

- Smith, T. W. 1982. "House Effects and the Reproducibility of Survey Measurements: A Comparison of the 1980 GSS and the 1980 American National Election Study." *Public Opinion Quarterly* 46: 54–68.
- Suchman, L., and B. Jordan. 1990. "Interactional Troubles in Face-to-Face Survey Interviews." *Journal of the American Statistical Association* 85: 232–241.
- Tartter, V. C. 1980. "Happy Talk: Perceptual and Acoustic Effects of Smiling on Speech." *Perception and Psychophysics* 27: 24–27.
- Traweek, S. 1988. *Beamtimes and Lifetimes*. Cambridge, MA: Harvard University Press.
- Turner, C. F. 1984. "Why Do Surveys Disagree? Some Preliminary Hypotheses and Some Disagreeable Examples." Pp. 159–214 in C. Turner and E. Martin (Eds.), *Surveying Subjective Phenomena*. New York: Russell Sage.
- van der Zouwen, J., W. Dijkstra, and J. H. Smit. 1991. "Studying Respondent-Interviewer Interaction: The Relationship between Interviewing Style, Interviewer Behavior, and Response Behavior." Pp. 419–37 in P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowitz, and S. Sudman (Eds.), *Measurement Errors in Surveys*. New York: Wiley.
- Weiss, C. 1968. "Validity of Welfare Mothers' Interview Responses." *Public Opinion Quarterly* 32: 622–633.
- Weiss, C. 1970. "Interaction in the Research Interview: The Effects of Rapport on Response." Pp. 17–20 in *Proceedings of the Social Statistics Section of the American Statistical Association* (August). Washington, DC: American Statistical Association.
- West, C. 1984. *Routine Complications: Troubles with Talk between Doctors and Patients*. Bloomington: Indiana University Press.
- Wilner, P. 1985. "The Main Drift of Sociology between 1936 and 1984." *History of Sociology* 5: 1–20.
- Whalen, J., and D. H. Zimmerman. 1998. "Observations on the Display and Management of Emotion in Naturally Occurring Activities: The Case of 'Hysteria' in Calls to 9-1-1." *Social Psychology Quarterly* 61: 141–159.
- Whalen, J., D. H. Zimmerman, and M. R. Whalen. 1988. "When Words Fail: A Single Case Analysis." *Social Problems* 35: 335–362.
- Zimmerman, D. 1992. "The Interactional Organization of Calls for Emergency Assistance." Pp. 418–469 in P. Drew and J. Heritage (Eds.), *Talk at Work: Social Interaction in Institutional Settings*. Cambridge: Cambridge University Press.

CHAPTER 16

How Uniform Is Standardization? Variation Within and Across Survey Research Centers Regarding Protocols for Interviewing

Jocelyn S. Viterna and Douglas W. Maynard

INTRODUCTION

The quality of survey data relies heavily upon standardization in the survey interview. According to Fowler and Mangione (1990:14), "in all sciences, meaningful measurement occurs by applying the same procedure across a set of situations so that differences in the readings that result can be compared and interpreted as indicating real differences in what is being measured." To achieve the "same procedure" in survey research, interviewers are trained to expose each respondent to the survey instrument in a precise and replicable—that is, standardized—manner. This allows researchers to assume that any variation in the resulting answers is directly attributable to differences in what is being measured among respondents.

Recently, scholars have questioned whether or not the ideals of standardization are attainable given the human interaction of a survey interview (Beatty, 1995). Maynard and Schaeffer (2000; this volume, Chapter 1) argue that full standardization or *mechanical objectivity* (Porter, 1995) is never fully attainable because interviewers must constantly employ tacit, commonsense knowledge (Garfinkel, 1967; Polanyi, 1958) to negotiate and complete the interview. If this is true, then the question is not whether centers for survey research depart from standardization, because by necessity they will have to. The question is how much such departure occurs and to what extent it interferes with the quality of survey data.

Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview
Edited by Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora Cate Schaeffer,
and Johannes van der Zouwen
ISBN 0-471-35829-0 Copyright © 2002 by John Wiley & Sons, Inc.

In this chapter, we ask a question related to concerns about the attainability of standardization: "How uniform are concepts of standardization across survey centers?" We also explore how consistent survey centers are in their implementation of standardization. To anticipate our main finding, survey centers are inconsistent in their requirements for standardization. We account for this variation by reference to practitioner's use of tacit and commonsense knowledge. We extend this analysis by using a sociology of science framework and pointing to contingencies of survey interviewing and the influence of local cultures for devising solutions to these contingencies. Some survey centers appear to relax standardization guidelines purposefully and carefully, while still pursuing the goal of generating good-quality data. Other centers with the same goal are stricter in their adherence to standardization. By documenting and analyzing the disparities of standardization requirements, our aim is to add clarity to the debate about the importance and feasibility of standardization within the survey interview.

The variation among survey organizations in how their interviewers perform is a topic of importance in the field of survey research (and indeed to all social scientists who rely on survey centers to collect their data), yet few previous studies have offered insight into what variation occurs and how that variation affects data quality. Our analysis is an exploratory and preliminary attempt to address this gap. In this study, we document what interviewers are *trained* and *supervised* to do, not what they *actually* do on the job. Because our results are based almost entirely on training programs and procedures, we must be careful in extrapolating our findings to how training directives are enacted within the actual survey interview. However, based on a study by Fowler and Mangione (1990), we believe that interviewer training and supervision are crucial factors in determining whether and how standardization guidelines are implemented in actual interviews. We argue that the variation in standardization guidelines presented to an interviewer during training means different practices during the actual collection of data.

HISTORICAL EMERGENCE OF STANDARDIZATION

When interviewers were first used for academic research in the 1930s, standardization was not the norm. Researchers "viewed interviewers as an escape from the restrictions of closed, self-administered questionnaires and were anxious to use this potential freedom" (Beatty, 1995:148). When polling centers introduced standardization, academics were critical because it hindered the development of *rapport* with the respondent (Bradburn, Sudman, and Associates, 1979; Lazarsfeld, 1944; Richardson, Dohrenwend, and Klein, 1965). Previously, well-trained interviewers had been expected to standardize the *meaning* of the question and were given freedom to modify the question *wording* in the way they deemed best for each individual situation (Converse, 1987:40).

Although standardization was not initially the norm, it quickly began to achieve dominance as a research technique largely due to concerns about quality of data (O'Muircheartaigh, 1997:13-14). From its inception in 1941, the National Opinion Research Center (NORC) required interviewers to read

questions verbatim (see Williams, 1942). By the 1950s two other pioneering survey research centers, the Bureau of Applied Social Research at Columbia University, and the University of Michigan Survey Research Center, had followed suit (Converse, 1987). Studies of interviewing by Herbert Hyman (1954) and his colleagues convinced many practitioners that standardized control would improve the reliability of survey data. In 1975, an influential study by Goudy and Potter (1975) added a theoretical basis to the practical appeal. Not only did they find no correlation between response rates and interviewer rapport, but they also found that an emphasis on rapport at the expense of standardization may actually bias the results of a study.

Today, standardization is prized in nearly all survey operations in the United States and in many countries around the world. Given the pervasiveness and ubiquity of survey research, which spans academic disciplines, national boundaries, government and business organizations, public and private sectors, and so on (Maynard and Schaeffer, 2000), it is important to discover whether official regard for standardization (*standardization in theory*) compares to what interviewers are trained and allowed to do (*standardization in practice*). Our study is confined to one nation (the United States) and one kind of survey center (university affiliated), but has wider implications.

DATA AND METHODS

We examine 12 academically based survey research centers. At five of these centers, the first author gathered data through participant observation in interviewer training sessions.¹ These sessions varied in length from 2.5 hours to 2 full days and included instructions on basic interviewing technique, CATI (computer-assisted telephone interviewing) commands and use, and general center regulations. Detailed notes were taken on what information was presented in the training, how it was reinforced through activities or exams, what questions were asked by participants, how they were answered, which interviewing procedures were corrected by supervisors during "mock" interviews, and which were allowed to pass without comment.

At four of the five centers visited, staff members who oversaw interviewer hiring, training, and supervision were interviewed. These interviews dealt with the history, budget, hiring procedures, ongoing training programs, interviewer monitoring and evaluation requirements, and the mission statement of the survey center. In addition, any discrepancies between what was taught during interviewer training and what was taught by accepted texts on standardization procedures were discussed. For example, staff members were asked how interviewers were taught to balance building rapport with maintaining professionalism or (if their training or manuals indicated this) why they advocated faster instead of slower rates of speaking.

Finally, the training manuals of these five centers, plus an additional seven training manuals from survey centers throughout the central and eastern United States,² were reviewed and coded to add breadth to the analysis. Inspection of the

manuals from the sites visited, and participant observation at those sites, revealed a strong correspondence between the amount of coverage a topic received in the training manual and the emphasis that topic was given in the actual training program.³ Therefore, we regard training manuals from nonvisited sites as a good indicator of training emphases at those sites and the additional manuals as reliable data for analyzing variation in interviewer training practices.

Data collection, transcription, and coding were completed by the first author in consultation with the second author. Coding involved organizing the data from each center (ethnographic notes, interview transcriptions, and manuals) around predesignated standardization guidelines, then determining how closely each particular center adhered to the standardization guidelines set forth in the literature. The analysis of these materials involved a bottom-up set of investigative procedures akin to analytic induction or grounded theory (Charmaz, 1983; Glaser and Strauss, 1967). We discuss the reasons for our coding decisions at particular points in the chapter.

