviewer variance such as the intraclass correlation ρ_{int}, the term can refer to any of these ways of describing the impact of the interviewer.

The intraclass correlation ρ_{int} — the ratio of variance between interviewers to total variance — is an especially common index for measuring interviewer effects. Together with the size of interviewers' caseloads, ρ_{int} influences the design effect $deff_{int}$. The design effect indicates the increase in variance (and resulting decrease in effective sample size) due to the interviewer effect and can be expressed as $deff_{int} = 1 + \rho_{int}(m-1)$, where m is the average interviewer caseload. A design effect of 2 cuts the effective sample size in half. Some studies examine $deff_{int}$ and others its square root, $deft_{int}$. Because they are unit free, ρ_{int} and $deff_{int}$ are comparable across questions and studies. Study implementation affects the impact of ρ_{int} on survey estimates: if the average interviewer caseload is 35, even a small ρ_{int} of 0.01 increases the variance of a statistic by about one-third. Thus, even a small value of ρ_{int} can have substantial consequences for the variance of survey estimates (e.g., Collins, 1980). (See Frankel, this volume, for an introduction to design effects and effective sample size.)

The literature on interviewer effects examines outcomes including nonresponse error (or its components, refusals and noncontacts) and measurement error (including changes in response distributions, measures of reliability, and measures of accuracy that compare survey reports to external records). It also considers other measures of data quality (such as the number of words recorded for open-ended responses), cost measures, and operational measures (such as behaviors by respondents, like breaking appointments, or behaviors by interviewers from paradata (Couper, 2005)).

13.4. Interviewers' Effects on Nonresponse Error

Interviewers are among the most important tools survey researchers have for increasing participation. Interviewers track and locate sample members, explain the purpose of a study, answer questions, and address householders' concerns (Groves et al., 2004). Interviewer-administered surveys historically have yielded higher response rates than other data collection modes, with response rates for face-to-face

surveys higher than those for telephone surveys. However, both face-to-face and telephone interviewers vary — sometimes substantially — in their ability to secure interviews (Campanelli & O'Muircheartaigh, 1999; Groves & Fultz, 1985) and this section explores factors that explain some of that variation.

Because interviewers may affect the various components of nonresponse differently — including refusals, noncontacts, and other types of noninterviews (e.g., inability to complete the interview due to language barriers) — it is important to examine these outcomes separately. However, most research in this area focuses on the impact of interviewers on persuading selected sample members to participate in the interview, and most studies report interviewer effects for cooperation rates, refusal rates, or overall response rates, which combine refusals with noncontacts. In the next section we review findings about the effects of interviewers on noncontact in general and then turn to more detailed examination of the effects of interviewer characteristics on nonresponse.

13.4.1. Effects on Noncontacts

Noncontact rates are probably similar across interviewers in a telephone facility, where the probability of contacting a phone number depends heavily on the number and the timing of attempts, primarily determined by the facility's calling procedures. Locating sample members for a face-to-face survey, however, is effortful. Although area probability frames select sample members in clusters to minimize travel time, interviewers may still need to travel substantial distances to locate their sample. For list samples of mobile populations or those without stable residences, such as youth in foster care, tracking and locating respondents can be difficult and require multiple attempts.

Pickery and Loosveldt (2002) explored the relationship between refusals and noncontacts in personal interviews from the Belgian Election Studies. They found a positive association between the two outcomes: interviewers who got fewer refusals also obtained fewer noncontacts, but the significance and strength of the relationship was tempered after removing an outlier. O'Muircheartaigh and Campanelli (1999) obtained the same pattern of results in the second wave of a panel study. Thus, there is some, but limited, evidence that face-to-face interviewers who work hard at locating sample members also work hard at recruiting them.

13.4.2. Characteristics of the Interviewer

13.4.2.1. Observable personal characteristics Although a number of studies have examined the impact of interviewers' social characteristics on nonresponse, the main effects of these characteristics on participation appear to be small. Often interviewer characteristics that are significantly related to nonresponse in bivariate comparisons are reduced or eliminated when other variables are controlled

(e.g., Campanelli & O'Muircheartaigh, 1999). This suggests that interviewers' social characteristics may be related to other factors that affect nonresponse, such as the interviewer's experience or the characteristics of the respondents they are assigned to interview, rather than shaping nonresponse directly. Groves and Couper (1998) predict that interviewers' social characteristics are most likely to exert effects by interacting with other variables, as happens when interviewers and respondents are matched on race or gender (see Section 13.5), but that even significant interactions are likely to be explained or mediated by other variables. None of the studies we reviewed involved an experimental design that matched interviewers and respondents randomly with respect to social characteristics and explored effects on nonresponse.

Several studies have examined the influence of interviewer race, gender, age, or voice. Surveys, particularly in-person surveys, sometimes match the race of the interviewer and respondent. This may be unintentional, an artifact of segregation and the practice of hiring local interviewers and assigning them to areas close to where they live (see Groves & Couper, 1998), or intentional, in the belief that it will increase participation or improve response validity — particularly for surveys on sensitive or threatening topics (but see "Race and ethnicity" under Section 13.5). Only nonexperimental studies have examined the impact of the race of the interviewer on participation, and none reported a significant effect (Merkle & Edelman, 2002; Singer, Frankel, & Glassman, 1983). Merkle and Edelman tested for, but did not find, an interaction between the race of the interviewer and that of the respondent in their analysis of participation in exit polls from the nineties. Singer et al. found no effect of the interviewers' race on screening or response rates in a telephone survey of leisure activities that included sensitive questions on alcohol and drug use.

Female interviewers are often described by respondents as possessing attributes that might increase householders' willingness to participate in surveys, such as friendliness (Fowler & Mangione, 1990) and approachability (Morton-Williams, 1993). Most studies do not find that these attributes result in substantially higher participation (see Lessler & Kalsbeek, 1992; Pickery & Loosveldt, 2002); but those studies that do find a significant difference report slightly higher response rates among female than male interviewers (Campanelli & O'Muircheartaigh, 1999; Hox & de Leeuw, 2002). For example, in their cross-national study of 9 countries and 32 interviewer-administered surveys, Hox and de Leeuw found a significant but small gender effect that favored female interviewers. A finding by Groves, O'Hare, Gould-Smith, Benki, and Maher (2008) is complementary: telephone interviewers with voices rated as sounding less masculine obtained higher response rates.

While some studies find no relationship between interviewer age and nonresponse (e.g., Morton-Williams, 1993; Pickery & Loosveldt, 2002), the preponderance of evidence suggests that older interviewers obtain slightly higher response rates (Campanelli & O'Muircheartaigh, 1999; Hox & de Leeuw, 2002; Merkle & Edelman, 2002; Singer et al., 1983). Future studies are needed to control for confounding variables such as interviewer experience. Unknown is how respondents perceive interviewers' age or how such perceptions vary by the respondent's age, questions particularly relevant for telephone studies.

Voice quality includes such features as loudness, pitch, and speech rate. Research about voice quality poses many technical and analytical challenges and the small number of studies examining its relationship to participation is inconclusive. In examining why interviewers in a panel study of consumer attitudes consistently achieved different response rates, Oksenberg, Coleman, and Cannell (1986) reported lower refusal rates among interviewers rated as speaking more quickly, loudly, and distinctly, and in a higher pitch. Later investigation, which used ratings of judges to simulate the perceptions of respondents, demonstrated that participation is related to the perceived rather than the actual acoustic properties of interviewers' voices (van der Vaart, Ongena, Hoogendoom, & Dijkstra, 2006). The most recent study of interviewers' voice characteristics found no meaningful relationships between participation and either ratings by listeners or the actual acoustic properties of interviewers' voices (Groves et al., 2008).

13.4.2.2. Unobservable personal characteristics Researchers have examined the effects of several unobservable personal characteristics of interviewers on participation, including experience, personality, knowledge, attitudes, and expectations. Empirical evidence supports the conclusion that interviewing experience is positively related to cooperation for both face-to-face and telephone surveys (Couper & Groves, 1992; Durbin & Stuart, 1951; Groves & Couper, 1998; Groves & Fultz, 1985; Hox & de Leeuw, 2002; Pickery & Loosveldt, 2002). For example, Groves and Couper (1998) reported that interviewers with five or more years of experience were better able to anticipate and overcome such barriers as negative attitudes towards surveys and delaying tactics (e.g., "I'm too busy"). Limited evidence also suggests that the interviewer's level of experience affects the rate of noncontacts. Results from the Family Resources Survey demonstrated that more experienced interviewers achieved lower noncontact rates by working harder and making more calls to households (Campanelli et al., 1997).

Two important methodological challenges must be borne in mind when interpreting findings about interviewers' experience. The first is selection: effects could be due to experience per se or to the fact that less successful interviewers are more likely to quit or be let go. The second concerns the measurement of experience: as duration (e.g., total number of years as an interviewer or at a particular organization) versus breadth (e.g., number of distinct surveys worked) (Groves & Couper, 1998). Studies that find no relationship of experience to nonresponse often measure it using indicators other than years as an interviewer. For instance, no effect — or even a negative relationship — of experience with response rates was found with experience measured as the number of organizations for which interviewers had worked (Groves & Couper, 1998) or the number of surveys on which they had worked (Merkle & Edelman, 2002).

Research exploring the relationship between personality measures (e.g., self-monitoring) and success in obtaining cooperation has not found strong effects (see summary in Groves & Couper, 1998). In contrast, some interviewers possess knowledge of the target population or cultural knowledge about the target population's language and customs (de Leeuw & Collins, 1997; Groves & Couper, 1998), which

probably increases response rates. Interviewers vary in their expectations about their ability to secure interviews in ways that are related to nonresponse, although it is impossible to determine whether such associations are causal. For example, confidence in their ability to secure an interview is positively related to interviewers' ultimate success in doing so (Campanelli et al., 1997; Groves & Couper, 1998; Singer et al., 1983). In summary, interviewers who report more positive attitudes about persuasion and persuasion strategies, express greater belief in confidentiality, ascribe to the importance of refusal conversions, and express "willingness to proceed as usual in the face of obstacles" tend to obtain higher response rates (Groves & Couper, 1998; Hox & de Leeuw, 2002).

