7

Evaluating Survey Questions and Instruments

Designing a good survey instrument involves selecting the questions needed to meet the research objectives, testing them to make sure they can be asked and answered as planned, then putting them into a form to maximize the ease with which respondents and interviewers can do their jobs. This chapter describes steps for designing good survey instruments.

Every survey requires either an interview schedule, which constitutes a script for survey interviewers, or a questionnaire that respondents will read and fill out themselves. These documents, either in paper form or as programs for a computer, will be referred to generically as survey instruments.

Understanding what a good question is and how to use questions as measures, as discussed in Chapter 6, is certainly the foundation of good survey instrument design. There is, however, a series of very practical steps needed to produce a good data collection instrument. This chapter presents a summary of those steps. Converse and Presser (1986), Bradburn and Sudman (1992), Fowler (1995), Groves and colleagues (2004), Presser and colleagues (2004), and Willis (2005) provide longer, more detailed discussions of such steps.

Survey instrument design has two components: deciding what to measure and designing and testing questions that will be good measures. The first step usually is to define the survey objectives, though those objectives may be revised based on subsequent question testing. Then the process of choosing and testing questions takes place. The steps involved in a survey instrument development process may include the following:

- focus group discussions
- drafting a tentative set of questions
- critical review to detect common flaws
- individual cognitive interviews (not replicating proposed data collection procedures)
- putting questions into a survey instrument
- pretesting using an approximation of proposed data collection procedures

DEFINING OBJECTIVES

A prerequisite to designing a good survey instrument is deciding what is to be measured. This may seem simple and self-evident, but it is a step that often is overlooked, to the detriment of the results. One valuable first step is to write a paragraph about what the survey is supposed to accomplish. In designing a survey instrument, researchers often are tempted to add related questions that do not contribute to achieving the project's goals. A check against such temptations is to have a good statement of the purposes, against which the inclusion of a particular area of inquiry can be measured. Second, one should make a list of what should be measured to accomplish the goals of the project. These should not be questions; they should be variables to be measured, listed in categories or areas that make sense.

An analysis plan should be developed to go with the list of variables to be measured. Presumably, a good start already will have been made in connection with the design of the sample. The researcher will have had to think through which subgroups in the population require special estimates. At this point, however, the researcher should refine those ideas, so that there is a clear list of (a) which variables are designed to be dependent variables, for which measures of central tendency (e.g., means or distributions) are to be estimated; (b) which variables are needed as independent variables in order to understand distributions and patterns of association; and (c) which variables may be needed as control or intervening variables to explain patterns observed and to check out competing hypotheses.

These three documents, a statement of purposes, a list of the kinds of variables to be measured, and a draft of an analysis plan, are essential components to developing a survey instrument.

PRELIMINARY QUESTION DESIGN STEPS

Focus Groups

Before writing a draft of a structured set of questions, it almost always is valuable to conduct focused discussions with people who are in the study population about the issues to be studied. The primary purpose of these discussions is to compare the reality about which respondents will be answering questions with the abstract concepts embedded in the study objectives.

Example. The goal is to measure the number of visits to doctors. A group discussion could be focused on what counts as a visit to a doctor. Two key

concepts are "visit" and "doctor." Participants could be asked about the various contacts they had related to doctors (e.g., telephone consultations, trips to have X-rays or laboratory tests, inoculations) and whether or not they considered these contacts to be visits. They also could be asked about the various people they contacted related to their health (e.g., psychologists, psychiatrists, physician assistants, ophthalmologists, optometrists, physical therapists) and asked about whether or not they considered these individuals to be doctors.

This discussion alone could provide critical information of at least three types:

1. The kinds of contacts people have that possibly could be considered visits. This information would help the researcher refine the objectives and refine question wording to make it clear what is and is not to be included. For example, do we want to include telephone consultations? If a nurse practitioner is seen in a doctor's office, does that count?

2. What people know. For example, is everyone clear that a psychiatrist is an M.D., but a psychologist is not? What assumptions can be made about people's knowledge and perceptions of the background, training, or credentials of health care providers?

3. Comprehension of some key words or terms. Does the word *doctor* mean an M.D., or is it more generic (like Kleenex), referring to professionals in white coats delivering health-related services? Do alternative terms, such as *health care provider* or *health care professional,* have consistent meaning for respondents?

Focus group discussions are best with six to eight people. The general protocol is to discuss people's perceptions, experiences, and perhaps feelings related to what is to be measured in the survey. The number of groups that are valuable will vary, but virtually every survey instrument will benefit from at least a couple of focus group discussions at an early stage in the survey instrument development process.