STANDARDIZATION CONTINUUM

Variation was found among survey centers in nearly every facet of interviewer protocol. To appreciate patterning in such variation, we developed a continuum and portray two ideal types on it. The ideal type of survey center on the right of the continuum is characterized by strict standardization and a lack of interviewer autonomy. Here, interviewers are given very little information about the purpose of a study and are required to adhere to a detailed and complex set of rules designed to cover any situation that might arise during an interview. Interviewers speak only the words specifically written on the questionnaire or specifically approved in the training manual. They must follow regulations even in situations where personal judgment suggests that such adherence is confusing or irritating the respondent, or may even lead to inaccurate data.

In the ideal type of survey center on the left of the continuum, interviewers are expected to follow proper standardization techniques but are given permission, sometimes implicitly and sometimes explicitly, to relax the rules if they feel these techniques may lead to a wrong answer or to the respondent's refusal to complete the interview. Here, general training is secondary to the extensive and comprehensive project-specific instructions, when supervisors discuss the purpose and goals of each individual question and the study as a whole. The goal of obtaining accurate data takes priority over proper technique, and the interviewer is considered competent to determine when it is appropriate, and perhaps even necessary, to bend the rules. Our continuum was not derived from but is roughly related to the two poles that O'Muircheartaigh (1999:52-53) describes as having pervaded orientations to the survey interview: It is through interviewer autonomy that respondents are "empowered" (as one pole), while through "detailed specifications of (uniform) permissible interactions" standardization is achieved (as the other pole).

To locate a center on the continuum between the two ideal types, each one is classified according to five components of standardized interviewing procedure. Fowler and Mangione's *Standardized Survey Interviewing* guide (1990) identifies these components.⁴ For each characteristic, we determine if a center's policies are more consistent with maintaining strict standardization or with emphasizing interviewer autonomy. If the policies define and advocate adherence to protocol, we assign a +1 to the center, thus moving that center one step closer to the standardized ideal type on the right. Likewise, if the policies are consistent with interviewer decision making, we assign a score of -1 to the center, and it moves one step further left, closer to the autonomous ideal type. In situations where no information is available on a particular interviewing component, we also assign a score of -1 to the center because it implicitly leaves that particular interviewing behavior to the interviewer's discretion.

In our later discussion, we add the scores for the 12 centers on each of the 5 components. This gives a condensed indication of where the center is located on the standardization continuum. We add the scores across centers for each component, which creates a summary score for each component. When the sum is near zero, this may indicate lack of consensus regarding how autonomous or standardized interviewers should be on that component. When the sum is strongly in the plus range, it indicates that the centers have consensus in the direction of standardization. When the sum is strongly in the minus range, centers are convergent in the direction of autonomy. As we will see, the tendency among our centers is toward autonomy.

Contrary to past discussions in the standardization debate (Beatty, 1995:150-151), we contrast *standardization* with *interviewer autonomy* rather than with *rapport*. Although strict standardization can (but does not necessarily) inhibit rapport, and greater autonomy can (but not necessarily) enhance rapport, developing rapport is not the only reason for relaxing standardization guidelines. For example, relaxed standardization guidelines often allow the interviewer to aid the respondent in understanding survey questions. For purposes here, it is sufficient to contrast standardization with interviewer autonomy.

Although some investigators may not agree that standardization is naturally opposed to interviewer autonomy, we argue that this is a helpful distinction, especially given the exploratory nature of our study. Granted, even the most standardized of interviewers must practice some autonomous decision making. For example, an interviewer must decide when a respondent's answer needs to be probed for clarification, and what probes will generate a codable answer without biasing data. The distinction, however, is that within a highly standardized center, even the autonomous choices allowed to the interviewer must be taken from a tightly constrained set of alternatives geared according to question type. In a less standardized center, the instructions for probing will be much less specific, and interviewers have more latitude in how they probe. Thus, we argue that for this preliminary look at standardization, interviewer autonomy provides a useful contrast to interviewer standardization, since standardization necessarily implies more limitations on the autonomous decision making of an interviewer.

ANALYSIS OF STANDARDIZATION COMPONENTS

To analyze variation in standardization, we examine five aspects of standardized survey interviewing procedures identified by Fowler and Mangione (1990) as relevant to the substantive portion of the survey interview:⁵ pace, feedback, professional objectivity, probing, and monitoring procedures. For each component, we first discuss the existing literature for that particular component, then analyze the variation within the collection, and conclude with suggestions for how such criteria can be used to locate the center on the standardization continuum. The initial classification of each component is based on information obtained from center manuals. Supporting information from observations and interviews is given after the analysis of manuals and is clearly indicated in the text as separate from the data from the manuals.

Pace

Pace refers to the speed at which the interviewer reads the questions to the respondent in the substantive part of the interview. Although no direct link between pace and data quality has been documented, 9 of the 12 centers participating in this study include instructions for pace in their training manuals. Two theories of pace are consistently found in the data. On one hand, 5 centers follow the suggestion of Fowler and Mangione (1990:71) that "having interviewers speak slowly is one important and practical way in which researchers can increase the standardization across interviews." This suggestion is followed in manuals by such statements as, "Read at a slow, two-word-per-second pace to make sure all respondents concentrate on the questions and hear them clearly" (Center O)⁶ or "Reading each question slowly and thoughtfully suggests to the respondent that he/she take the same care in answering the questions correctly" (Center C). Center supervisors indicate that insisting on a slow pace also reduces the possibility that the interviewer may change the wording of the question or make coding mistakes, thereby further increasing the level of standardization.

On the other hand, four centers direct their interviewers to speak either in normal, conversational tones or to use faster-than-normal rates of speech during the substantive part of the interview. Examples:

Interviewers are judged as being more competent and professional, more friendly to the respondent, and less likely to be refused data when they have the following voice qualities: greater variation in pitch, relatively loud voice, *relatively fast rates of speaking*, and clearer pronunciation (our emphasis). (Center J)

Speak slowly enough to ensure every word is heard but quickly enough to ensure the interview moves ahead promptly. (Center R)

The common thread among these centers is greater autonomy for the interviewer. The interviewer is not given a static rate (i.e., read at a two-word-per-second pace) and may be allowed to infer what respondents need or desire and then to act accordingly. Furthermore, the goals of slow pacing—to ensure fewer

Table 16.1 Reading Pace

Recommended Pace for Reading Questionnaire	Centers												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
Questions should be read at a conversational pace OR at a faster-than-normal pace OR the pace of the interview is left to the interviewer's discretion.	X			X*				X*	X	X	X	X*	-7
Questions should be read at a slower-than-normal pace.		X	X		X	X	X						5
Standardization score	-1	+1	+1	-1	+1	+1	+1	-1	-1	-1	-1	-1	-2

mistakes and more consistent delivery across interviewers and thus increase standardization—may be lost when interviewers are allowed to personally define a "conversational" or "relatively fast" rate of speech.

The recommended pace for reading questions provides the first measure of where survey centers fit on the standardization continuum (see Table 16.1). For centers that specify a slow pace, and thus a more standardized means of collecting data, a +1 has been assigned, moving the center toward the right end of the continuum. For centers allowing a "normal" or "faster" pace, so that interviewers may decide what is best for the interview at hand, a -1 has been assigned, moving the center to the left. Centers that did not mention pace in their manuals or training sessions were also assigned a -1, but an asterisk is used to indicate that the interviewer's autonomy in regulating pace is implicitly (by omission of statement about pacing) and not explicitly allowed.⁷ In the lower right-hand corner of the table, the negatives and positives are summed as an indication of overall consistency between centers on this topic.⁸ The summary score for pace is -2, indicating low consistency among survey centers. Roughly speaking, a little over half the centers in this study allow the interviewer to determine the appropriate pace for the interview.

Feedback

In everyday conversations, if one asks a question it is common to acknowledge the answer by responding reciprocally, sometimes approving or disapproving of an answer, or registering it with an "oh." These *third-turn* comments in an interviewing sequence (i.e., QUESTION-ANSWER-ACKNOWLEDGMENT) are considered by ethnomethodologists to be the crucial mechanism by which participants recognize institutionalized activity frameworks (Heritage, 1984:290). In conversation, the content of the third turn after a question-answer sequence can indicate to the answerer a variety of things, such as whether the answer is heard, understood, accepted, correct, good, and so on. In institutional or

organizational settings, therefore, the third turn is often regulated (see discussion in Chapter 1, this volume). For instance, in news interviews, after obtaining an answer from a subject, the host refrains from third-turn commentary, thereby indicating "neutrality" and acting as a conduit for the answer to reach an "over-hearing" audience (Clayman, 1988; Heritage and Greatbatch, 1991). In educational testing interviews, administrators are permitted to use "okay," "thank you," and the like but are not to use evaluative terms indicating the correctness of an answer (Marlaire and Maynard, 1990; Maynard and Marlaire, 1992). By contrast, teachers asking *known information* questions in the classroom produce third turns that indicate whether a student's answer is right or wrong (Mehan, 1979).

The survey interview is most like the testing situation; survey designers accept the importance of third-turn utterances in acknowledging a respondent's answer but recognize that fine gradations in these responses potentially influence the respondent and create measurement error. A few studies demonstrate a significant relationship between the use of value-laden feedback and an apparent bias in the respondent's answers (Fowler and Mangione, 1990; Marquis, Cannell, and Laurent, 1972). The suggestion is that when respondents give answers appearing to be thoughtful, complete, and within the constraints of the given question, they are rewarded with a "value neutral" positive comment about their interviewing performance, not about the content of the answer (Cannell, Oksenberg, and Converse, 1977; Marquis, Cannell, and Laurent, 1972).