13.4.3. Methodological Issues

The number of studies examining how interviewers affect sample persons' participation in surveys is relatively small, for both methodological (Groves & Couper, 1998) and logistical reasons. First, even when they exist, interviewer effects can be statistically difficult to detect because of the small number and relative homogeneity of interviewers in any given study, which make it difficult to determine the impact of interviewer characteristics. Although the composition of interviewing staffs is rarely documented, informal observation suggests that interviewers in a particular survey organization and a specific mode (face-to-face or telephone) may be fairly homogenous with regard to race, gender, and age and possibly other attributes, such as personality (e.g., those who choose to be interviewers and survive in the job may tend to be outgoing). Second, geography-based assignments of cases to interviewers in face-to-face studies conflate effects of interviewers with the effects of the areas they work. This is particularly problematic for studies of nonresponse because it is harder to locate and secure interviews with respondents in some geographic areas, such as cities. Other problems mentioned previously — the selection processes operating on the pool of interviewers and tension between the controls needed for strong designs and realistic field practices — can also pose problems or distort inferences. Third, the composition of nonrespondents and the causes of nonresponse may vary over populations and over time, particularly as response rates decline. Thus, some results that appear inconsistent may be due to undocumented or unanalyzed differences in the processes leading to participation.

13.4.4. Interaction in the Survey Introduction

Once interviewers locate a household, by phone or in-person, they must introduce themselves and the study and attempt to secure an interview. This entire interaction occurs very quickly — usually in under a minute in a telephone interview (Oksenberg, Coleman, & Cannell, 1986) and in five minutes or less in a face-to-face interview (Groves & Couper, 1994). There is good reason to expect that aspects of this interaction affect respondents' decisions to participate.

Research on interviewer-respondent interaction during survey introductions can be grouped into three phases. The earliest research focused on how manipulating the content and length of scripted introductions affected cooperation. Later research drew on the social psychological literature on compliance, helping behavior, and attitude change to predict different conditions under which interviewers would be more effective in getting respondents to cooperate. Current efforts focus largely on exploring effects of the actual interaction between interviewers and respondents on nonresponse, specifically on techniques known as "tailoring" and "maintaining interaction."

13.4.4.1. Standardized survey introductions Survey introductions vary in their content (e.g., whether they mention the sponsor's name, refer to sample selection procedures, provide example questions, include likely interview length, or address confidentiality concerns), the amount of information they provide (e.g., the level of detail about the survey's topic), and how scripted they are (e.g., whether interviewers are required to read a script or are allowed to improvise).

In an early telephone survey, Dillman, Gallegos, and Frey (1976) found that neither personalization (mentioning the sample member's name), varying the study's description, nor offering to share results was linked to significant differences in response rates. O'Neil, Groves, and Cannell (1979) varied what the interviewer said after interviewers delivered a standard introduction: one condition immediately presented example questions, another provided verbal feedback, and a third gave a lengthy description of the survey organization. Although response rates differed significantly among some conditions, the overall results were not readily interpretable. Oksenberg and Cannell (1988) observed that although some interviewers consistently obtained higher participation in a telephone survey of consumer sentiments, response rates did not improve when underperforming interviewers used introductions modified from higher-performing ones. Singer and Frankel (1982) reported no significant differences in response rates among experimental groups when interviewers presented different amounts of information about the study's content or informed respondents about the study's purpose.

An interesting extension of this line of research (Houtkoop-Steenstra & van den Bergh, 2000) tested differences in response to a telephone survey after four different introductions. Three of them differed in the amount and content of scripted material, while the fourth included an agenda that permitted interviewers to describe features of the survey using their own words. Results confirmed previous findings about the ineffectiveness of scripted introductions: interviewers using the agenda-based introduction secured higher levels of participation than those using scripted introductions. These results replicate those of Morton-Williams (1993) who tape-recorded doorstep introductions and found that requiring interviewers to follow a script significantly decreased cooperation.

Overall, results from experimental studies that vary the content and amount of scripted information read to sample members have not shown strong effects. Few include manipulation-checks or direct observation of interviewers' behavior, however, which tempers confidence in their results. This is particularly unfortunate

because the behavior of selected sample members during introductions may require quick responses from the interviewer, who may then deviate from the experimental procedure. Research either assumes that the manipulation was implemented as intended or asks interviewers for post-interview reports about their behavior. Morton-Williams'(1993) finding that interviewers did not follow scripts even when recorded suggests that assuming that interviewers follow the script may not be wise. Recording of actual survey introductions to use in a manipulation check was difficult when these studies were conducted, but this is no longer difficult.

13.4.4.2. Social psychology and survey introductions Drawing on social psychological principles of compliance, helping, and attitude change, Groves, Cialdini, and Couper (1992) developed a framework for predicting which techniques in an interviewer's request to participate in a survey will be more effective. They describe ways that interactional partners can increase compliance with a request — through appeals to authority, reciprocation, social validation, scarcity, consistency, and saliency — and derive predictions about which techniques will obtain higher cooperation. For example, applying the principle of reciprocation (e.g., individuals will be more likely to comply with a request that involves a concession or gift) to a survey context, predicts that interviewers who counter respondents' refusals to complete long questionnaires by reducing the scope of the task (e.g., asking respondents to complete a few questions for nonresponse adjustments) will likely persuade respondents to complete the smaller task.

Analyzing spontaneously occurring persuasion attempts, Dijkstra and Smit (2002) reported that telephone interviewers who addressed respondents' concerns about interview length by mitigating ("maybe we can do it faster") or decreasing ("I'm sure we can manage it within ten minutes") had fewer refusals. However, they found no support for appeals to authority (e.g., refusal rates were higher after appeals to university sponsorship than personal appeals) or social validation (e.g., appeals that "many" people are needed or that most people enjoy the interview). Hox and de Leeuw (2002) examined self-reported interviewer behaviors and also found that appeals related to social validation actually decreased response rates. In sum, there is little empirical support for the effectiveness of the social psychological techniques from either interviewers' self-reports of their use or direct analysis of interviewer-respondent interactions. Groves and Couper (1998) assert that this lack of support highlights the importance of examining what interviewers and respondents actually do. Research currently underway does just that (e.g., Maynard, Freese, & Schaeffer, 2008).

13.4.4.3. Tailoring and maintaining interaction Results from focus groups with interviewers about their techniques for securing interviews (reported by Groves et al., 1992 and replicated by Snijkers, Hox, & de Leeuw, 1999) and analysis of audio-recordings of interviewer-respondent doorstep interactions (Morton-Williams, 1993) suggest that the techniques of "tailoring" and "maintaining interaction" might promote successful recruitment of sample members. Tailoring is a technique employed by expert interviewers who respond to cues in their immediate setting — verbal, nonverbal, and visual — to produce utterances and behavior that respond to

sample persons' utterances and behavior. To be effective at tailoring, an interviewer must be able to readily deploy a large and varied repertoire of persuasion techniques (Maynard & Schaeffer, 2002b). When maintaining interaction, interviewers continue engaging respondents in conversation to obtain more information for tailoring and reduce the likelihood that the respondent will refuse to participate at a given turn in the interaction. Tailoring and maintaining interaction are reciprocal: Interaction must be maintained in order for tailoring to occur (see Groves & Couper, 1998 for a comprehensive overview).

Operationalizing and measuring these techniques is difficult, but there is support for their importance. For example, following every contact with a household, Groves and Couper (1994, 1996, 1998) had interviewers record details about their interaction with the sample person such as whether the contact person commented on time (e.g., "I'm too busy"), made negative comments about surveys (e.g., "I don't trust surveys"), or asked questions (e.g., "What's the purpose of the survey?"). Subsequent analyses found that negative comments about surveys and time-delay statements (about not having time to complete the survey) were associated with nonresponse, while asking questions was positively related to responding. Further, interviewers who changed their behavior in response to a refusal during their initial contact with the household were more likely to secure participation in a subsequent contact, but not significantly so. Couper and Groves (2002) analyzed recorded contacts, suggesting that tailoring can increase participation when respondents make negative or time-delay statements. To replicate many of these findings, Campanelli et al. (1997) analyzed tape recordings from doorstep introductions by face-to-face interviewers. Interviewers who tailored their participation appeals to the respondent and his or her concerns were more effective in converting respondents who were initially undecided. Dijkstra and Smit (2002) found only moderate support for the effectiveness of tailoring in their analysis of interaction in a telephone survey, however. Although response rates were higher among interviewers who tailored when compared to those who made no attempts at persuasion, tailoring was not more effective than other persuasion attempts. Dijkstra and Smit also studied interviewer attempts to maintain interaction through behaviors like repeating a sample member's utterances or posing questions. Attempts that simply prolonged interaction did not lead to more cooperation; instead, the content of the interviewers' utterances, especially those that provided information to sample members, was effective.

Maynard and Schaeffer (1997) explored how requests for participation were organized between interviewers and respondents, examining transcripts from recorded telephone introductions using conversation analysis. They described how both interviewers and respondents work to maintain interaction. Based on analyses, the authors speculated that interviewers who used standard (untailored) responses to address respondents' concerns and questions were not effective. Using similar analytical techniques, Maynard and Schaeffer (2002a) found that interviewers often work optimistically and effectively in responding to sample members' questions about the survey's length and purpose. But such optimism was not effective when responding to sample members who resisted proposals to be called back or who were not interested. These latter findings suggest that persuasion attempts are ineffective for a set of "hard core" refusers (see also Campanelli et al., 1997).

Groves and McGonagle (2001) provide the most plausible evidence to date on the effectiveness of tailoring and maintaining interaction. They conducted focus groups with experienced interviewers to cull concerns raised by respondents, organized the concerns into themes, and developed responses to the concerns. From these efforts study-specific nonresponse training was developed for interviewers working on two different surveys. Interviewers practiced using the themes, concerns, and responses to quickly analyze and address respondents' concerns using their own words. Results demonstrated that the training — possibly because it improved tailoring or for other reasons — produced significantly higher response rates, with the rise in rates concentrated among interviewers who were less successful before the intervention. For other research about training in avoiding refusals see Cantor, Allen, Schneider, Hagerty-Heller, and Yuan (2004) and O'Brien, Mayer, Groves, and O'Neill (2002).

Interest in increasing interviewers' competence in recruiting respondents has led to experimentation with virtual human technology, developing realistic simulations that allow interviewers more practice (Link, Armsby, Hubal, & Guinn, 2006). Such forms of training will become more effective as we continue to understand which behaviors of interviewers actually secure participation (Maynard et al., 2008).