Drafting Questions

Armed with a list of what is to be measured, the researcher attempts to find the single question or set of questions needed to create measures of the variables on the list. Many questions, such as those dealing with background or demographic issues, are standard to many surveys. Reviewing the questions in the General Social Survey carried out by the National Opinion Research Center at the University of Chicago may be useful. Many surveys are also available online through the Inter-university Consortium of Political and

Social Research (ICPSR) at the University of Michigan. McDowell (2003) is a valuable resource for those doing health-related surveys. Copies of original survey instruments from any of the major survey organizations also are useful as references. From these, the researcher can glean ideas about how specific questions are phrased, how to generate standardized questions, and how to format survey instruments.

Taking advantage of the work that others have done is very sensible. Of course, it is best to review questions asked by researchers who have done previous work on the study topic. In addition, if questions have been asked of other samples, collecting comparable data may add to the generalizability of the research. The mere fact that someone else has used a question before, however, is no guarantee that it is a very good question or, certainly, that it is an appropriate question for a given survey. Many bad questions are asked over and over again because researchers use them uncritically. All questions should be tested to make sure that they "work" for the populations, context, and goals of a particular study.

PRESURVEY EVALUATION

Critical Systematic Review

Once a set of questions is drafted, a good next step is to subject them to a critical systematic review. Lessler and Forsyth (1996) produced a list of issues to look for in a set of questions. Fowler and Cosenza (2008) also proposed a list of standards for questions that can be applied prior to testing. While neither list is exhaustive, both lists identify a set of question characteristics that are indicative of problem questions. An extension of this has been developed by Graesser and colleagues (Graesser, Cai, Louwerse, & Daniel, 2006), who have a list of 12 question problems that can be identified using an online evaluation tool. Using one of these lists can help to identify questions that need revision; it also can flag issues for attention during the next phases of testing.

Cognitive Laboratory Interviews

Once a set of questions has been drafted, critically reviewed, and revised as warranted, the next step is to find out if they are questions people consistently can understand and can answer. Focus group discussions should provide some insights into comprehension issues, but they do not provide a forum for evaluating specific wording or the difficulty of the response task. At early stages of framing questions, the researcher also can learn a great deal by trying out questions on friends, relatives, and coworkers. Early versions of

most survey instruments contain questions that are confusing, that cannot be read as written, and that are virtually unanswerable by anyone.

Once questions are in draft form, but before subjecting them to a formal field pretesting, a more formal kind of testing, commonly called *cognitive testing,* is a valuable next step (DeMaio & Rothgeb, 1996; Forsyth & Lessler, 1992; Fowler, 1995; Lessler & Tourangeau, 1989; Presser et al., 2004; Willis, 2005; Willis, DeMaio, & Harris-Kojetin, 1999). Although cognitive interviews take a variety of forms, there are certain features that they usually share. First, respondents are volunteers who have a willingness to expend more time than the data collection itself actually involves in order to help the researchers understand how the questions work. Often respondents are paid and are brought into a laboratory setting, where the interviews can be observed or videotaped.

These interviews usually are not done by regular interviewers. In some cases, interviewers are cognitive psychologists; in other cases, interviews are done by the investigators themselves or senior interviewer supervisors. In most cases, interviewers are thoroughly knowledgeable about the objectives of each question, so that they can detect issues that arise in the way that respondents understand questions or form answers to questions.

A typical protocol calls for asking respondents a set of proposed questions, then in some way gathering information about how the respondents understood the questions and about the way in which they answered them. Sometimes respondents are asked to "think aloud" while they are preparing their answers. In other cases, respondents are asked a set of follow-up questions about the way they understood each question and about issues related to their answers. Two of the most common tasks are

1. to ask respondents to say in their own words what they think the question is asking
2. to ask respondents to explain how they chose a particular answer over others

The point is to get enough information about the respondents' comprehension and preparation of responses to evaluate whether they performed the task in the way the researcher wanted. There are four specific kinds of questions that most cognitive testing is designed to answer:

1. Are questions consistently understood?
2. Do respondents have the information needed to answer the questions?
3. Do the answers accurately describe what respondents have to say?
4. Do the answers provide valid measures of what the question is designed to measure?

There are limits to what can be learned from laboratory interviews. Usually few such interviews are done (often fewer than 10), because they are labor

intensive and, in most organizations, can only be conducted by a small number of people. Second, the interviews are conducted under artificial conditions; tasks that volunteers are able and willing to perform may not be handled by a cross-section sample interviewed in their homes. Nonetheless, such interviews are increasingly seen as an essential step in the design and evaluation of a survey instrument. Questions that are not consistently understood or answered in a laboratory setting certainly will not work any better in an actual survey. Problems of comprehension and difficulties with the response task are not identified as reliably in field pretests as they are in laboratory interviews, where the answering process can be examined.