Ten of 12 centers in this study made specific reference to feedback use in their training materials. The proposed use of feedback varied markedly along three different measures: frequency, purpose, and content. First, feedback varied in its permitted frequency; only two centers specified how often feedback should be given. At one of these, interviewers were instructed to give feedback after every acceptable answer. At the other, interviewers were instructed to give feedback after 30–50 percent of the respondent's acceptable responses. The remaining 10 centers did not indicate when or how often feedback should be used.

In Table 16.2, centers that specifically indicate how often feedback should be used are given a score of +1, moving them further to the right on the standardization continuum because all respondents in these centers should receive roughly the same amount of feedback. Survey centers that do not regulate the frequency of feedback are assigned a value of -1, moving them further to the *autonomous* side of the continuum since the interviewer chooses how much feedback is desirable. Survey centers that do not mention feedback at all are again differentiated in the table with an asterisk and assigned a -1 since the omission of feedback directions implicitly allows the interviewer to choose the frequency of feedback. The score of -8 indicates that there is very high consistency in the regard for frequency of feedback usage across survey centers. Most centers allow the interviewer to determine how often feedback will be given.

A second mode of variation in feedback usage is found in how or whether a center defines its purpose. Of the 10 centers that recommended the use of neutral feedback, 4 indicated that feedback should be given after acceptable answers and withheld after unacceptable answers. These are given a score of +1

Table 16.2 Frequency of Feedback Usage

Frequency of Usage	Centers												
	1	2	3	4	5	6	7	8	9	10	11	12	Total
Frequency of Usage													
Training manual does not indicate how frequently feedback should be used.	X												-10
Training manual specifically indicates how often value-neutral feedback should be used.		X											2
Standardization score	-1	+1	-1	-1	-1	+1	-1	-1	-1	-1	-1	-1	-8

Table 16.3 Purpose of Feedback

Purpose of Feedback	Centers												
	1	2	3	4	5	6	7	8	9	10	11	12	Total
Training manual advocates use of feedback as a means of "building rapport" with the respondent OR makes no specific mention of feedback.			X	X	X		X*	X*	X				-8
Training manual advocates use of feedback as a way of "training" the respondent.						X				X			4
Standardization score	+1	+1	-1	-1	-1	+1	-1	-1	-1	+1	-1	-1	-4

in Table 16.3, indicating an orientation to standardization on the part of these centers. The remaining 6 centers stated only that the purpose of feedback is to build rapport with the respondent by demonstrating that the interviewer is listening, interested, and nonjudgmental. Thus we infer that feedback would be permitted for unacceptable as well as acceptable answers. These centers receive a score of -1 in Table 16.3 for their orientation to autonomy. The two centers not mentioning feedback in their manuals are also assumed to allow feedback for both acceptable and unacceptable answers and are scored -1. This results in a summary score of -4 and indicates that most centers (2/3) allow interviewers to use feedback as they see fit. Nevertheless, one-third of the centers advocate stricter use of feedback to shape respondent performance.

A third mode of variation in feedback use is found in the suggested feedback phrases themselves. Interestingly, many of the same phrases appeared in the majority of the survey manuals in this study with word-for-word replication from one center to the next.⁹ Table 16.4 lists all feedback phrases from all manuals, along with the number of manuals that advocate use of that phrase.

Given the number of shared feedback phrases, it is surprising to find that the *suggested* feedback at some centers is the *forbidden* feedback at others. At three of the centers visited, interviewers were instructed that saying "yes" or "right" was forbidden because it indicated agreement. At one of these three centers, saying "okay" was also forbidden for the same reason. Yet as can be seen in Table 16.4, four other centers recommend in their manuals using "yes" and four recommend using "okay" when giving feedback.¹⁰ In short, some centers advocate the use of feedback that, in the eyes of other centers, may compromise the data.

Recommendations about feedback phrases are difficult to use as a measure to locate a center on our continuum. Two conclusions, however, are warranted. First, there needs to be further research to see if and how feedback affects data

Table 16.4 Frequency of Suggested Feedback Phrases

Acceptable Feedback Options from Survey Center Interviewer Training Manuals	Number of Manuals Advocating Use of This Feedback Phrase
I see.	10
That's useful(helpful) information.	7
That's useful (helpful) to know.	
It's useful to get your ideas on this.	
It is important to find out how you (people) feel about this.	7
It is important to get your ideas on this.	
It is important to get your opinion on that.	
It is important to find out what people think about this.	
Thanks/Thank you.	6
Uh-huh/Uh-hmmm/Mm-hmmm	6
Okay/All Right	4
Yes	4
That's interesting	3
That helps our research a lot.	2
That is helpful for our research.	
That is important for our research.	
Right	1
Thanks, that's the sort of EXACT answer we need.	1
That's important to know	1

quality. Although survey centers and experts on survey design believe that types of feedback do affect respondents' answering (per discussion above), there is very little systematic investigation of this matter and almost none since the 1970s.¹¹ Second, survey center manuals do recognize the possibility of such an effect and offer sometimes conflicting advice about how interviewers should perform their feedback. This indicates a *leaning* toward standardization within centers and inconsistency across centers. The main form of agreement across centers is in the recommendation to use "I see."

In addition to finding variation in the feedback phrases listed in the manuals, we observed differences in the attention accorded to feedback in training programs. Three of the five centers visited gave extensive coverage to the issue of feedback both in the training manuals and in the training sessions, while two did not. Those centers that addressed feedback provided a list of suggested phrases for the interviewers to post in their carrels. At two of these three centers interviews were observed in progress and interviewers appeared to follow their training very closely in both centers. No feedback was heard that was not on the recommended list.

In the other two centers visited, the topic of neutral feedback received only one paragraph in the training manual. Further, neither of these centers made any mention of feedback in their training sessions, nor did they provide lists of "appropriate" feedback for interviewers to use while conducting an interview. It was in these two centers that strongly evaluative feedback was heard. In one instance, experienced interviewers said, "great," "very good," and "good," in mock interviews with their supervisors. Supervisors did not reprimand these interviewers, nor did they discuss inappropriate feedback throughout the rest of the training program or otherwise indicate that such feedback was a violation of good protocol. Indeed at one of these centers an experienced interviewer used phrases including "interesting," "good point," "yeah, that's true," and "no, these aren't the greatest questions" as third-turn utterances after the respondent's answers.

Two implications derive from our findings about feedback. First, lack of regulation at some centers may mean that interviewers are not exposing respondents to uniform interview situations. Second, even when feedback is tightly regulated, centers disagree about what "neutral" feedback is and how and when it should be used.

Professional Objectivity

Ten survey centers discuss in their manuals the need for *professional objectivity*, being *interpersonally neutral*, or establishing a *balanced rapport* with the respondent. Such discussions are consistent with Fowler and Mangione's (1990:48-49) recommendation. For an interview to be standardized, the interviewer must not "volunteer personal information to the respondent about life situations, views, or values" and "should be careful that the feedback provided to respondents does not imply any evaluation or judgment about the content of the respondent's answers."

Nevertheless, eight survey center manuals instruct interviewers to convey warmth and pleasantness in their tone of voice, two centers instruct their interviewers to be friendly, and one even suggests using humor to build better rapport

with the respondent. To some extent, these suggestions to be both impersonal and friendly form a dilemma and can create strain for the interviewer.¹² This dilemma is acknowledged in five interviewer manuals. Two examples:

Interviewers perform a delicate balancing act between being congenial with a respondent and maintaining a professional distance. A respondent will be more cooperative when the interview is a pleasant experience. However, maintaining a distance is essential to avoid biasing the interview. If a respondent thinks of the interviewer as a newly found "friend," there is the possibility that the answers may not be as truthful as they would be if a professional rapport were maintained. (Center C)

In every interview, interviewers must play several roles: that of an individual who builds a friendly and trusting relationship with each respondent; and that of a technician who gathers information to select a respondent and who applies standard techniques to the interview. (Center N)

However, no manual specifies or explicates how interviewers are to deal with this dilemma. Rather, learning how to balance rapport and neutrality appears to be something interviewers learn on the job through listening to other interviewers around them and through feedback from their supervisors (Goodwin, 1998; Heil, 1999; Lavin and Maynard, this volume, Chapter 15).¹³

When survey staff were asked how they train interviewers to deal with the contradiction between remaining neutral and establishing rapport, they consistently asserted that interviewers learn what an acceptable level of rapport is through on-the-job experience. For example:

It's not unusual for one of us to go over and... get with an interviewer and say, "well, that was ok, what you were doing was helping the flow of the interview and that sort of thing, but when they engage in talking about their dog, you know you really should have worked to disengage from that conversation and move on." We take it very much on a case to case basis. We don't have hard and fast rules about it, we don't tell people that if they laugh they're fired; likewise, we certainly discourage... them getting very friendly with people. (Center E)

In the four survey centers where the first author observed interviewers at work, she found wide variation in the definition of *professionalism*. In two centers, no emotional response was allowed from interviewers whatsoever. Supervisors corrected an interviewer for laughing at one center; and at the other, supervisors who heard interviewer laughter when monitoring would identify it as "nonstandard feedback" on evaluation forms. At a third center, however, laughter and casual banter were not sanctioned and were, accordingly, audible parts of interviews. Although interviewers did not share opinions or personal information, they did relate to the respondents in a friendly, even personal manner. At a fourth center, although the interviewers were more formal, some casual banter and discreet chuckles were noticeable. Intrigued, Viterna asked one of the interviewers if, for example, they were allowed to laugh at a respondent's joke. He replied, "Oh, yeah. You'd have to, no matter how lame the joke was." While he went on to point out that rapport-building mechanisms should never influence an

Table 16.5 Levels of Professional Objectivity

Professional Objectivity Guidelines	Centers												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
Survey center instructs interviewers to be flexible, friendly, warm, and conversational when asking questions.	X*		X					X*	X	X	X	X	-7
Survey center instructs interviewers to be neutral, serious, and professional at all times.		X		X	X	X	X						5
Standardization score	-1	+1	-1	+1	+1	+1	+1	-1	-1	-1	-1	-1	-2

interviewer's remarks or alter the wording of the questions, he and other interviewers had learned that building rapport was more important than being formal and objective at all times.