13.4.5. Interviewer Effects on Nonresponse in Longitudinal Surveys

For longitudinal surveys, the central additional question about interviewers and nonresponse is whether using the same interviewer from the previous wave increases retention. Some early and nonexperimental studies showed increasing respondent attrition when interviewers changed between waves (see summary by Campanelli & O'Muircheartaigh, 1999), but it is likely that other variables affecting attrition were related to whether or not the interviewer changed. Campanelli and O'Muircheartaigh (1999) addressed these methodological shortcomings by employing an interpenetrated sample design that randomly assigned households to interviewers within groupings of PSUs for the second wave of the British Household Panel Survey (BHPS). They found variation in refusal rates (but not noncontact rates) among interviewers, but in contrast to previous research, none of the variation was due to interviewer continuity: having the same interviewer in both waves was unrelated to participation in the second wave.

Evidence also indicates that the effect of the subsequent interviewer may be small or insignificant, net of the effects of the interviewer who initially convinced the respondent. Pickery, Loosveldt, and Carton (2001) analyzed participation in the second wave of the Belgian Election Studies, which did not randomly assign respondents to interviewers. They found that the first-wave interviewer-respondent contact was more important than that in the second wave for predicting retention in the panel, presumably because the initial interviewer created a "positive or pleasant experience" for the respondent that carried over to the next wave (see also Nederhof, 1987). Lepkowski and Couper (2002) developed a model of factors to use in predicting retention in the second waves of America's Changing Lives Survey and the National Election Surveys. They showed that interviewers' assessments of

respondents' survey experience at Wave 1 — such as comments that the respondent "enjoyed the interview," was "too busy," or demonstrated "reluctant behavior" — predicted cooperation at Wave 2 (see also Kalton, Lepkowski, Montanari, Maligalig, 1990). These findings complement those of Groves and Couper (1994, 1996) and Campanelli et al. (1997) who reported that negative comments by respondents during an initial interviewer contact in a cross-sectional study predicted subsequent refusal.

13.5. Interviewers' Effects on Measurement Error

The interviewer's role in measurement is both relied on and viewed with skepticism and even suspicion. For some researchers, "interviewer effect" refers specifically to interviewers' contribution to variance as measured in studies of face-to-face and telephone interviewing. But researchers have also examined in other ways how characteristics of the interviewer and of the interaction between the interviewer and respondent affect measurement error.

13.5.1. Interviewer Variance and Mode of Interview

Early studies that estimated interviewer variance for face-to-face interviews were influential in the development of survey practice. When telephone interviewing was introduced, researchers were quick to assess the level of interviewer variance in this new mode.

Groves (1989, p. 365) summarized details from seven studies (published between 1962 and 1985) that used face-to-face interviewing, and he reported that mean ρ_{int} for these studies ranged from 0.005 to 0.102, with half the studies having mean ρ_{int} greater than 0.02. In a more recent experiment about interviewers in the BHPS that used interpenetrated sample units, O'Muircheartaigh and Campanelli (1998) found that the distribution of ρ_{int} for about 820 variables was roughly comparable to that of ρ_s (the intraclass correlation due to clustering in the sample design) and that 3 in 10 of the estimates of ρ_{int} were significantly different from 0 ($p < 0.05$) compared with 4 in 10 of the estimates of ρ_s. In their study (1999–2000) of crime victimization in Germany that used interpenetrated samples, Schnell and Kreuter reported a median design effect ($deft_{int}$) of roughly 1.4, with most of the design effects attributed to the interviewer rather than to the clustering of the sample (2000, p. 90, 2005, p. 400). Similarly, in their analysis of fear of crime in the British Crime Survey, Brunton-Smith and Sturgis (2009) report that the impact of clustering within interviewer is greater than that of geographic clustering.

In centralized telephone surveys, ρ_{int} should be lower than in face-to-face designs because interviewers are supervised and monitored more closely than field interviewers are, and because instruments used in face-to-face surveys are often more complicated than those used in telephone surveys. Groves and Magilavy (1986) reported a range from 0.0018 to 0.0184 in several studies that used centralized telephone interviewing, with half of the studies having mean ρ_{int} greater than 0.0086. But assessing the consequences of even these small values of ρ_{int} requires taking the size of the caseload of the interviewers into account, and telephone interviewers often have large caseloads. For the studies Groves and Magilavy summarized, the impact of these interviewer effects ranged from essentially none to an approximate doubling of the variance.

Although they cannot be present in self-administered instruments delivered by mail, interviewer effects have sometimes been observed in self-administered instruments completed in the presence of interviewers (Campbell, 1981; O'Muircheartaigh & Campanelli, 1998, p. 69), but are not always present (see Tourangeau, Rasinski, Jobe, Smith, & Pratt, 1997), a finding with potential implications for computer-assisted self interviewing (CASI) and its variants audio-CASI and video-CASI (see Krysan & Couper, 2003).

Because there are relatively few studies of interviewer variability and some date from earlier regimes of interviewing practices or used poorly documented interviewing practices, it is difficult to draw strong conclusions about which features of survey design and implementation affect interviewer variance, the circumstances under which those features will exert their effects in current survey implementations, or the impact of interviewer effects on substantive models. Nevertheless, research suggests that a lower "presence" of the interviewer (but see discussion in Krysan & Couper, 2003) is associated with lower interviewer variance. Although studies of interviewer variance will always be few because of the difficulties they pose, it is important that they continue because the complexity of studies and the practices of interviewing have changed over time and will continue to do so.

13.5.2. Characteristics of the Interviewer

Some attempts to locate the sources of interviewer effects on survey responses examine the impact of observable personal characteristics (race and ethnicity, gender, age, and voice), while others consider unobservable role characteristics (experience and expectations). Because the behavior of the interviewer is so contingent on both task structure and respondent behavior, we discuss research about the effect of the interviewer's behavior in a separate section ("Interviewer Effects and Survey Questions").

To foreshadow conclusions of the research summarized later, interviewer and respondent characteristics are social facts and come into play when interviewers and respondents interact. An interviewer characteristic is most likely to affect responses to questions that make the characteristic salient or relevant in the interaction, activate stereotypes, or evoke the respondent's concerns with affiliation, relative status, or deference. Effects do not always appear for such questions, however, and sometimes appear for questions without these features.

13.5.2.1. Race and ethnicity Early studies based on local (e.g., city) samples documented that the race of the interviewer sometimes had a substantial influence on answers given by black respondents. These effects could be interpreted as reflecting social circumstances (e.g., larger effects among rural than urban blacks), but the

effects were not all large or easily interpreted (see summary in Hyman, 1975). A response to this finding was to attempt to match the race of interviewer and respondent in studies of blacks' racial attitudes (Schuman & Converse, 1971, p. 44). This strategy assumes that such matching increases response validity, but makes it impossible to estimate the impact of the interviewer's race or control it in analysis. Schuman and Converse's Detroit Area Study (DAS) about racial attitudes among blacks used 25 professional black interviewers and 17 white graduate student interviewers and randomly assigned clusters to interviewers of each race. They concluded that race-of-interviewer effects were restricted mostly (but not entirely) to questions measuring militancy and hostility toward whites: levels were higher among respondents interviewed by black interviewers and effects were largest among respondents with less education or income; overall, interviewer race explained about a quarter of the variance in a 12-question scale measuring hostility. Large effects were also found for a question about entertainers, which suggested that mechanisms other than simple deference or ingratiation could produce such effects. Subsequent investigation using DAS data demonstrated that whites were also vulnerable to race-of-interviewer effects (Hatchett & Schuman, 1976). Schaeffer (1980), taking into account the clustering of respondents within interviewers and controlling for characteristics of interviewers' assignments to the extent possible, found effects for both black and white respondents in the General Social Survey (GSS), a national study, as did Cotter, Cohen, and Coulter (1982) in a telephone survey that used a local sample. Anderson et al. (1988a) reported that false claims of voting by black respondents were more likely when the interviewer was black rather than white, demonstrating that matching interviewer and respondent characteristics could decrease validity. In contrast, designs that cross interviewer and respondent race do not assume that racial matching increases validity, allow the possibility that the race-of-interviewer effect in itself reflects the climate of race relations in the population, and take on the task of documenting when that climate affects survey answers (e.g., Callegaro, DeKeulenaer, Krosnick, & Daves, 2005; Hill, 2002; Wolford, Brown, Marsden, Jackson, & Harrison, 1995).

Another line of studies considers how interviewer race affects substantive conclusions and considers the influence of interviewer ethnicity in addition to that of race (e.g., Hurtado, 1994). For example, Schaeffer (1980) described changes in the racial composition of the GSS interviewing staff over time, conjecturing that they might lead to apparent trends in racial attitudes, a speculation subsequently confirmed in the National Election Surveys (Anderson, Silver, & Abramson, 1988b; see also Krysan, 2002). Analyzing the National Black Election Survey, Davis (1997) argued that race-of-interviewer effects biased estimates of the relationship between racial consciousness and support for Jesse Jackson. Differences in interviewer effects by ethnicity affected group comparisons in data from New Zealand (Davis & Scott, 1995). For other examples, see Finkel, Guterbock, and Borg (1991) and Hox et al. (1991).

Developments in video-CASI technology make it possible to control the race of a virtual interviewer who appears on a laptop screen, regardless of the race of the interviewer who presents the laptop to the respondent. In their laboratory study, Krysan and Couper (2003) found that when effects occurred, black respondents

expressed more racial liberalism to black interviewers, whether virtual or live. White respondents, however, were more racially conservative to virtual black than virtual white interviewers. This might be due to the activation of racial stereotypes, as might the answers by white respondents to the question about entertainers reported by Schuman and Converse. The activation of racial stereotypes affecting cognition is an additional possible mechanism for race-of-interviewer effects that complements suggestions of other investigators that interviewer-respondent race differences may induce "test anxiety" when answering questions about knowledge (e.g., Davis & Silver, 2003). If black respondents are more vulnerable to such race-of-interviewer effects and most interviewers are white, black-white test score differences would be overestimated (Huang, 2009).