The cognitive laboratory interview has most often been used to test interview protocols. The same issues of comprehension and difficulty of the response task, however, apply to self-administered forms. Although standard tests of self-administered forms, as described below, often involve debriefing questions similar to those used in cognitive interviews, respondent comprehension is more apparent when the question-and-answer process is carried out orally. Thus, to test questions designed to be self-administered, an oral cognitive interview may be an effective way to identify problems that will not be picked up in the standard pretest.

DESIGN, FORMAT, AND LAYOUT OF SURVEY INSTRUMENTS

Once a set of questions is close to ready for final pretesting, the questions need to be put into a form to facilitate interviewer or self-administration. A first step is simply to order the questions. Many researchers like to start with relatively easy, straightforward questions that help get the respondent into the survey. Questions requiring a good deal of thought, or those believed to be sensitive, often are reserved for the middle or later sections of survey instruments. A good practical step is to number questions in sections: A1, A2, B1, B2, and so forth. In this way, when questions are added or deleted, it is not necessary to renumber every question.

Whether the survey is to be interviewer administered or self-administered, the goal of the layout and format of the questionnaire should be to make the tasks of the interviewer and the respondent as easy as possible. For an interviewer-administered survey instrument, the following are some rules that will help achieve that goal:

1. Adopt a convention that differentiates between the words that interviewers are to read to respondents and words that are instructions. A common

convention is to use uppercase letters for instructions and lowercase for questions to be read aloud.

2. If an interview uses a paper-and-pencil form, and is not computer assisted, establish a clear convention for handling instructions to skip questions that do not apply to a particular respondent. A common convention is to put INSTRUCTIONS IN CAPITAL LETTERS. The instructions should be keyed to a particular response and tell the interviewer where to go to ask the next questions. Of course, computer-assisted instruments will make skips automatically, based on the answers that are entered.

3. Put optional wording in parentheses. Conventions such as (his/her) or (husband/wife) are easy for interviewers to handle smoothly if they are alerted by the parentheses. A similar convention uses all caps (e.g., SPOUSE) when the interviewer must supply a word that is not provided in the question itself. Computer assistance often enables optional wording to be filled in, rather than have the interviewer adapt the wording to the situation.

4. Check to make sure that all the words that an interviewer has to say are, in fact, written. This includes not only the phrasing of the questions but transitions, introductions to questions, needed definitions, and explanations.

For self-administered questionnaires, the same kind of general principles apply; that is, the main goal is to make the questionnaire easy to use. If anything, the formatting of a self-administered questionnaire is more important. In contrast to interviewers, respondents do not receive the benefit of training, they usually are not motivated to do the job well, and they are not selected on the basis of their ability to handle questionnaires. Five guiding principles are as follows:

1. A self-administered questionnaire mainly should be self-explanatory. Reading instructions should not be necessary, because they will not be read consistently.

2. Self-administered questionnaires mainly should be restricted to closed answers. Checking a box, clicking on a response, or circling a number should be the only tasks required. When respondents are asked to answer in their own words, the answers usually are incomplete, vague, and difficult to code, and therefore they are of only limited value as measurements.

3. The question forms in a self-administered questionnaire should be few in number. The more the instrument can be set up so that the respondent has the same kinds of tasks and questions to answer, the less likely it is that respondents will become confused; also, the easier the task will be for the respondents.

4. A questionnaire should be laid out in a way that seems clear and uncluttered. Photo reduction (or other strategies for putting many questions on a

page) actually reduces the response rate compared with when the same number of questions are spaced attractively over more pages.

5. Provide redundant information to respondents by having written and visual cues that convey the same message about how to proceed. If people possibly can be confused about what they are supposed to do, they will be. Work on making everything simple and clear.

Most of the principles outlined above also apply to computer-assisted instruments. In addition, helping respondents and interviewers handle common navigational challenges such as how to correct previous answers or what to do if respondents want to skip a question is an important part of the design work. The key to addressing all these issues is to find the problems through testing. Good descriptions of testing procedure are found in Dillman and Redline (2004), Tarnai and Moore (2004), Hansen and Couper (2004), and Baker, Crawford, and Swinehart (2004).

FIELD PRETESTS

Once a survey instrument has been designed that a researcher thinks is nearly ready to be used, a field pretest of the instrument and procedures should be done. The purpose of such pretests is to find out how the data collection protocols and the survey instruments work under realistic conditions.

Pretesting an Interview Schedule

The traditional pretest done by conscientious survey organizations usually consists of experienced interviewers taking 20 to 50 interviews with respondents drawn from a population the same as, or similar to, the population to be included in the survey. Interviewers are asked to play two roles in such pretests: They are interviewers, carrying out the procedures, and they are observers of the data collection process who are asked to report back to the researchers about any ways in which the procedures and survey instruments could be improved. It probably is most typical for this feedback process to take place in a group debriefing session, though on occasions interviewers report back individually.