To operationalize professional objectivity as an element of standardization, the wording in the professionalism section in each manual was analyzed. If a manual made reference to being serious, professional, speaking in a "matter of fact way," or, in general, keeping respondents "focused" or on "track," then a score of +1 was assigned. If the manual did not mention the issue of seriousness, professional distance, or focus, and instead emphasized warmth, friendliness, "communicating sympathy" or "building a friendly, trusting relationship," then we assigned a score of -1. Centers that did not discuss either issue were assigned a -1 as well, since in the absence of specific instructions for neutrality, we assumed that interviewers are permitted to be as engaging as possible with the respondent. See Table 16.5.

Probing

Our discussion has centered on how interviewers ask questions and respond to answers in a standardized fashion. Now we turn our attention to situations where the respondent breaches standardization by not answering a question in a way assumed by standardization. For example, a respondent may answer a question by saying "I don't know" or may give a response that does not fit into a designated response category. Given the likelihood of unpredictable situations, how are interviewers to obtain a "good" answer while still remaining standardized? Survey researchers have developed a well-known but complex procedure called *nondirective probing* to deal with such situations. In theory, nondirective probing allows interviewers to standardize the procedure for achieving a good answer by limiting and specifying the type of exchange that may take place, even if the actual exchange itself varies from situation to situation.

Probing Generally, Probing Closed Questions, Probing Open Questions

For purposes of clarity, we classify all probes into three categories: General probes, probes for closed questions, or probes for open questions. First, general probes are ones that can be used any time respondents express uncertainty and withhold an answer. For instance, if a respondent answers a question with "I don't know," the interviewer may try to elicit a specific answer by allowing an expectant pause, by saying "Generally,..." and repeating the question, or by assuring the respondent that "there are no right or wrong answers, we are only interested in your opinion." Furthermore, if a respondent asks for clarification of the meaning of the question, the interviewer is not allowed to give personal definitions, but

Table 16.6 Frequency of Suggested "General Probes"

Recommended General Probes from Survey Center Interviewer Training Manuals	Number of Manuals Advocating Use of This Probe
Repeat the question/repeat the answer categories.	10
Whatever it means to you./Whatever _____ means to you.	7
Whatever you think of as _____.	
Anything you would call _____.	
Whatever interpretation seems best to you _____.	
It's important that the question be answered as best you can in terms of the way it's stated.	
An expectant pause.	6
Generally, what would you say?	6
Generally/Overall/For people in general/On average/On the whole/Usually/ Taking everything into consideration/In the country as a whole/ In most cases/	
Most of the time/Generally speaking... (then repeat the question or the response categories).	
In your opinion... (repeat question or response categories).	5
Of course no one knows for sure (but we are interested in people's opinions).	
Of course there are no right or wrong answers.	
There are no right or wrong answers, just your ideas on it.	
We're just interested in what you think.	
We are just trying to get people's ideas on this.	
What do you think?	2
What do you expect?	
I see.	2
I understand, but I need to know... (repeat question).	
We all hope, but...	1
Yes, but...	1

is expected to prompt the respondent into giving a codable response by saying, "Whatever interpretation seems best to you."

As with feedback, the same probes were often recommended by different survey centers. Table 16.6 lists all phrases found in the 12 interviewer training manuals that could be classified as general probes, along with the number of manuals that advocate use of each particular probe.

The second category of probes includes those that correspond to closed questions, which are designed to have a limited set of responses from which respondents must choose their answer. The most common probes for closed questions work to help the respondent fit her answers into one of the predesignated response categories. Take, for example, the hypothetical question, "How often do you typically brush your teeth: less than once a day, once a day, twice a day, three times a day, or more than three times a day?" If the respondent answers, "two or three times a day," the interviewer would then probe with "Would it be closer to two or three?" before coding the response. Table 16.7 lists all probes for close-ended questions as consolidated from the 12 interviewing manuals in this study. A high level of consistency of probes across centers is evident.

The third and final category of probes involves open questions, which ask the respondent to answer a question without predesignated answer categories. To obtain as much information as possible from the respondent, interviewers are often instructed to follow an answer to an open question with a probe such as "tell me more" or "why do you feel that way?" In addition to asking for more information, probes for open questions are often designed to obtain sufficient information for accurate coding. "Could you be more specific?" is a common probe for encouraging respondents to clarify their answers. Probes for open questions are summarized in Table 16.8.

Table 16.7 Frequency of Suggested Closed-Question Probes

Recommended Closed-Question Probes from Survey Center Interviewer Training Manuals	Number of Manuals Advocating Use of This Probe
Which would be closer (to the way you feel)?	9
Which is closer?	
Which comes closest?	
What category comes closest to your feeling?	
Which figure do you think comes closest?	
Would it be closer to _____ or _____?	
I need to fit your answer into one of our categories. Please choose the one that's closest to the way you feel.	
What is your best estimate/guess?	2
What would be your best estimate?	
Could you give me your answer in terms of (repeat response categories).	2
We need a specific (number, unit, etc).	

Table 16.8 Frequency of Suggested Open-Question Probes

Recommended Open-Question Probes from Survey Center Interviewer Training Manuals	Number of Manuals Advocating Use of This Probe
Are there any other reasons you feel that way?	9
Do you have any other reasons for feeling as you do?	
(Are there) Any other reasons?	
(Is there) Anything else?	
What else (can you think of)?	7
What others?	
Thinking about _____ a bit more, what else is there about it that comes to mind?	
Please tell me more.	
Would/could you tell me (a little) more (about that)?	
Would/could you tell me more about your thinking on that?	
Could you give me more information about your thought on _____?	
I'd like to know more about your thinking about that.	
How do you mean (that)?	
What do you mean (by that)?	
What/How do you mean exactly?	7
Could you be (a little) more specific (about that)?	
Please be more specific.	6
An expectant pause.	
Repeat what the respondent has said.	4
What do you have in mind?	
Will/would you tell me what you have in mind.	4
Why do you say you feel like that?	
Why do you feel that way?	2
We'd like to know how you feel.	
I see.	2
I understand, but I need to know... (repeat question).	
What do you think causes that?	1
Why do you think that is so?	
I don't understand what you mean.	1
Ok/yes/uh-huh/that's interesting... (followed by expectant pause).	

Overall, there are some widely accepted probes. That is, at least three-quarters of the centers agree that general probing should be done by repeating the question or answer categories, that probing for closed questions should be done with a question that asks for an answer that is "closest" to a predetermined category, and that probing for open questions should be done by asking for "other reasons" or "what else" in regard to an answer.

Beyond this agreement on probes and the need to implement nondirective probing—the importance of which is mentioned in 11 of the 12 centers in this study—training about probing *practices* differs greatly from center to center. Three measures of variation in probing practices demonstrate this lack of uniformity: the degree of specificity in regard to probing, the permissible quantity of probing, and the interviewer's autonomy in deciding whether or not to repeat all response categories.

Specificity of Probing

Deciding whether or not to probe the answer given by the respondent is the first issue interviewers face in the probing process. It is very difficult for survey center staff to design guidelines specific enough to cover all situations where probing might be necessary. Thus, we found very little information in the training manuals explaining what constitutes a complete response and what requires a probe. Instead, learning when to probe an answer appears to be an on-the-job task. Take, for example, a question such as, "Would you consider your professors helpful, neither helpful nor unhelpful, or not at all helpful," and imagine the respondent replied, "Oh, they are just super!" In one variation (at two of our centers), interviewers would be instructed simply to code "helpful" and proceed with the questionnaire. It would be taken as obvious that the respondent would choose this option, even if he did not use the exact wording of the response categories, and repeating the response options would only slow down the interview and irritate the respondent. In a second variation (one survey center), the interviewer would be instructed to say, "so that would be helpful?" and if the respondent answered yes, the interviewer would then code "helpful" and proceed. With this probe, the interviewer acknowledges the answer given by the respondent, and still gives a quick reminder that the respondent should answer with one of the predesignated responses.¹⁴ In a third variation (one center), however, the interviewer would be required to repeat the categories: "So would you say helpful, neither helpful or unhelpful, or not at all helpful." In other words, interviewers at this center could not code "helpful" until re-reading all response categories and eliciting the respondent's unaided choice of one. Accordingly, even when an interviewing procedure such as probing is standardized with an instruction to probe inadequate answers, the local interpretation of such an instruction may implicate strategies that are not uniform from center to center.