13.5.2.2. Gender The only gender-of-interviewer effect that Groves and Fultz (1985) found in their analysis of the Survey of Consumer Attitudes was that respondents expressed greater optimism about their economic situation to male telephone interviewers. Johnson and Moore (1993) found no effect of interviewer gender in telephone interviews about the sale and use of pornographic materials. In face-to-face interviews, O'Muircheartaigh and Campanelli (1998) found a significant relationship between interviewer variance and the gender of the interviewer for fewer than 10 percent of the questions with significant interviewer variation that they examined. In contrast, Kane and Macaulay's (1993) analysis of a national telephone survey found that depending on interviewer gender, men expressed different opinions about gender inequality in employment while women varied in their support for collective action and group interests (see also Huddy et al., 1997). Like race-of-interviewer effects, gender-of-interviewer effects can be considered a social fact. How respondents react to interviewer gender may vary not only by the topic of the question, but by geographic or cultural region. For example, effects may be stronger in rural than urban areas, but may balance at the national level (Flores-Macias & Lawson, 2008).

13.5.2.3. Other observable personal characteristics Singer et al. (1983) found no significant effects of interviewer age on the quality of responses or on item nonresponse rates. In contrast, O'Muircheartaigh and Campanelli (1998) found some significant variance linked to interviewer age for about a quarter of those questions determined to have significant interviewer variance. The age homogeneity of a given interviewing staff may limit the ability of researchers to detect influences of interviewer age.

We found only two studies that examined the impact of the interviewer's voice on measurement: In a field experiment (which lacked a manipulation check), Barath and Cannell (1976) found that "yes" answers in a check list were more common if the interviewer read the items with a rising as compared with a falling intonation. This finding was later contradicted by data from a national survey (Blair, 1977), which found that coders' ratings of intonation agreed in only 83.5 percent of the cases examined and that agreement was more likely when intonation was more pronounced.

13.5.2.4. Unobservable role characteristics A handful of studies examine whether unobservable interviewer role characteristics affect measurement. Experienced interviewers obtained lower reports of substance use on self-administered instruments (Hughes, Chromy, Giacoletti, & Odom, 2002) and had higher item nonresponse for income questions (Bailar, Bailey, & Stevens, 1977) than inexperienced interviewers. (For other negative effects of experience see O'Muircheartaigh & Campanelli, 1998; van Tilburg, 1998.) As interviewers conduct more interviews within a given survey, they may conduct interviews more quickly and perceive respondents as less interested (Olson & Peytchev, 2007).

Interviewer expectations have been a source of concern since early surveys (Hyman, 1975). Interviewers who expect that asking threatening questions will be difficult may obtain higher item nonresponse or lower levels of reporting (Sudman, Bradburn, Blair, & Stocking, 1977; Singer & Kohnke-Aguirre, 1979; Singer et al., 1983). Interviewer expectations are undoubtedly affected by their tenure and type of experience, both in total and with a given study, but this has not been studied.

13.5.2.5. Methodological issues The concept of the "presence" of an interviewer during an interview needs substantial refinement (Krysan & Couper, 2003). This is clear from theoretical and technological developments, as well as empirical results: effects sometimes appear in telephone and self-administered as well as face-to-face data collection. In addition, differences in the characteristics of field and telephone interviewers and changes in staff composition over time are largely undocumented and make it difficult to generalize from these studies. For example, although our informal observations suggest that field staff in the 1970s and 1980s were predominantly middle-aged women, telephone interviewers may currently be more balanced in terms of gender (as they are at the University of Wisconsin Survey Center), if not age.

13.5.3. Interviewer Effects and Survey Questions

Because the size of interviewer effects varies over survey questions, it is important to understand how the survey topic and questionnaire design influence interviewer effects. Observations and judgments by interviewers for topics as varied as the respondent's skin color or whether the respondent's living room was carpeted have shown substantial interviewer effects (Gray, 1956; Hill, 2002; see also Fecso, 1991). Using nonexperimental data, but controlling for objective indicators of the respondent's political knowledge, Leal and Hess (1999) documented that interviewers' ratings of respondents' political knowledge were influenced by respondents' social characteristics. This resembles earlier findings discussed by Hyman (1975) about the impact of interviewers' expectations for consistent answers. Such effects appear common: O'Muircheartaigh and Campanelli (1998, p. 69) documented significant interviewer effects for 79 percent of items based on interviewers' observations.

Interviewer effects probably vary by question topic. Bailey, Moore, and Bailar (1978, p. 20), for example, reported greater interviewers' influence on the variability

of estimates of assaultive violence without theft than those of assaultive violence with theft. Bailar, Bailey, and Stevens (1977, p. 339) reported substantial interviewer effects on item nonresponse for income. A later study that used both face-to-face and telephone interviews, however, reported only a small intraclass correlation for item nonresponse more generally in a study of well-being (Hox et al., 1991, p. 449). van Tilburg (1998) found interviewer effects on measured social network size for a sample of elderly respondents in the Netherlands, while Marsden (2003) reported similar effects with a simpler instrument and a GSS sample.

Collins (1980) argued that apparent interviewer effects sometimes result from how interviewers and respondents use the instrument. We suspect that this is fairly common. (See Fuchs, 2002 for an experiment about household roster grids showing how instrument design affects interviewer and respondent behavior.) That is, in addition to a question's content, instructions and design features of questions can moderate or amplify interviewer effects. Effects can occur, for example, if an instrument requires interviewers to determine if respondents are "blind or partially sighted" or "deaf or hard of hearing" or requires respondents to volunteer that a question does not apply to them (Collins, 1980, pp. 86, 89). In such cases, a standardized solution is use a filter question to route all respondents through the instrument.

These examples suggest the need to examine systematically how interviewer effects vary by characteristics of questions. The requirements for such studies are daunting, as suggested earlier. Early studies reported that open-ended questions showed interviewer effects (e.g., Gray, 1956; Collins, 1980); and research indicated that the number of mentions to open items in particular could be affected (Hox et al., 1991, pp. 449–450; Groves & Magilavy, 1986, p. 260). Mangione, Fowler, and Louis (1992) refined the discussion by examining specific features of questions, using a design in which each interviewer's assignment was a probability subsample of the total sample. They coded 130 questions into contrasting categories (open vs. closed, sensitive vs. not, opinion vs. factual, and difficult vs. easy). About 30 percent of the questions had significant interviewer effects (1992, p. 301). Drawing conclusions from their analysis is complicated by the relatively small number of questions, overlap among the characteristics of the items, and the possibility of interaction effects. However, they found some evidence that interviewers could have more of an impact on answers to difficult questions or to those that are both open and nonsensitive. Their strongest argument was that interviewer effects arise from how questions affect interviewer-respondent interaction (a finding we return to later).

Schnell and Kreuter (2005) applied the Mangione et al. (1992) categories to 118 questions about crime. They found that interviewer effects were larger for sensitive questions than nonsensitive ones and for nonfactual than factual ones. Their results also suggested greater interviewer effects for open than for closed questions and for difficult than easy ones (though the last difference was quite small). Schnell and Kreuter constructed an index of "harmful" question properties ranging from 0–4 and found that interviewer effects increased with the index (2005, pp. 403–404).

Analyzing 820 questions in the BHPS, O'Muircheartaigh and Campanelli (1998, p. 69) found significant interviewer effects for about a quarter of attitude questions (slightly more for Likert scales and slightly fewer for other attitudinal questions),

factual, and "quasi-factual" questions. Examining whether features of questions (threatening or not, difficult or not) mediate interviewer effects on nonresponse, Pickery and Loosveldt (2001) found pervasive interviewer effects on nonresponse of all types; question difficulty helped to explain interviewer effects on item nonresponse.

These studies — all of which use face-to-face interviews and standardized administration — illustrate well the challenges of isolating the role of the instrument in creating interviewer effects. Apart from the cost of such experiments, designing effective studies is hampered by rudimentary systems for classifying survey questions (about which consensus may be low, as Mangione, Fowler, and Louis note).

13.5.4. Interaction in the Survey Interview

Sudman and Bradburn (1974) pointed out long ago that in the social situation of the survey interview the three principals are the interviewer, the respondent, and the task. The interaction among them is the mechanism creating many, if not most, interviewer effects. Most contemporary survey interviews attempt to promote production of the "paradigmatic" interaction sequence in which the interviewer asks a scripted question, the respondent gives a properly formatted answer, and the interviewer acknowledges the answer and moves on (Schaeffer & Maynard, 1996). The standardized style of interviewing was designed to reduce interviewer effects and increase the reliability of survey data, although critics worry that it may reduce the validity of measurement (Suchman & Jordan, 1990; Schaeffer, 1991). Because telephone interviews can be recorded easily, much of what we know about interaction in the interview is based on telephone interviews or laboratory studies. Recording interviews on laptops is now feasible and we expect more studies of interaction in face-to-face interviews in the future.

13.5.4.1. Standardization and styles of interviewing Standardization refers to a set of practices commonly used in surveys conducted to describe a population. Practices in early survey interviews were probably very different from current methods of standardization. Converse (1987) described practices that would now be termed "field coding," in which interviewers listened to a respondent's answer and selected an answer category that seemed closest. Later practices summarized by Hyman (1975, p. 84) were similar: interviewers should read questions exactly as written and in the order in which they appear, record answers verbatim, and "code a reply in the answer box that most nearly corresponds to the actual words" to reduce biases arising from the interviewer's expectations. The National Opinion Research Center's "Basic Instructions to Interviewers" (Williams, 1942) read much like modern standardization, but practice at the time appears to have used field coding, questions about social characteristics were not scripted, and social characteristics were sometimes inferred by interviewers (Converse, 1987, pp. 332–335). Field coding requires interpretation by interviewers, particularly when used with subjective questions. It is thus not surprising that early models of survey error gave significant attention to the interviewer (e.g., Hansen, Hurwitz, & Bershad, 1961).

Most contemporary survey researchers would probably accept the principles of standardization described by Fowler and Mangione (1990, p. 35): "1. Read questions as written. 2. Probe inadequate answers nondirectively. 3. Record answers without discretion. 4. Be interpersonally nonjudgmental regarding the substance of answers." Although reading questions as worded appears to be a foundation for all varieties of standardization, organizations vary substantially in how they operationalize all the principles (Viterna & Maynard, 2002). For example, if the respondent volunteers the answer to a question before it is asked, organizations that allow "verification" authorize the interviewer to refashion the question to confirm the answer with the respondent (e.g., "The next question asks how old you are, and I think you said you are 50. Is that correct?").

The response categories an interviewer reads or probes with can affect the distribution of responses (e.g., Smit, Dijkstra, & van der Zouwen, 1997). Some organizations balance the rigidities of standardization against conversational practices by using probing practices like "tuning." Tuning probes only with categories close to an answer already volunteered by the respondent, instead of all categories; so if the respondent says "not too often" the interviewer might probe with the relevant categories: "So would you say seldom, rarely or never?" (van der Zouwen, 2002, p. 59; see also the description of "zeroing in" in Fowler & Mangione, 1990, p. 41).

Some organizations also train interviewers about concepts in the survey as advocated in the "best practices" for training interviewers of the American Association for Public Opinion Research (AAPOR, http://www.aapor.org/bestpractices). One way to do this is through detailed "question-by-question specifications." Although organizations may expect interviewers to deploy this knowledge when needed, they may not specifically train interviewers in how to determine when the respondent needs the information or how to convey it in a way compatible with principles of standardization. Finally, unscripted behaviors, such as laughter, may also vary across survey organizations (Lavin & Maynard, 2001).

In addition to variation in implementing the principles of standardization, there are failures of standardization, which can also be associated with measurement error (see discussion later). The inability of interviewers to read a question as worded may indicate problems in question design or the need for more training (van der Zouwen & Dijkstra, 2002; van der Zouwen & Smit, 2004; Ongena & Dijkstra, 2006). Some failures of standardization reflect the tension between constraints of standardization and the pressure of conversational practices (Maynard & Schaeffer, 2002c). For example, interviewers might read the first question in a series as worded, but then begin to modify or skip questions as they try to take into account information provided by the respondent (Oksenberg, Cannell, & Blixt, 1996, p. 19). Conversational practices appear in the standardized interview in other ways. Interviewers may interpret hesitations or pauses as indicators of comprehension problems (Schaeffer & Maynard, 1995, 2002). When they are unsure how to answer a question, respondents may provide "reports" about their circumstances rather than give a properly formatted answer (e.g., "yes" or "no" to a "yes/no" question) (Schaeffer & Maynard, 1995, 2002, 2008), and interviewers routinely code some answers (such as "yes" when the respondent says "probably") using their tacit knowledge of what such answers "mean" (Hak, 2002).

When standardization is criticized because it undermines rapport, it is helpful to remember that "rapport" fell into disrepute (Goudy & Potter, 1976) in part due to concerns about rapport's influence on bias (e.g., Weiss, 1968; Dohrenwend, Colombotos, & Dohrenwend, 1968). The various components of the vague construct, "rapport" — including motivation, friendliness, affiliation, ingratiation, and empathy — still need to be systematically analyzed, identified as properties of the interviewer or the respondent, and investigated (but see Lavin & Maynard, 2001). Several studies illustrate the usefulness of these distinctions. Respondents in a panel survey interviewed up to four times by the same interviewer — who might have had greater "rapport" with the interviewer as a result — gave lower reports of drug use than those interviewed by different interviewers (Mensch & Kandel, 1988), suggesting that "rapport" might increase bias. In contrast, motivated respondents reported more events presumably because they worked harder (Cannell, Miller, & Oksenberg, 1981). Similarly, the personal style of interviewing implemented by Dijkstra and van der Zouwen (1987) attempted to increase motivation by allowing the interviewer to reflect things the respondent had expressed, particularly feelings or signs of distress (e.g., "I understand what your moving to your house meant to you"). Respondents interviewed with the personal style drew more accurate maps and gave fewer socially desirable answers (Dijkstra, 1987) than respondents interviewed with a more formal scripted style, in a study that randomly assigned interviewers and respondents to treatments.

Another criticism of standardization is that it decreases validity by denying interviewers interactional resources for diagnosing and correcting comprehension problems. A "conversational" style of interviewing (Schober & Conrad, 1997), like other forms of standardized interviewing, required interviewers to read questions as worded, but also authorized interviewers to say "whatever they wanted to assure that the respondent had understood the question as the survey designer had intended" (1997, p. 584). An experiment using questions from federal surveys about housing, employment, and purchases compared this flexible style with a strict standardization, in which interviewers were prohibited from providing definitions. Respondents interviewed with the flexible style also read approximately 200 words of instructions explaining how survey concepts could differ from everyday concepts and stressing the importance of asking questions (1997, p. 600); thus they were potentially more engaged and motivated than most survey respondents. When answering questions about a complicated scenario, respondents interviewed with the flexible style were substantially more accurate than those interviewed using strict standardization. The flexible interviews were also substantially longer, however. Recent implementations of a complex instrument, the event history calendar (EHC) (Belli, James, VanHoewyk, & Alcser, 2009), use a different flexible or "conversational" style of interviewing: question wording is much less scripted and respondents are encouraged to report across topics so that memories in one life domain can improve the accuracy of reports in another (Belli, Shay, & Stafford, 2001; Belli, Smith, Andreski, & Agrawal, 2007). Overall, in comparison to conventional standardized questions, this particular implementation of the EHC resulted in somewhat higher interviewer variance for a number of variables (Sayles, Belli, & Serrano, 2008), but the impact on total error has not been assessed.

13.5.4.2. Interaction and the quality of survey data Studies of interviewer effects reviewed earlier (e.g., Collins, 1980; Mangione et al., 1992) documented that conditions in which standardization or the paradigmatic question-answer sequence fail — whether because of problems in the design of the question, a poor fit between the respondent and the question, or a need for probing — are associated with increased interviewer variation, that is, decreased reliability. The impact of interaction in the survey interview — whether standardized or not — on the reliability of data has been investigated in studies using reinterviews. Its consequences for validity have been examined in studies that compare answers to external records or compare a retrospective report in one wave of a panel study to an earlier contemporaneous report.

When survey answers about health services use were compared to records, interviewers' deviations from the scripted wording were associated with more accurate answers, although the difference was significant for only one variable (Dykema, Lepkowski, & Blixt, 1997). But when interviewers asked about joint legal custody, deviations from the script were associated with much lower accuracy (Schaeffer & Dykema, 2004). Dykema (2004, 2005) examined twelve questions about child support or family matters that could be checked against records, and found no relationship between the accuracy of answers and interviewers' reading of the question for seven questions. For three questions (about joint legal custody, whether visitation privileges had been legalized, and the date of the first court order) deviations from the script were associated with lower accuracy while for two (the amount of child support owed and the amount paid), deviations were associated with greater accuracy. These findings suggest that different types of questions (e.g., those about statuses or dates as opposed to those requiring estimation of totals) might evoke different behaviors from interviewers and that different behaviors lead to greater accuracy for different types of questions. Moreover, the effects of deviating from scripts on accuracy probably depend on exactly what changes interviewers make to the wording of scripted questions. Hess, Singer, and Bushery (1999) found that interviewers' accuracy in reading questions did not affect test-retest reliability, but deviation from the script was substantial for only two of their 34 questions. Other behaviors, however, are consistently associated with less accurate answers including probing by interviewers and respondents' expressions of uncertainty or delay (Mathiowetz, 1999; Dykema et al., 1997; Hess, Singer, & Bushery, 1999; Draisma & Dijkstra, 2004; Schaeffer & Dykema, 2004).

13.6. Managing Survey Interviewers: Training and Supervision

Few have studied the many issues that arise in managing survey interviewers. Studies we have found examine some aspects of training and supervision. The "best practices" for surveys described by AAPOR (see website specified earlier) include three items relevant for interviewers: "Train interviewers carefully on interviewing techniques and the subject matter of the survey. Check quality at each stage. Maximize cooperation or response rates within the limits of ethical treatment of human subjects."

Most survey organizations have a general interviewer training, which teaches how to use equipment (e.g., Wojcik & Hunt, 1998) and practices of standardized interviewing (Viterna & Maynard, 2002) and also provides a briefing for each project. The on-line training for the Behavioral Risk Factors Surveillance Surveys illustrates one implementation of interviewer training (http://www.cdc.gov/BRfss/training/interviewer/index.htm). General training involves practice such as role playing with other trainees and conducting "mock" interviews that are monitored and reviewed. Such training might be supplemented periodically by special training in avoiding refusals (see discussion earlier) and coaching as part of monitoring and quality control.

Fowler and Mangione (1990) studied effects of the length of interviewer training and reported that standardized interviewing technique was worst for interviewers trained for less than a day. Five or 10 days of training added little improvement over two days, however, except in probing (see also Billiet & Loosveldt, 1988). In practice, the length of training depends on the complexity of the study and the tasks the interviewer must perform as well as the budget. Findings by Groves and McGonagle (2001) and Campanelli et al. (1997) strongly indicate the need for both study-specific training, so that interviewers can describe the study's purpose and content effectively and quickly, and training on nonresponse techniques. One model of training is immersive and includes role-playing, interaction with and advice from expert interviewers, and audio recordings.

Monitoring interviewer performance involves both verifying that interviews have taken place to prevent falsification (particularly for face-to-face surveys), and reviewing how successfully standardized practices are implemented (American Statistical Association, 2003; Cannell, Lawson, & Hausser, 1975). Criteria and methodology for evaluating interviewers continue to be refined, usually for telephone surveys (Durand, 2004; Kelly, Link, Petty, Hobson, & Cagney, 2008; Pickery & Loosveldt, 2004; Steve, Burks, Lavrakas, Brown, & Hoover, 2008; Tarnai & Moore, 2008). Because laptops can now record field interviews, it is now possible to use similar standards and techniques to evaluate field and telephone interviewers (Biemer, Herget, Morton, & Willis, 2000).

Interviewers are increasingly recognized as research agents who must be trained that following study protocols is part of protecting human subjects. Simultaneously, interviewers are themselves increasingly under observation as part of the measurement process. Modeling interviewers' impact on data requires interviewer identifiers (to index interviews conducted by the same interviewer), interviewer characteristics, and, for face-to-face surveys, data from the sample frame or other sources (such as the Census) on characteristics of interviewers' caseloads. Interviewers now are regularly asked to sign both confidentiality pledges and forms assenting to use of data on their personal characteristics and recordings of their voices. Descriptions of burdens of the interviewer's role seldom take the interviewer's point of view (but see Japec, 2008), but the challenges of interviewing become more complicated as study designs do.

13.7. Conclusion

In the future, some researchers may attempt to reduce or eliminate interviewer effects

avatars (Conrad & Schober, 2008), that would presumably embody some version of desirable interviewer behavior.