The specificity of probing instructions differs in the manuals as well. Although we have organized probing phrases according to question type for the purposes of this chapter, not all survey centers provide the same or similar categorizations for their interviewers. In some survey centers, a short list of three or four probes is simply accompanied by directions to use the probes if the respondent provides an answer that is vague, unclear, or incomplete. The manual does not relate these probes to question type, nor does it limit interviewers to using only probes in the interviewing manual. Given the lack of direction in deciding which probe to use, these centers are scored a -1 on the continuum; interviewers are given much autonomy in determining how to probe a given situation.

Table 16.9 Probing Specificity

Probing Specificity	Centers												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
Training manual does not categorize probes by use OR does not discuss probing at all.	X			X			X	X*	X	X	X		-7
Training manual explicitly dictates what are acceptable probes and which probes should be used for which situation.		X	X		X	X						X	5
Standardization score	-1	+1	+1	-1	+1	+1	-1	-1	-1	-1	-1	+1	-2

This is in contrast to survey centers that provide interviewers with a list of specific probing options organized by the type of question for which each probe is designed. In these centers, when interviewers believe a probe is necessary, they must select a probe from a list that will meet the needs of the particular situation and then read that probing phrase verbatim.¹⁵ These centers are assigned a score of +1 because they establish specific guidelines regarding which exact phrase should be used in each type of situation. The results of this coding can be found in Table 16.9. As shown by the summary score of -2, there is strong variation across centers with respect to the specificity of probing guidelines. While five centers give very specific probing instructions, seven centers do not, and permit the interviewer to decide how best to handle inadequate answers.

How Many Times to Probe

A second measure of probing standardization concerns how many times interviewers are instructed to probe a respondent's answer before giving up and moving on with the interview. Four of the centers in this study give strict guidelines. Three state that the interviewer should only probe once so "it won't seem like you're badgering the R," and the fourth instructs its interviewers to probe twice. These centers all receive a score of +1 because the number of probes is prespecified.

On the other hand, eight centers allow interviewers significant autonomy in determining how many times they probe a respondent's answer. Two of these centers suggest probing only once but leave it to the discretion of the interviewer to determine if a second probe is a viable option without damaging rapport. In one of these, a trainer stated, "You must probe once and may probe twice, but if the respondent is obviously not going to answer, just code a refusal or a don't know and move on. . . it's not worth getting hung up on" (Center Y). Two other centers instruct interviewers to probe until they have received a complete answer or until the respondent has become frustrated. The remaining four centers in this category do not mention how often interviewers should probe a respondent's answer, and

Table 16.10 How Many Times to Probe a Respondent's Answer

Probing Frequency	Centers												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
Training manual allows interviewer to decide how many probes would be beneficial given the specifics of the situation.		X	X*	X	X	X	X*	X*	X*				-8
Training manual explicitly specifies how many times an interviewer may probe a respondent's answer.	X									X	X	X	4
Standardization score	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	+1	+1	-4

are designated with an asterisk to indicate that the interviewer's autonomy comes from omission of rules rather than explicit permission to be autonomous. The results of this classification are found in Table 16.10. The summary score of -4 indicates some agreement among survey centers in allowing interviewers limited autonomy in deciding how often to probe an incomplete answer.

Repeating Response Categories

A final means of evaluating the standardization of probing concerns closed questions or those with predetermined response options. If a respondent does not choose an answer that matches the answer categories, then the interviewer must probe to find out which alternative is closest to the way the respondent feels. For example, if a question asked the respondent to rate her work environment as very good, good, fair, or poor, and the response was, "not very good," the interviewer would need to probe to find out which category is closest to the respondent's view. In such circumstances, Fowler and Mangione (1990:40) state that it is important to re-read all possible responses to the closed question because respondents "respond to the number of categories and the position of a category on a scale, as well as to the words, when classifying themselves. A truncated version of a set of responses is not the same stimulus, and it will affect the answers."¹⁶ This line of thought is reflected in four survey center manuals where interviewers are instructed always to repeat the entire set of response categories, even if they feel certain responses have already been eliminated. These centers receive a score of +1 on the standardization continuum.

Four other centers instruct interviewers to re-read only those response options that have not already been clearly eliminated by the respondent. To demonstrate, in the above example about work environment, when the respondent answers "not very good," the interviewer would probe, "so would you say fair or poor?" This results in less overall standardization because interviewers have discretion to determine whether or not a response is codable and which responses have or

Table 16.11 Repeating Response Categories

Instructions for Repeating Response Categories	Centers												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
The interviewer is allowed to determine which response categories need to be repeated to determine the correct answer when probing a close-ended question.	X		X		X*	X		X*	X	X		X*	-8
Training manual states that the entire set of response categories must be re-read when probing a close-ended question.		X		X			X					X	4
Standardization score	-1	+1	-1	+1	-1	-1	+1	-1	-1	-1	+1	-1	-4

have not been eliminated. These survey centers receive a score of -1 on the standardization continuum.

Table 16.11 shows that four centers did not take a position on how many response categories must be repeated when probing close-ended questions, therefore also leaving interviewers to decide the number of response categories to be repeated. Two-thirds of the centers in this study allow their interviewers to decide how many response categories must be repeated to code an appropriate response to a closed question. While survey centers lean toward autonomy in repetition, there still is variation, with some centers regulating the matter. Overall, centers agree that probing should be nondirective and prescribe common phrases to achieve this. However, they differ in the degree of specificity in using probes, how many times to probe, and whether and which response categories should be repeated.

Monitoring Procedures

The centralization of interviewing, made possible by CATI and other telephonic surveying, allows supervisors to monitor interviewers frequently and give immediate feedback regarding their technical performance. This is often cited as one of the greatest benefits of telephone survey interviewing over on-location, face-to-face interviews. In a room near to or part of the space where interviewers work, supervisors use headphones to listen to the interviewer and respondent, and may also watch, on their own computers, the interviewer's computer screen as she pages through the instrument.

According to Fowler and Mangione (1990:133), frequent and detailed supervision cannot take the place of good interviewer training. If an interviewer is inadequately trained before beginning her job, then there is little chance for any real improvement through monitoring and supervision. Still, the monitoring of

interviewers is essential for maintaining high interview quality. The more experienced interviewers become, the more they rely on their own judgment about how to interview. Otherwise put, interviewing maturity leads to lower standardization levels (Bradburn, Sudman, and Associates, 1979; Fowler and Mangione, 1990).

Monitoring interviewing technique is only one way in which interviewers are evaluated in survey centers. Other ways include assessing efficiency (e.g., number of interviews per hour), cooperation (e.g., number of completions versus number of people to whom the interviewer spoke), and individual reliability (e.g., attitude, dependability, timeliness, and attendance). Ten of the 12 centers in this study have reported that interviewers are consistently evaluated on all these measures. Despite this claim, however, there is much variation. The frequency of online monitoring and whether or not the evaluation forms focus on standardization provide two key dimensions on which to measure uniformity (or lack thereof) in how evaluation is carried out among the centers.

For monitoring to be effective in encouraging standardized interviewing techniques, it should be frequent and without exact scheduling (so that interviewers know only *that* they will be monitored but not *when*). For example, at one survey center, monitoring is a part of every interview shift, and interviewers receive immediate feedback at least once or more on a shift about their interviewing technique and the way they handle specific problems. Knowing their words will be monitored at least once and possibly more on any shift, interviewers appeared disciplined in their interviewing. At another center, things are different. Throughout the three evenings the first author spent at this center, the monitoring room was never occupied and the lights were off, clearly signaling to interviewers that they did not need to be concerned about being observed and checked. Accordingly, they appeared less disciplined in their behavior during interviews, providing nonneutral feedback, spontaneous definitions, and casual banter in interaction with their respondents.

Frequency of Monitoring

Using information in the manuals and from a follow-up questionnaire, survey centers were classified according to the frequency of online monitoring. All centers that checked each interviewer at least once a shift received a +1 score. Centers that monitored less than once a shift per interviewer received a score of -1. Centers that did not report monitoring procedures in their manuals are indicated with an ND (no data) and are not assigned a value on the continuum, since we cannot infer interviewer autonomy without more knowledge of the monitoring practices in these centers. Table 16.12 demonstrates that most centers (7 out of 10 who reported) do not require once-per-shift monitoring of telephone interviews.

Of the 12 centers in this study, 9 reported that interviewers with more experience were monitored less frequently than newer interviewers. (No information was available for the other 3 centers). These centers all indicated that the minimum monitoring guidelines (i.e., once a shift, once a month, etc.) applied to all interviewers, including those with much experience, but new interviewers were given extra monitoring sessions and extra feedback whenever time permitted. Sometimes this practice was official and printed in the center's manual; other times it

Table 16.12 Frequency of On-line Monitoring

Frequency of On-line Monitoring	Centers												Total	
	1	2	3	4	5	6	7	8	9	10	11	12		
Survey center does not require interviewers to be monitored at least once a shift.				X		X	X	X	X	X	X			-7
Survey center requires that each interviewer is monitored at least once a shift.	X	X			X									3
Survey center did not report monitoring frequency.			ND										ND	2(*)
Standardization score	+1	+1	ND	-1	+1	-1	-1	-1	-1	-1	-1	ND	ND	-4

was informally stated by survey center supervisors. Either way, the practice is at odds with Fowler and Mangione's (1990) argument that more monitoring cannot make up for poorly trained interviewers, that more experienced interviewers tend to relax standardization guidelines more often than less experienced interviewers, and that supervision of experienced interviewers should not be relaxed.