But interviewers will continue to be an important part of survey designs when samples are complex, response rates are critical, and measurement is demanding. Population surveys collect extremely varied types of data, perhaps increasingly so. Interviewers can be asked to take environmental, biometric, or anthropometric measurements; and expert practitioners of other forms of data collection, such as phlebotomists or clinic staff, are sometimes asked to conduct interviews for population surveys. Researchers will continue to try to describe and understand how interviewers accomplish tasks such as listing samples and identifying sample units, recruiting and persuading respondents, filling out complex forms such as household rosters or EHCs, and identifying and solving respondents' comprehension problems. The tension between validity and reliability or flexibility and consistency, and the alternation between conversational practices and standardization (see Maynard & Schaeffer, 2002c) may be resolved in changing ways as the consequences of current practices for interviewers, respondents, and data quality are better understood. We have described many relationships between interviewer characteristics and outcomes, but those relationships are mediated by the interaction between the parties to the interview in ways that we are still only beginning to understand.

This review suggests the need for research concerning the interviewer's role in both recruitment and measurement. The skills associated with getting sample members to participate in a survey and the lessons of an interviewer's experience in this task — including flexibility and responsiveness — are in tension with the skills required by standardization during the interview — adhering to the text of a questionnaire and a set of prescribed rules. Although Presser and Zhao (1992) found no relationship between an interviewer's success in recruiting and performance on measurement tasks in a telephone survey, their findings are in need of replication and extension, particularly for face-to-face interviews. Tension between securing participation and data quality may also arise: Sample members who make negative comments about participating but who are ultimately converted may provide lower quality data (Campanelli et al., 1997; Couper, 1997). Future research about the interviewer's role in measurement should consider how to refine interviewing practices used with less standardized instruments (such as EHCs and grids), improve methods for diagnosing and correcting comprehension problems, and motivate respondents. Studies about methods for training interviewers — both in recruiting respondents and in more complex interviewing techniques — are difficult to design and execute, but increasingly needed.

Although such knowledge is always provisional, our review suggests a few guidelines and cautions. Training and monitoring interviewers in all the tasks they are asked to perform — making observations, sampling, recruiting, screening, and administering instruments — are essential. Recruiting and persuading respondents is likely to be more successful if interviewers use flexible introductions (which they have been well trained to deliver so that they describe the study accurately and cover other elements of informed consent) rather than read a script, and are trained to maintain interaction and answer respondents' concerns quickly and responsively. Measurement is likely to be more reliable if instruments are designed to clarify survey concepts and permit standardized administration, and may be more valid if

Future research on interviewing needs strong study designs and analytical approaches, as illustrated, for example, by O'Muircheartaigh & Campanelli (1998). The strength of the evidence about the effects of interviewers in the studies we review here varies substantially across studies. Disentangling potentially confounding effects (e.g., distinguishing effects of interviewers and the areas to which they are assigned for face-to-face surveys) requires study designs that use interpenetrated samples, supplementary data about the sample frame, data analysis techniques like multilevel modeling, and manipulation checks for experiments. We have seen that for some characteristics (or some respondents) evidence calls into question the assumption that matching interviewers-respondents on relevant characteristics increases the validity of survey responses, and racial matching in particular may be challenging to implement (e.g., Cobb, Boland-Perez, & LeBaron, 2008). An alternative approach would be to cross interviewer and respondent characteristics when feasible (so that some interviewer-respondent pairs match and some do not) and estimate effects of interviewer characteristics, compensating for the lack of random assignment by controlling for other characteristics of interviewers and their workloads (but compare the discussion of matching interviewer and respondent race in Krysan, 2002).

Our review also suggests that researchers document characteristics of interviewers. The comprehensive list of items suggested for disclosure in AAPOR's "best practices" (which goes beyond the guidelines for minimum disclosure) proposes describing "interviewer characteristics" but is not more specific. At the least, information summarizing the distribution of information about the race, gender, age, and experience of the corps of interviewers and the size of their caseloads would be desirable. Identifiers for interviewers, characteristics of interviewers, and characteristics of sample units that could be associated with interviewer characteristics (when relevant) should be included in data sets more routinely than they are now, so that analysts can study how this key component of the measurement process — the interviewer — affects the reliability and validity of survey data. Finally, we look forward to assessments of the impact of interviewers that examine total error — bias and variable errors as well as errors of nonobservation and observation.

Acknowledgment

We would like to thank Colm O'Muircheartaigh, Stanley Presser, Roger Tourangeau, and the editors for their extremely helpful comments.

References

American Statistical Association. (2003). Interviewer falsification in survey research: Current best methods for prevention, detection, and repair of its effects. Available at http://www.amstat.org/sections/SRMS/falsification.pdf

Anderson, B. A., Silver, B. D., & Abramson, P. R. (1988a). The effects of race of the interviewer on measures of electoral participation by blacks in SRC national election studies. *Public Opinion Quarterly*, *52*(1), 53–83.

Anderson, B. A., Silver, B. D., & Abramson, P. R. (1988b). The effects of the race of the interviewer on race related attitudes of black respondents in SRC/CPS national election studies. *Public Opinion Quarterly*, *52*(3), 289–324.

Bailar, B. A., Bailey, L., & Stevens, J. (1977). Measures of interviewer bias and variance. *Journal of Marketing Research*, *24*, 337–343.

Bailey, L., Moore, T. F., & Bailar, B. A. (1978). An interviewer variance study for the eight impact cities of the national crime survey cities sample. *Journal of the American Statistical Association*, *73*, 16–23.

Barath, A., & Cannell, C. F. (1976). Effect of interviewer's voice intonation. *Public Opinion Quarterly*, *40*(3), 370–373.

Belli, R. F., James, S. A., VanHoewyk, J., & Alcser, K. (2009). The implementation of a computerized event history calendar questionnaire for research in life course epidemiology. In: R. F. Belli, F. P. Stafford & D. F. Alwin (Eds), *Calendar and time diary methods in life course research* (pp. 225–238). Thousands Oaks, CA: Sage.

Belli, R. F., Shay, W. L., & Stafford, F. P. (2001). Event history calendars and question list surveys: A direct comparison of interviewing methods. *Public Opinion Quarterly*, *65*(1), 45–74.

Belli, R. F., Smith, L., Andreski, P., & Agrawal, S. (2007). Methodological comparisons between CATI event history calendar and conventional questionnaire instruments. *Public Opinion Quarterly*, *71*(1), 603–622.

Biemer, P., Herget, D., Morton, J., & Willis, G. (2000). The feasibility of monitoring field interviewer performance using Computer Audio Recorded Interviewing (CARI). *The American Statistical Association 2000 proceedings of the section on survey research methods* (pp. 1068–1073).

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.

Blair, E. (1977). More on the effects of interviewer's voice intonation. *Public Opinion Quarterly*, *41*(4), 544–548.

Brunton-Smith, I., & Sturgis, P. (2009). A new approach to separating interviewer from area variability in face-to-face interviewer surveys. Paper presented at the annual meetings of the American Association for Public Opinion Research, 14–17 May, Ft. Lauderdale, FL.

Callegaro, M., DeKeulenaer, F., Krosnick, J. A., & Daves, R. P. (2005). Interviewer effects in a RDD telephone pre-election poll in Minneapolis 2001: An analysis of the effects of interviewer race and gender. *The American Statistical Association proceedings of the Survey Research Methods Section (SRMS)* (pp. 3815–3820).

Campanelli, P., & O'Muircheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey nonresponse. *Quality & Quantity*, *33*(1), 59–76.

Campanelli, P., Sturgis, P., & Purdon, S. (1997). Can you hear me knocking: An investigation into the impact of interviewers on survey response rates. *The Survey Methods Centre at SCPR*.

Campbell, B. A. (1981). Race-of-interviewer effects among southern adolescents. *Public Opinion Quarterly*, *45*(2), 231–244.

Cannell, C. F., Lawson, S. A., & Hausser, D. L. (1975). *A technique for evaluating interviewer performance*. Ann Arbor, MI: Survey Research Center of the Institute for Social Research, University of Michigan.

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. In: S. Leinhardt (Ed.), *Sociological methodology* (Vol. 11, pp. 389–437). San Francisco, CA: Jossey-Bass.

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

Cobb, P. D., Boland-Perez, C., & LeBaron, P. (2008). Race-matching: Interviewers' reactions to the race-matching process. Survey Practice (December). Available at http://surveypractice.org/2008/12/19/race-matching/, 13 January 2009.
Collins, M. (1980). Interviewer variability: A review of the problem. *Journal of the Market Research Society, 22*(2), 77–95.
Conrad, F. G., & Schober, M. F. (Eds). (2008). *Envisioning the survey interview of the future*. Hoboken, NJ: Wiley.
Converse, J. M. (1987). *Survey research in the United States: Roots and emergence 1890–1960*. Berkeley, CA: University of California Press.
Cotter, P. R., Cohen, J., & Coulter, P. B. (1982). Race-of-interviewer effects in telephone interviews. *Public Opinion Quarterly, 46*(2), 278–284.
Couper, M. P. (1997). Survey introductions and data quality. *Public Opinion Quarterly, 61*, 317–338.
Couper, M. P. (2005). Technology trends in survey data collection. *Social Science Computing Review, 23*(4), 486–501.
Couper, M. P., & Groves, R. M. (1992). The role of the interviewer in survey participation. *Survey Methodology, 18*(2), 263–277.
Couper, M. P., & Groves, R. M. (2002). Introductory interactions in telephone surveys and nonresponse. In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 161–178). New York, NY: Wiley.
Davis, D. W. (1997). The direction of race of interviewer effects among African-Americans: Donning the black mask. *American Journal of Political Science, 41*(1), 309–322.
Davis, D. W., & Silver, B. D. (2003). Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Political Science, 47*(1), 33–45.
Davis, P., & Scott, A. (1995). The effect of interviewer variance on domain comparisons. *Survey Methodology, 21*, 99–106.
de Leeuw, E., & Collins, M. (1997). Data collection methods and survey quality: An overview. In: L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds), *Survey measurement and process quality* (pp. 199–220). New York, NY: Wiley-Interscience.
Dijkstra, W. (1983). How interviewer variance can bias the results of research on interviewer effects. *Quality and Quantity, 17*(3), 179–187.
Dijkstra, W. (1987). Interviewing style and respondent behavior: An experimental study of the survey interview. *Sociological Methods and Research, 16*(2), 309–334.
Dijkstra, W., & Smit, J. H. (2002). Persuading reluctant recipients in telephone surveys. In: R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds), *Survey nonresponse* (pp. 121–134). New York, NY: Wiley.
Dijkstra, W., & van der Zouwen, J. (1987). Styles of interviewing and the social context of the survey-interview. In: H. J. Hippler, N. Schwarz & S. Sudman (Eds), *Social information processing and survey methodology* (pp. 200–211). New York, NY: Springer-Verlag.
Dillman, D. A., Gallegos, J. G., & Frey, J. H. (1976). Reducing refusal rates for telephone interviews. *Public Opinion Quarterly, 40*(1), 66–78.
Dohrenwend, B. S., Colombotos, J., & Dohrenwend, B. P. (1968). Social distance and interviewer effects. *Public Opinion Quarterly, 32*, 410–422.
Draisma, S., & Dijkstra, W. (2004). Response latency and (para)linguistic expression as indicators of response error. In: S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds), *Methods for testing and evaluating survey questionnaires* (pp. 131–148). New York, NY: Springer-Verlag.
Durand, C. (2004). How to measure interviewer performance in telephone surveys. *Quality and Quantity, 39*, 763–778.

Durbin, J., & Stuart, A. (1951). Differences in response rates of experienced and inexperienced interviewers. *Journal of the Royal Statistical Society, Series A (General), 114*(2), 163–206.
Dykema, J. (2004). *Analysis of factors influencing errors in self-reports about child support and other family-related variables*. Doctoral Dissertation, Department of Sociology, University of Wisconsin-Madison.
Dykema, J. (2005). An investigation of the impact of departures from standardized interviewing on response errors in self-reports about child support and other family-related variables. Paper presented at the annual meeting of the American Association for Public Opinion Research, May 12–15, Miami Beach, FL.
Dykema, J., Lepkowski, J. M., & Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: analysis of interaction coding in a validation study. In: L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds), *Survey measurement and process quality* (pp. 287–310). New York, NY: Wiley-Interscience.
Fecso, R. (1991). A review of errors of direct observation in crop yield surveys. In: P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz & S. Sudman (Eds), *Measurement errors in surveys* (pp. 327–346). New York, NY: Wiley.
Finkel, S. E., Guterbock, T. M., & Borg, M. J. (1991). Race-of-interviewer effects in a preelection poll: Virginia 1989. *Public Opinion Quarterly, 55*(3), 313–330.
Flores-Macias, F., & Lawson, C. (2008). Effects of interviewer gender on survey responses: Findings from a household survey in Mexico. *International Journal of Public Opinion Research, 20*, 100–110.
Fowler, F. J., Jr., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error*. Newbury Park, CA: Sage.
Fuchs, M. (2002). The impact of technology on interaction in computer-assisted interviews. In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 471–491). New York, NY: Wiley.
Goudy, W. J., & Potter, H. R. (1975–1976). Interview rapport: Demise of a concept. *Public Opinion Quarterly, 39*(4), 529–543.
Gray, P. G. (1956). Examples of interviewer variability taken from two sample surveys. *Applied Statistics, 5*(2), 73–85.
Groves, R. M. (1989). *Survey errors and survey costs*. New York, NY: John Wiley & Sons.
Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56*(4), 475–495.
Groves, R. M., & Couper, M. P. (1994). *Householders and interviewers: The anatomy of pre-interview interactions*. Unpublished manuscript.
Groves, R. M., & Couper, M. P. (1996). Contact-level influences on cooperation in face-to-face surveys. *Journal of Official Statistics, 12*(1), 63–83.
Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York, NY: Wiley.
Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: Wiley.
Groves, R. M., & Fultz, N. H. (1985). Gender effects among telephone interviewers in a survey of economic attitudes. *Sociological Methods and Research, 14*(1), 31–52.
Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly, 50*(2), 251–266.
Groves, R. M., & McGonagle, K. A. (2001). A theory-guided interviewer training protocol regarding survey participation. *Journal of Official Statistics, 17*(2), 249–266.

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

Hak, T. (2002). How interviewers make coding decisions. In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 449–470). New York, NY: Wiley.

Hansen, M. H., Hurwitz, W. N., & Bershad, M. A. (1961). Measurement errors in censuses and surveys. *Bulletin of the International Statistical Institute*, *38*(2), 359–374.

Hatchett, S., & Schuman, H. (1976). White respondents and race-of-interviewer effects. *Public Opinion Quarterly*, *39*(4), 523–528.

Hess, J., Singer, E., & Bushery, J. (1999). Predicting test-retest reliability from behavior coding. *International Journal of Public Opinion Research*, *11*(4), 346–360.

Hill, M. E. (2002). Race of the interviewer and perception of skin color: Evidence from the multi-city study of urban inequality. *American Sociological Review*, *67*(1), 99–108.

Houtkoop-Steenstra, H., & van den Berg, H. (2000). Effects of introductions in large-scale telephone survey interviews. *Sociological Methods and Research*, *28*(3), 251–280.

Hox, J., & de Leeuw, E. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In: R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds), *Survey nonresponse* (pp. 103–120). New York, NY: Wiley.

Hox, J. J., de Leeuw, E. D., & Kreft, I. G. G. (1991). The effect of interviewer and respondent characteristics on the quality of survey data: A multilevel model. In: P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz & S. Sudman (Eds), *Measurement errors in surveys* (pp. 439–462). New York, NY: Wiley.

Huang, M. (2009). Race of the interviewer and the black-white test score gap. *Social Science Research*, *38*(1), 29–38.

Huddy, L., Billig, J., Bracciodieta, J., Hoeffler, L., Moynihan, P. J., & Puglian, P. (1997). The effect of interviewer gender on the survey response. *Political Behavior*, *19*(3), 197–220.

Hughes, A., Chromy, J., Giacoletti, K., & Odom, D. (2002). Impact of interviewer experience on respondent reports of substance use. In: J. Gfroerer, J. Eyerman & J. Chromy (Eds), *Redesigning an ongoing national household survey: Methodological issues, substance abuse and mental health services administration* (pp. 161–184). Rockville, MD: Office of Applied Studies.

Hurtado, A. (1994). Does similarity breed respect: Interviewer evaluations of Mexican-descent respondents in a bilingual survey. *Public Opinion Quarterly*, *58*(1), 77–95.

Hyman, H. H. (1975[1954]). *Interviewing in social research*. Chicago, IL: The University of Chicago.

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

Johnson, T. P., & Moore, R. W. (1993). Gender interactions between interviewer and survey respondents: Issues of pornography and community standards. *Sex Roles: A Journal of Research*, *28*(5/6), 243–261.

Kalton, G., Lepkowski, J., Montanari, G. E., & Maligalig, D. (1990). Characteristics of second wave nonrespondents in a panel survey. In: *The American Statistical Association proceedings of the section on survey research methods* (pp. 462–467). Alexandria, VA: American Statistical Association.

Kane, E. W., & Macaulay, L. J. (1993). Interviewer gender and gender attitudes. *Public Opinion Quarterly*, *57*(1), 1–28.

Kelly, J., Link, M. W., Petty, J., Hobson, K., & Cagney, P. (2008). Establishing a new survey research call center. In: J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link & R. L. Sangster (Eds), *Advances in telephone survey methodology* (pp. 317–339). New Jersey: Wiley.

Krysan, M. (2002). Data update to racial attitudes in America, an update and website to complement racial attitudes in America: Trends and interpretations, revised edition, by Howard Schuman, Charlotte Steeh, Lawrence Bobo and Maria Krysan, 1997, Harvard University Press. Available at http://tigger.uic.edu/krysan/writeup.htm. Accessed 13 January 2009.

Krysan, M., & Couper, M. P. (2003). Race in the live and the virtual interview: Racial deference, social desirability, and activation effects in attitude surveys. *Social Psychology Quarterly*, *66*(4), 364–383.

Lavin, D., & Maynard, D. W. (2001). Standardization vs. rapport: Respondent laughter and interviewer reaction during telephone surveys. *American Sociological Review*, *66*(3), 453–479.

Leal, D., & Hess, F., (1999). Survey bias on the front porch: Are all subjects interviewed equally? *American Politics Quarterly*, *27*, 468–487.

Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In: R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds), *Survey nonresponse* (pp. 259–272). New York, NY: Wiley.

Lessler, J. T., & Kalsbeek, W. D. (1992). *Nonsampling error in surveys*. New York, NY: Wiley.

Link, M. W., Armsby, P. P., Hubal, R. C., & Guinn, C. I. (2006). Accessibility and acceptance of responsive virtual human technology as a survey interviewer training tool. *Computers in Human Behavior*, *22*(3), 412–426.

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

Manheimer, D., & Hyman, H. (1949). Interviewer performance in area sampling. *The Public Opinion Quarterly*, *13*(1), 83–92.

Marsden, P. V. (2003). Interviewer effects in measuring network size using a single name generator. *Social Networks*, *25*(1), 1–16.

Mathiowetz, N. A. (1999). Expressions of respondent uncertainty as indicators of data quality. *International Journal of Public Opinion Research*, *11*(3), 289–296.

Maynard, D. W., Freese, J., & Schaeffer, N. C. (2008). Requesting as a social action: Implications for nonresponse and 'leverage-saliency' in the survey Interview. Paper Presented at the Annual Meetings of the American Sociological Association, 2 August, Boston, MA.

Maynard, D. W., & Schaeffer, N. C. (1997). Keeping the gate: Declinations of the request to participate in a telephone survey interview. *Sociological Methods and Research*, *26*(1), 34–79.

Maynard, D. W., & Schaeffer, N. C. (2002a). Opening and closing the gate: The work of optimism in recruiting survey respondents. In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 179–204). New York, NY: Wiley.

Maynard, D. W., & Schaeffer, N. C. (2002b). Refusal conversion and tailoring. In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 219–239). New York, NY: Wiley.

Maynard, D. W., & Schaeffer, N. C. (2002c). Standardization and its discontents. In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 3–45). New York, NY: Wiley.

Mensch, B. S., & Kandel, D. B. (1988). Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effects. *Public Opinion Quarterly*, *52*(1), 100–124.