Survey staff at the centers would probably agree with Fowler and Mangione (1990) that supervision does not take the place of solid interviewer training. Instead, extra monitoring sessions for newer interviewers may help them develop the *judgment* necessary to deal with situations that fall outside of what interview protocols cover. Furthermore, survey centers often regard more experienced interviewers as capable of such judgment:

We're encouraging people, especially for experienced folks, to give it your best judgment, give it a try, and just take a note, and at the end of the interview you can make an outline, a thumbnail sketch or whatever of what happened and usually that's fine. Cause usually a lot of times when we get questions people will know the answers; they're just looking for a little reinforcement. (Center Y)

Content of Monitoring

Another approach for comparing the standardization of monitoring practices involves analyzing the content of the monitoring forms used by supervisors to evaluate interviews. Seven of the 12 centers in this study provided copies of their monitoring forms, which vary greatly in size and emphasis.¹⁷ For the purposes of this study, we considered monitoring to encourage standardization when all the core factors highlighted by Fowler and Mangione (1990)—pace, feedback, probing, professional objectivity, and proper question reading¹⁸—were mentioned. Of these five factors, proper question reading and probing are included on all 7 survey centers' monitoring forms. Five centers include a component for professional behavior, 4 include a component for proper pace, and only 2 include

Table 16.13 Inclusion of Core Evaluative Factors on Monitoring Forms

Inclusion of Core Evaluative Factors	Centers												Total	
	1	2	3	4	5	6	7	8	9	10	11	12		
Survey center is missing one or more of the five core evaluative factors on the monitoring form.				X	X	X	X				X	X		-6
Survey center includes all five core evaluative factors on the monitoring form.		X												1
Monitoring form is not available for survey center.	ND		ND					ND	ND			ND	ND	5(*)
Standardization score	ND	+1	ND	-1	-1	-1	-1	ND	ND	-1	-1	ND	ND	-5

a component for proper feedback. Survey centers that include all 5 of the factors on their monitoring forms are assigned a score of +1 on the standardization continuum. Those centers that omit one or more of the 5 core issues are assigned a score of -1.¹⁹ The 5 centers that did not provide monitoring forms for analysis are again indicated by an ND signifying that no data was available for analysis. As shown by the consistency score of -5 in Table 16.13, the pattern is not to include all core standardization issues on the monitoring form.

Informally, supervisors recognize that fully standardized performance (as framed by their monitoring forms) is not possible:

Usually I stress that it [the monitoring form] is based on a model interview that's perfect. The form is, at least. You know, there is an expected amount of error. I mean, it's not just the interviewer who's necessarily to blame for not doing a perfect session, it's also the respondent having difficulty with questions and having to steer them. A perfect [standardized interview] can be extremely difficult if you have to do some persuasion... they're refusing, things like that. (Center Y)

Accordingly, survey center staff recognize that at times complete standardization of interviewer behavior is impossible and that interviewers must use their judgment to determine how best to handle the interview.

In sum, monitoring interviewer technique is important for maintaining standardization within a survey enterprise, but survey centers vary greatly in how often they monitor their interviewers and whether or not they hold their interviewers responsible for key standardization practices such as appropriate feedback or interview pace. Furthermore, interviews with survey staff indicate that in some centers, supervisors view monitoring as an opportunity to help interviewers develop judgment, not just to reprimand them for improper technique. This shows a move toward developing interviewer autonomy on the job, even when such

autonomy may not be evident by reviewing a center's formal guidelines for standardization.

DISCUSSION

Summing positive points for each procedure that favored standardization and negative points for each procedure that favored interviewer autonomy, it is possible to observe patterns intrinsic to specific survey centers and across them (see Table 16.14). We highlight four important tendencies exhibited in Table 16.14.

First, the additive picture of the standardization continuum for individual centers in Table 16.14 shows the extent to which they favor interviewer autonomy over standardization. Summing down columns, a score in the minus range indicates that on at least half the measures presented in this chapter, the center leans more toward autonomy than standardization. Ten out of 12 centers have this tendency. Only 2 centers (centers 2 and 6) are in the other direction and are strictly standardized on at least half of our measures.

Second, while the autonomy permitted interviewers is prevalent, it is also relatively constrained. As Figure 16.1 shows, of the 10 centers that fall to the left of the continuum's center, 5 are in the -1 through -3 range and are close to that center. Another set of 3 centers are at -5 on the continuum. This indicates that 8 centers are relatively standardized, rather than relatively autonomous, on at least some measures. Similarly, one center (6) is only at +1 on the continuum, which means that it is close to being balanced in terms of autonomy and standardization. In short, 9 out of 12 centers vary in their own orientations to standardization and autonomy. Only 3 centers are consistently one way or the other—2 centers (8 and 9) strongly favor autonomy and one center (2) implements strict standardization.

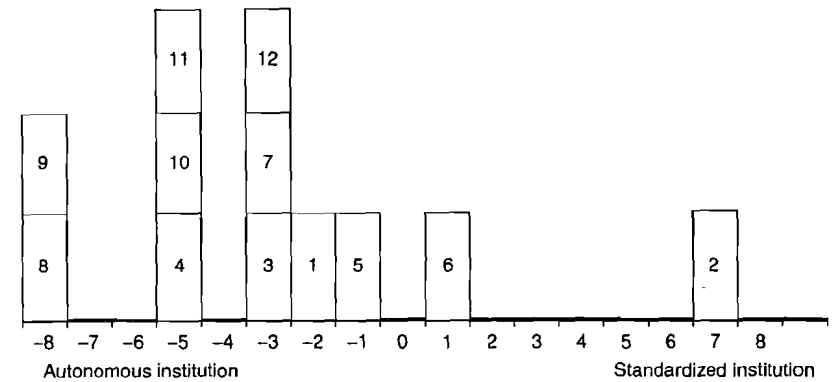


Figure 16.1 The standardization continuum. (Numbers in boxes identify institution.)

Third, this analysis demonstrates the variation across centers in how autonomy is granted. Recall that centers that neglected to mention a particular standardization measure were assigned a negative score for permitting interviewer autonomy, but marked with an asterisk to indicate that the autonomy was permitted implicitly (by lack of instruction) rather than explicitly (by official documentation). Table 16.15 provides a summary of the number of omitted guidelines related to those considered here—pacing, feedback, professional objectivity, probing, and monitoring—by center.

Interestingly, the two centers furthest to the left of the continuum—centers 8 and 9—are very different in their guideline specificity. It is logical to conclude that autonomy at center 8 may be attributed to underspecification or even undermanagement, while autonomy at center 9 more likely reflects a thoughtful position on the part of the survey center staff to allow their interviewers more individual discretion.

Fourth, Table 16.14 allows a look at standardization across standardization components and centers. By summing scores horizontally for each component, it is apparent that each one is negative. This indicates that not one of the practices (pace, feedback, etc.) is strictly standardized in even half the survey centers evaluated. That is, the majority of our centers favor interviewer autonomy on each component of standardization. Frequency of feedback is the component that is most frequently left to the interviewer's autonomous discretion, but feedback

Table 16.14 Total Standardization Scores by Center

Standardization Components	Centers												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
1.1 Pace	-	+	+	-*	+	+	+	-*	-	-	-	-*	-2
2.1 Feedback: Frequency	-	+	-	-	-	+	-*	-*	-	-	-	-	-8
2.2 Feedback: Purpose	+	+	-	-	-	+	-*	-*	-	+	-	-	-4
3.1 Professional objectivity	-*	+	-	+	+	+	+	-*	-	-	-	-	-2
4.1 Probes: Specificity	-	+	+	-	+	+	-	-*	-	-	-	+	-2
4.2 Probes: Repetition	+	-	-*	-	-	-	-*	-*	-*	+	+	+	-4
4.3 Probes: Response categories	-	+	-	+	-*	-	+	-*	-*	-	+	-*	-4
5.1 Monitoring: Frequency	+	+	ND	-	+	-	-	-	-	-	-	ND	-4
5.2 Monitoring: Core factors	ND	+	ND	-	-	-	-	ND	ND	-	-	ND	-5
Total	-2	+7	-3	-5	-1	+1	-3	-8	-8	-5	-5	-3	

Table 16.15 Omitted Guidelines across Centers

	Centers											
	1	2	3	4	5	6	7	8	9	10	11	12
Number of standardization guidelines NOT mentioned in the center's manual	1	0	1	1	1	0	3	7	2	0	0	2

purpose, probing repetition, and monitoring frequency scores are also strongly negative. This indicates that, regardless of the measure used, survey centers on the whole simply do not uniformly adhere to rigorous forms of standardization. Nevertheless, several summations move closer to the center of the continuum. These include pace, professional objectivity, and probing specificity. Accordingly, Table 16.14 also indicates the variation across centers on these measures.

CONCLUSION

Our exploratory investigation suggests that survey centers are internally variable in their approach to resolving the standardization and autonomy dilemma, and that the overall tendency across centers is toward autonomy rather than standardization. This raises two key questions: Why does there appear to be a tendency toward interviewer autonomy?²⁰ What does this tendency say about the quality of data obtained from different survey centers?

First, why are standardization principles administered the way they are within the survey enterprise? With a rich literature on how to standardize survey interviewing, why are less rigid forms of these techniques used in practice? From talking to center personnel, one reason may be that simplification of procedure eases the training and monitoring of interviewers. A more relaxed approach to survey interviewing may also add to an interviewer's on-the-job comfort, and thus lessen the pervasive problem of job turnover that plagues many survey centers. Survey centers with a high number of omitted guidelines have these concerns.