Merkle, D. M., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In: R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds), *Survey nonresponse* (pp. 243–258). New York, NY: Wiley.

Morton-Williams, J. (1993). *Interviewer approaches*. England: Dartmouth.

Nederhof, A. J. (1987). When neutrality is negative. *Quality and Quantity*, *21*(4), 425–432.

O'Brien, E., Mayer, T. S., Groves, R. M., & O'Neill, G. E. (2002). Interviewer training to increase survey participation. *The American Statistical Association Proceedings of the Survey Research Methods Section (SRMS)*, American Statistical Association, Alexandria, VA (pp. 2502–2507).

Oksenberg, L., & Cannell, C. F. (1988). Effects of interviewer vocal characteristics on nonresponse. In: R. M. Groves, P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II & J. Waksberg (Eds), *Telephone survey methodology* (pp. 257–272). New York, NY: Wiley.

Oksenberg, L., Cannell, C., & Blixt, S. (1996[1991]). Analysis of interviewer and respondent behavior in the household survey. National Medical Expenditure Survey Methods 7, Agency for Health Care Policy and Research (AHCPR Pub. No. 96-N016). Public Health Service, U.S. Department of Health and Human Services, Rockville, MD.

Oksenberg, L., Coleman, L., & Cannell, C. F. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly*, *50*(1), 97–111.

Olson, K., & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly*, *71*(2), 273–286.

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

O'Muircheartaigh, C., & Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, *163*(3), 437–446.

O'Muircheartaigh, C., Eckman, S., & Weiss, C. (2002). Traditional and enhanced filed listing for probability sampling. In: *The American Statistical Association 2002 proceedings of the section on survey research methods* (pp. 2563–2567). Washington, DC: American Statistical Association.

O'Neil, M. J., Groves, R. M., & Cannell, C. F. (1979). Telephone interview introductions and refusal rates: Experiments in increasing respondent cooperation. In: *The American Statistical Association 1979 proceedings of the section on survey research methods* (pp. 252–255). Washington, DC: The American Statistical Association.

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

Pickery, J., & Loosveldt, G. (2001). An exploration of question characteristics that mediate interviewer effects on item nonresponse. *Journal of Official Statistics*, *17*(3), 337–350.

Pickery, J., & Loosveldt, G. (2002). A multilevel multinomial analysis of interviewer effects on various components of unit nonresponse. *Quantity and Quality*, *36*(4), 427–437.

Pickery, J., & Loosveldt, G. (2004). A simultaneous analysis of interviewer effects on various data quality indicators with identification of exceptional interviewers. *Journal of Official Statistics*, *20*(1), 77–90.

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

Presser, S., & Zhao, S. (1992). Attributes of questions and interviewers as correlates of interviewing performance. *Public Opinion Quarterly*, *56*(2), 236–240.

Sayles, H., Belli, R. F., & Serrano, E. (2008). *Interviewer variance between event history calendar and conventional questionnaire interviews*. Unpublished manuscript, University of Nebraska, Lincoln, Nebraska.

Schaeffer, N. C. (1980). Evaluating race of interviewer effects in a national survey. *Sociological Methods and Research*, *8*(4), 400–419.

Schaeffer, N. C. (1991). Conversation with a purpose—or conversation? Interaction in the standardized interview. In: P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz & S. Sudman (Eds), *Measurement errors in surveys* (pp. 367–392). New York, NY: Wiley.

Schaeffer, N. C., & Dykema, J. (2004). A multiple-method approach to improving the clarity of closely related concepts: Distinguishing legal and physical custody of children. In: S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds), *Methods for testing and evaluating survey questionnaires* (pp. 475–502). New York, NY: Springer-Verlag.

Schaeffer, N. C., & Maynard, D. W. (1995). Occasioning intervention: Interactional resources for comprehension in standardized survey interviews. Paper presented at the International Conference on Survey Measurement and Process Quality International Conference on Survey Measurement and Process Quality, April, Bristol, England.

Schaeffer, N. C., & Maynard, D. W. (1996). From paradigm to prototype and back again: Interactive aspects of cognitive processing in survey interviews. In: N. Schwarz & S. Sudman (Eds), *Answering questions: methodology for determining cognitive and communicative processes in survey research* (pp. 65–88). San Francisco, CA: Jossey-Bass.

Schaeffer, N. C., & Maynard, D. W. (2002). Occasions for intervention: Interactional resources for comprehension in standardized survey interviews. In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 261–280). New York, NY: Wiley.

Schaeffer, N. C., & Maynard, D. W. (2008). The contemporary standardized survey interview for social research. In: F. G. Conrad & M. F. Schober (Eds), *Envisioning the survey interview of the future* (pp. 31–57). Hoboken, NJ: Wiley.

Schnell, R., & Kreuter, F. (2000). Das DEFECT-projekt: Sampling-errors und nonsampling-errors in komplexen bevölkerungsstichproben. *ZUMA-Nachrichten*, *47*(S), 89–101.

Schnell, R., & Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics*, *21*(3), 389–410.

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, *61*(4), 576–602.

Schuman, H., & Converse, J. M. (1971). The effects of black and white interviewers on black responses in 1968. *Public Opinion Quarterly*, *35*(1), 44–68.

Singer, E., & Frankel, M. R. (1982). Informed consent procedures in telephone interviews. *American Sociological Review*, *47*, 416–427.

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

Singer, E., & Kohnke-Aguirre, L. (1979). Interviewer expectation effects: A replication and extension. *Public Opinion Quarterly*, *43*(2), 245–260.

Smit, J. H., Dijkstra, W., & van der Zouwen, J. (1997). Suggestive interviewer behaviour in surveys: An experimental study. *Journal of Official Statistics*, *13*(1), 19–28.

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

Steve, K. W., Burks, A. T., Lavrakas, P. J., Brown, K. D., & Hoover, J. B. (2008). Monitoring telephone interviewer performance. In: J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link & R. L. Sangster (Eds), *Advances in telephone survey methodology* (pp. 401–422). New Jersey: Wiley.

Suchman, L., & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association*, *85*(409), 232–253.

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

Sudman, S., Bradburn, N. M., Blair, E., & Stocking, C. (1977). Modest expectations: The effects of interviewers' prior expectations on responses. *Sociological Methods and Research*, *6*(2), 171–182.

Tarnai, J., & Moore, D. L. (2008). Measuring and improving telephone interviewer performance and productivity. In: J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japec, P. J. Lavrakas, M. W. Link & R. L. Sangster (Eds), *Advances in telephone survey methodology* (pp. 359–384). New Jersey: Wiley.

Tourangeau, R., Rasinski, K., Jobe, J., Smith, T., & Pratt, W. (1997). Sources of error in a survey of sexual behavior. *Journal of Official Statistics*, *13*, 341–365.

van der Vaart, W., Ongena, Y., Hoogendoom, A., & Dijkstra, W. (2006). Do interviewers' voice characteristics influence cooperation rates in telephone surveys? *International Journal of Public Opinion Research*, *18*(4), 488–499.

van der Zouwen, J. (2002). Why study interaction in survey interviews? In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 47–66). New York, NY: Wiley.

van der Zouwen, J., & Dijkstra, W. (2002). Testing questionnaires using interaction coding. In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 427–448). New York, NY: Wiley.

van der Zouwen, J., & Smit, J. H. (2004). Evaluating survey questions by analyzing patterns of behavior codes and question-answer sequences: A diagnostic approach. In: S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds), *Methods for testing and evaluating survey questionnaires* (pp. 109–130). New York, NY: Springer-Verlag.

van Tilburg, T. (1998). Interviewer effects in the measurement of personal network size: A nonexperimental study. *Sociological Methods and Research*, *26*, 300–328.

Viterna, J. S., & Maynard, D. W. (2002). How uniform is standardization? Variation within and across survey research centers regarding protocols for interviewing. In: D. W. Maynard, H. Houtkoop-Steenstra, J. van der Zouwen & N. C. Schaeffer (Eds), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 365–401). New York, NY: Wiley.

Weiss, C. H. (1968). Validity of welfare mothers' interview responses. *Public Opinion Quarterly*, *32*, 622–633.

Williams, D. (1942). Basic instructions for interviewers. *Public Opinion Quarterly*, *6*(4), 634–641.

Wojcik, M. S., & Hunt, E. (1998). Training field interviewers to use computers: Past, present, and future trends. In: M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, W. L. Nicholls II & J. M. O'Reilly (Eds), *Computer assisted survey information collection* (pp. 331–350). New York, NY: Wiley.

Wolford, M. L., Brown, R. E., Marsden, A., Jackson, J. S., & Harrison, C. (1995). Bias in telephone surveys of African Americans: The impact of perceived race of interviewer on responses. *The American Statistical Association 1995 proceedings of the Survey Research Methods Section (SRMS)* (pp. 795–804).

Chapter 14

Telephone Surveys

Paul J. Lavrakas

Telephone survey methods have undergone serious development only in the past 30 years. Prior to then, the percentage of households with telephones in the United States, Europe, and elsewhere was too low to justify use of the telephone as a representative survey mode for general population surveys. However, by the late 1970s, U.S. household telephone coverage exceeded 90%. By the mid-1980s, telephone surveying became commonplace for academic, commercial, and government surveys of the general public.

In the mid-2000s, however, the ability of researchers to reach representative samples of the U.S. public via telephone began to face serious new challenges. The main reasons for this were (a) changes in how Americans use telephones, (b) changes in U.S. government telecommunication regulations, and (c) business practices in the United States related to mobile cell phones. Most of these challenges do not exist in Europe and other countries, nor do they threaten the ability of U.S. telephone surveys to reach special target populations (e.g., an organization's customers or clients).

This chapter provides an overview of telephone survey methods, including (a) this mode's advantages and disadvantages, (b) sampling frames, (c) selecting respondents within a household, (d) gaining cooperation, (e) measurement issues, and (f) post survey adjustments that need to be performed prior to analyzing data. It also outlines the steps required to conduct a telephone survey. The relatively recent impact of cell phones on telephone surveys in the United States is discussed throughout.

14.1. Advantages and Disadvantages of Telephone Surveys

14.1.1. Advantages

Although many fail to recognize it, an important advantage of telephone surveying over other survey modes is that it provides the opportunity for continuous quality

Handbook of Survey Research, Second Edition

ISBN: 978-1-84855-224-1