However, often the standardization components in this analysis are not simply ignored but rather carefully implemented to permit some interviewer autonomy. For example, interviewers are instructed to speak in normal conversational tones instead of a two word-per-second pace. And interviewers are told to use feedback to *establish rapport* rather than to *train* the respondent. Interviewers are allowed to probe as often as they deem necessary. And an interviewer's judgment is often given weight over technique when being monitored by supervisors. In short, survey center staff may be deliberate in how they institute principles of standardization in the direction of more interviewer autonomy. Centers 4, 10, and 11 are all located firmly on the autonomous side of the continuum, yet each of these center's manuals specifically addressed each component analyzed in this chapter.

Consistent with a theme in this volume, accordingly, an understanding of *tacit knowledge* may help explain why survey administrators permit increased autonomy for interviewers. It appears that some survey centers, rather than simply ignoring standardization or letting it slide, recognize the need for interviewers to exercise their commonsense as they interact with respondents. In other words, those survey centers that recognize broader areas of autonomy may recognize the contingencies of the interview process and believe that quality of data is improved when interviewers are taught to handle these contingencies with some degree of flexibility (Groves, 1989:404).²¹

Our speculations about reasons for relatively more autonomy should not imply an endorsement of any policy. Instead, we want to make sociological sense of

patterns that appear in our data. A matter related to the recognition of tacit knowledge is what sociologists have found in regard to reproducibility and replicability in natural science. That is, studies of science and technology, including those of Collins (1985:57-58) and Shapin and Schaffer (1985:281), refer to the *capriciousness* of knowledge flow regarding methodology and have documented how difficult it is for practices and procedures to "travel" from one lab to another. We could hardly expect social science to be any less capricious in the flow of methodological knowledge than natural science. According to Converse's (1987) history of survey research in the United States, rules for standardization have proliferated to countless academic and polling organizations from a relatively few major centers, such as the National Opinion Research Center in Chicago, the Survey Research Center in Ann Arbor, and the Bureau for Applied Social Research in New York. While these or other survey centers may write extensive formulas for standardized interviewing, in the end it takes people to implement the formula, and they have to decide what counts as doing the "same" thing the formula prescribes in a place distanced from where the formula originates (Lynch, 1993). Such decision-making is overwhelmingly a situated activity that happens according to established routines in local cultures and in relation to circumstances of a potentially infinite variety that help condition how any survey center works to collect its data. Indeed, in the survey literature, this matter is discussed in relation to the influence on survey data of possibly distinctive practices involved in training and supervision at different survey organizations, that is, *house effects* (Hill, 1991; Smith, 1978, 1982; Turner, 1982).

Finally, to address our second question regarding quality of data, if researchers argue that greater autonomy may lead to improved data, and our findings indicate that survey administrators are endorsing autonomy, our study also suggests that the self-same survey, with identically scripted CATI questions, could be administered very differently depending on the survey center where it is placed. Interviewers could read questions at dissimilar paces, use disparate forms of feedback after respondents' answers, be more neutral and professional or more warm and friendly in demeanor, engage in distinctive probing practices, and be monitored more or less intensively. These are interactional matters and require investigations that treat them as such. Presently, despite Schaeffer's (1991; this volume, Chapter 4) call for "closer study of how conversational practices are enacted within the prescribed structure of the interview to create data," there is little research that addresses whether these specific interactional differences matter to the quality of data the survey collects.²² Given the importance of reliability in measurement, research is needed to determine whether or not the same instrument at centers with disparate orientations to interviewer standardization and autonomy generate findings that are in substantial agreement. In the past, it has been difficult to discriminate what may be house effects from other influences, including of question order and of historical events occurring between different administrations of the same survey (Smith, 1978; Turner, 1982). Perhaps the time has come for more systematic efforts to determine how uniform measurement is

from survey center to survey center and how consistent the results are from the use of identically written instruments in distinct organizations.

NOTES

1. The five survey centers in this group were selected because of convenience and proximity to Viterna's research base in Bloomington, Indiana. Given these constraints, the resulting mix among centers was quite diverse. The five centers had been in operation from as few as 2 years to as many as 57 years. Most of the surveys conducted by these centers were university affiliated, but many centers also conducted studies for local, state, and national government organizations, not-for-profit organizations, and even several marketing organizations. Resources varied drastically from center to center, from survey organizations with several hundred interviewers in multiple interviewing locations to a single cluster of rooms in a small building and 20 or 30 interviewers on the payroll.

2. We wrote to a dozen survey centers and used the manuals from the seven that responded.

3. For example, of the five centers the first author visited, two made little mention of feedback in their manuals and did not mention appropriate feedback at all throughout their training sessions. The remaining three centers all gave extensive coverage to feedback issues in both their manuals and during actual interviewer training. Likewise, four of the five centers specifically discussed when and how to probe in their manuals and provided lists of acceptable probes. These four centers also discussed probing extensively in their training programs. The fifth center only briefly mentioned probing in the manuals and provided no list of acceptable probes. During training at this center, trainees were told only that they were never to help the respondent understand a question by rephrasing or defining unclear parts. They never received information, verbally or in writing, discussing different situations that may require probing or what probes would be effective for various situations.

4. Fowler and Mangione (1990) is the most recent comprehensive source documenting standardized survey interviewing techniques. [Previous publications about interviewing include Bradburn, Sudman, and Associates (1979) and Cannell, Miller, and Oksenberg (1981).] Furthermore, Fowler and Mangione (1990) is cited in other recent sources on standardized survey interviewing (see, e.g., Lavrakas, 1993) giving it weight as an authoritative source for interviewing technique.

5. The substantive part of the interview contains the instrument's questions for gathering data. Specialized interviewing procedures are required in the substantive portion of the interview to ensure standardization, in contrast to the front end, or introduction, of the interview, where interviewers are characteristically allowed much more autonomy, and where the autonomy may improve interviewers' effectiveness at persuasion. See Houtkoop and van den Bergh (this volume, Chapter 9).

6. We identify survey centers with letters in our quotations and with numbers in our tables. The purpose of different identifiers is to ensure that our centers and their personnel remain anonymous.

7. We recognize that there is a fundamental difference between interviewer autonomy that is explicitly granted in a training program and interviewer autonomy that is implicitly given through the absence of a published procedure. One reflects a thoughtful position on behalf of survey center staff to permit autonomy, while the other may simply be a reflection of more relaxed management policy. Whether such distinctions matter substantively for data collection is not something we are able to address, and we simply collapse the explicit and implicit forms of orientation to pacing as both indicating autonomy.

8. A high absolute number indicates high consistency, with a high positive number indicating consistency in the direction of standardization and a high negative number indicating consistency in the direction of autonomy. A positive or negative total near zero indicates low consistency across centers.

9. Although personnel at several centers mentioned that their lists of "acceptable" feedback phrases may have been derived from practices used in the Center for Survey Research at the University of Michigan, no one was certain about when or where the phrases were initially created.

10. This is a total of seven centers, as three centers recommend "yes," three others recommend "okay," and a seventh center recommends both "yes" and "okay."

11. However, see the discussion of feedback in this volume, Chapter 1.

12. The contradiction between being "serious" (for purposes of standardization) and being "pleasant" (for purposes of rapport) often arises when respondents engage in displays of affect. For example, respondents may make a small joke or introduce laughter that operates to invite reciprocal laughter (Lavin and Maynard, this volume, Chapter 15). Or, a respondent may reveal a sad detail about her life circumstances, as when being asked her marital status, and she replies, "I'm recently widowed." If an interviewer has been instructed to show no emotion or reaction to a respondent's answer, no matter how affect-laden it may be, the strain can be intense.

13. See experimental research on interviewer style in Dijkstra (1987) and Dijkstra and van der Zouwen (1988).

14. In Chapter 1, this volume, see the discussion of predetermined clarifications, verification probes, tuning, and zeroing in as devices for nondirective probing in contexts where respondents have provided approximations that suggest, but do not exactly fit, answer categories.

15. For example, interviewers in Center O have a list of approved probes hanging in their cubicles. The list is organized into four major categories (closed probes, open probes, don't know probes, and refuse probes). The closed probes are further subdivided into probes for nonresponse, probes for two responses, probes for numbers, and probes for qualified answers. The open probes section is subdivided into probes for "anything else?," probes for limiting a response to one answer, and probes asking for more specificity.

16. As mentioned in Chapter 1 (this volume), research does confirm that the meaning of any individual category depends on the entire set of categories, suggesting the importance of re-reading all response options when probing (e.g., Schaeffer and Charmg, 1991; Smit, Dijkstra, and van der Zouwen, 1997).

17. Interviewers are evaluated with as many as 22 different components of interviewing technique or as few as 5. Two of the forms in this collection are organized by the number of the question in the actual interview, allowing supervisors to make comments and evaluate procedure on a question-by-question basis. The other five monitoring forms are organized by topic; supervisors give an overall impression of pace or probing without necessarily evaluating each individual question asked by the interviewer to the respondent. A comparison of monitoring forms also finds great variety in which components of interviewing are included for evaluation. Some monitoring forms ask about the overall flow of the interview, if a pleasant tone was used, if time was wasted between calls, or if the interviewer correctly answered the respondent's questions. Others stick strictly to the basics of question asking: reading verbatim and coding responses correctly.

18. Proper question reading includes reading questions verbatim, neither adding nor omitting words, and asking every question on the survey in its proper order. All centers in this study require all questions to be read verbatim in the order they appear.

19. Of the seven centers for which we have data, only one includes all five factors identified as key in this chapter. Three include four of the five factors, two include three of the five factors, and one only includes two of the five factors. Although we considered making the cutoff point for coding a center as standardized at four out of five, we hesitated to do so because of the lack of consensus on which factors were excluded. For example, two of the three centers with four out of five factors were missing any sort of evaluative framework for proper use of feedback, and the third was missing an analysis of professionalism. Since Fowler and Mangione (1990) give equal weight to each aspect, we decided that only centers that included all key issues would get positive points. If we were to recode the data so that centers with four out of five would receive positive instead of negative points, centers 5, 6, and 11 would all move two spaces further to the right on the continuum in Figure 16.1 (discussed below).

20. We refer to a "tendency" toward autonomy although there actually may be an "increase" in autonomy. Beatty (1995:151) observes that during the 1970s there were "attacks on rapport." Most prominent were definitional problems inherent in the term *rapport* and lack of evidence that rapport (however defined) improved interviewing productiveness. The result was that "research on controlling interviewer variation through standardization... had new opportunities to flourish" (Beatty, 1995:151). Furthermore, interviewing manuals, such as the University of Michigan Survey Research Center's *General Interviewing Techniques*, published in 1983, contained extended attention to standardizing the behavior of interviewers compared to earlier manuals. On the basis of such evidence, Beatty (1995:153, our emphasis) refers to "the increase in survey standardization." We speculate that the 1980s may have

been standardization's heyday, and that interviewing practice has subsequently moved away from the stricter versions of standardization, and this has increased interviewer autonomy. It is because there is no concrete evidence of this increase that we discuss a tendency toward autonomy.

21. Recent studies lend credence to the idea that increased autonomy might improve data quality in some respects. Suchman and Jordan (1990) argue that strict standardization could obtain reliability at the expense of validity, and Schober and Conrad (1997) conclude that allowing interviewers freedom to clarify may improve response accuracy. But Dykema, Lepkowski, and Blixt (1997) find that not reading the question as worded was associated with accuracy in only 1 out of 10 questions they examined. Even then, the association was in the opposite direction from what was predicted (i.e., the association was negative).

22. However, in regard to more general matters such as the "style" of the interviewer, see the discussion in this volume, Chapter 1, and, in particular the studies of Dijkstra (1987), Dijkstra and van der Zouwen (1988), and Schober and Conrad (1997).

REFERENCES

- Beatty, P. 1995. "Understanding the Standardized/Non-Standardized Interviewing Controversy." *Journal of Official Statistics* 11: 147-160.
- Bradburn, N. M., S. Sudman, and Associates. 1979. *Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research*. San Francisco: Jossey-Bass.
- Cannell, C. F., L. Oksenberg, and J. M. Converse. 1977. *Experiments in Interviewing Techniques: Field Experiments in Health Reporting: 1971-1977*. Hyattsville, MD: NCHSR.
- Charmaz, K. 1983. "The Grounded Theory Method: An Explication and Interpretation." Pp. 109-126 in R. M. Emerson (Ed.), *Contemporary Field Research: A Collection of Readings*. Prospect Heights, IL: Waveland.
- Clayman, S. 1988. "Displaying Neutrality in Television News Interviews." *Social Problems* 35: 474-492.
- Collins, H. M. 1985. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press.
- Converse, J. M. 1987. *Survey Research in the United States: Roots and Emergence 1890-1960*. Berkeley, CA: University of California Press.
- Dijkstra, W. 1987. "Interviewing Style and Respondent Behavior: An Experimental Study of the Survey Interview." *Sociological Methods and Research* 16: 309-334.
- Dijkstra, W., and J. van der Zouwen. 1988. "Types of Inadequate Interviewer Behavior in Survey Interviews." Pp. 24-35 in W. E. Saris and I. N. Gallhofer (Eds.), *Data Collection and Scaling*. New York: St. Martins.
- Dykema, J., J. M. Lepkowski, and S. Blixt. 1997. "The Effect of Interviewer and Respondent Behavior on Data Quality: Analysis of Interaction Coding in a Validation Study." Pp. 287-310 in L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo,

- N. Schwarz, and D. Trewin (Eds.), *Survey Measurement and Process Quality*. New York: Wiley-Interscience.
- Fowler, F. J., and T. W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, CA: Sage.
- Garfinkel, H. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Glaser, B. G., and A. L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- Goodwin, T. 1998. "Managing Participation in a Local Work Collectivity: Finding Space, Eavesdropping, and Standardization in the Survey Research Center." Master's thesis, Department of Sociology, Indiana University, Bloomington, IN.
- Goudy, W. J., and H. R. Potter. 1975. "Interview Rapport: Demise of a Concept." *Public Opinion Quarterly* 39: 529-543.
- Groves, R. M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.
- Heil, T. 1999. "The Interview Was Really Good Overall?: Advice-Giving and Advice-Reception in a Supervisory Feedback Session." Master's thesis, Department of Sociology, Indiana University, Bloomington, IN.
- Heritage, J. 1984. *Garfinkel and Ethnomethodology*. New York: Polity.
- Heritage, J., and D. Greatbatch. 1991. "On the Institutional Character of Institutional Talk: The Case of News Interviews." Pp. 93-137 in D. Boden and D. H. Zimmerman (Eds.), *Talk and Social Structure*. Berkeley: University of California Press.
- Hill, D. H. 1991. "Interviewer, Respondent, and Regional Office Effects on Response Variance: A Statistical Decomposition." Pp. 463-486 in P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*. New York: Wiley.
- Hyman, H. 1954. *Interviewing in Social Research*. Chicago: University of Chicago Press.
- Lavrakas, P. J. 1993. *Telephone Survey Methods: Sampling, Selection, and Supervision*. Newbury Park, CA: Sage.
- Lazarsfeld, P. F. 1944. "The Controversy Over Detailed Interviews—An Offer for Negotiation." *Public Opinion Quarterly* 8: 38-60.
- Lynch, M. 1993. *Scientific Practice and Ordinary Action: Ethnomethodology and Social Studies of Science*. Cambridge: Cambridge University Press.
- Marlaire, C. L., and D. W. Maynard. 1990. "Standardized Testing as an Interactional Phenomenon." *Sociology of Education* 63: 83-101.
- Marquis, K. H., C. F. Cannell, and A. Laurent. 1972. "Reporting Health Events in Household Interviews: Effects of Reinforcement, Question Length, and Reinterviews." In *Vital and Health Statistics, Series 2 (45)*. Washington, DC: U.S. Government Printing Office.
- Maynard, D. W., and C. L. Marlaire. 1992. "Good Reasons for Bad Testing Performance: The Interactional Substrate of Educational Exams." *Qualitative Sociology* 15: 177-202.
- Maynard, D. W., and N. C. Schaeffer. 2000. "Toward a Sociology of Social Scientific Knowledge: Survey Research and Ethnomethodology's Asymmetric Alternates." *Social Studies of Science* 30: 264-312.
- Mehan, H. 1979. *Learning Lessons*. Cambridge, MA: Harvard University Press.
- O'Muircheartaigh, C. A. 1997. "Measurement Error in Surveys: A Historical Perspective." Pp. 1-28 in L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz,

- and D. Trewin (Eds.), *Survey Measurement and Process Quality*. New York: Wiley-Interscience.
- O'Muircheartaigh, C. A. 1999. "CASM: Successes, Failures, and Potential." Pp. 39-62 in M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, and R. Tourangeau (Eds.), *Cognition and Survey Research*. New York: Wiley.
- Polanyi, M. 1958. *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago: University of Chicago Press.
- Porter, T. M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Richardson, S. A., B. S. Dohrenwend, and D. Klein. 1965. *Interviewing: Its Forms and Functions*. New York: Basic Books.
- Schaeffer, N. C. 1991. "Conversation with a Purpose—Or Conversation? Interaction in the Standardized Survey." Pp. 367-391 in P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*. New York: Wiley.
- Schaeffer, N. C., and H-W. Charng. 1991. "Two Experiments in Simplifying Response Categories: Intensity Ratings and Behavioral Frequencies." *Sociological Perspectives* 34: 165-182.
- Schober, M. F., and F. G. Conrad. 1997. "Does Conversational Interviewing Reduce Survey Measurement Error?" *Public Opinion Quarterly* 61: 576-602.
- Shapin, S., and S. Schaffer. 1985. *Leviathan and the Air Pump: Hobbes, Boyle, and the Experimental Life*. Princeton: Princeton University Press.
- Smit, J. H., W. Dijkstra, and J. van der Zouwen. 1997. "Suggestive Interviewer Behaviour in Surveys: An Experimental Study." *Journal of Official Statistics* 13: 19-28.
- Smith, T. W. 1978. "In Search of House Effects: A Comparison of Responses to Various Questions by Different Survey Organizations." *Public Opinion Quarterly* 42: 443-463.
- Smith, T. W. 1982. "House Effects and the Reproducibility of Survey Measurements: A Comparison of the 1980 GSS and the 1980 American National Election Study." *Public Opinion Quarterly* 46: 54-68.
- Suchman, L., and B. Jordan. 1990. "Interactional Troubles in Face-to-Face Survey Interviews." *Journal of the American Statistical Association* 85: 232-253.
- Turner, C. F. 1982. "Why Do Surveys Disagree? Some Preliminary Hypotheses and Some Disagreeable Examples." Pp. 159-214 in C. Turner and E. Martin (Eds.), *Surveying Subjective Phenomena*. New York: Russell Sage.
- Williams, D. 1942. "Basic Instructions for Interviewers." *Public Opinion Quarterly* 6: 634-641.