Pretests such as those described above are an essential part of the survey design process. A particularly important function is to test the usability of the instrument, both the questions and the layout, from the interviewers' perspective. However, such tests also have several limitations. The standards that

interviewers use for what constitutes a problem often are not specified well, and it is almost certain that interviewers are inconsistent in what they consider to be a problem. In addition, a group discussion is an imperfect way to gather systematic information about the pretest experience.

Researchers have added steps designed to make the pretest experience more systematic and more valuable. One simple innovation is to ask interviewers to fill out a brief rating form on each question in addition to reporting back in a group session. One such rating form asks interviewers to evaluate each question with respect to whether or not (a) it is easy to read as worded, (b) respondents understand the question in a consistent way, and (c) respondents can answer the question accurately (Fowler, 1995). Obviously, interviewers have to guess about whether or not respondents are understanding questions and answering accurately; however, they do this in any case. The advantage of a form is that interviewers are asked systematically to attend to these aspects of question design as well as the other, more practical aspects of the survey instrument to which they ordinarily attend. Also, having interviewers do these ratings makes it easier for investigators to summarize interviewer reports and identify question problems in a more consistent way.

A more important, and probably more useful, innovation with respect to the field pretest is the use of audio-recording and behavior coding to evaluate survey questions. With respondent permission, which is almost always granted, it is easy to record pretest interviews done either in person or over the telephone. Trained coders can then listen to those recordings and evaluate problems in the question-and-answer process in a consistent way.

Three behaviors have been shown to be most important in identifying problems with survey questions (Fowler & Cannell, 1996; Oksenberg et al., 1991): (a) whether or not the interviewer reads the question as worded, (b) whether or not the respondent asks for clarification, and (c) whether or not the respondent initially gives an inadequate answer that requires interviewer probing. It has been found that questions consistently produce or do not produce these kinds of behaviors in interviews; that is, there are questions that interviewers consistently misread, that lead respondents to ask for clarification, or that respondents consistently answer in an inadequate way. Such coding does not identify all questions that are not consistently understood by respondents. However, when one of these behaviors occurs in 15% or more of pretest interviews, it has been shown that a question is either highly likely to produce distorted data or distinctively susceptible to interviewer effects (Fowler, 1991; Fowler & Mangione, 1990).

An additional benefit of behavior coding of pretest interviews is that the results are systematic and can be replicated. Thus the question evaluation process is moved beyond the subjective opinions of researchers and

interviewers, and concrete, replicable evidence is produced about questions that are inadequate.

Trace files are a third source of information from a pretest of a computer-assisted interview. When an interview is computer-assisted, it is possible to retrieve the actual key strokes interviewers make. Those files can identify places where interviewers have to go back to previous screens and questions. Having to return to previous questions slows down an interviewer and often is a sign that question flow is not well designed. Looking at how "help" functions are used can provide clues to where help is needed and how "useful" various help functions are. Again, a plus of examining trace files is that the results are systematic and quantifiable (Couper, Hansen, & Sadowsky, 1997; Hansen & Couper, 2004).

Pretesting a Self-Administered Questionnaire

If anything, self-administered instruments require more pretesting than interviewer-administered survey instruments, simply because interviewers can solve some problems that the researchers did not solve in the design of the survey instrument. Unfortunately, pretesting a self-administered instrument is also somewhat harder, because problems of comprehension and difficulties with answering questions are less evident. Although people have used observation of how people fill out forms or interact with a computer as a way of trying to identify unclear or confusing questions and instructions, it is not as satisfactory as the audio recording and behavior coding of interviews to identify question problems.

Probably the best way to pretest a self-administered questionnaire is in person, with a group of potential respondents. If it is a computer-based survey, respondents can respond via individual laptops. First, respondents should fill out the questionnaire as they would if they were part of a survey. Then the researcher can lead a discussion about the instrument. One topic obviously is whether the instructions were clear. A second is whether or not the questions were clear. A third is whether there were any problems in understanding what kind of answers were expected or in providing answers to the questions as posed (Dillman & Redline, 2004).

In addition to group tests, the usability of a computer-based instrument often benefits from some one-on-one testing, in which some respondents are observed interacting with the computer and the questions (Tamai & Moore, 2004). Direct observation or videotaping can be use to identify trouble spots. Again, trace files can also be examined to identify navigational problems or places where respondents went back to correct answers.

Debugging a Computer-Assisted Instrument

Having interviewers or respondents test instruments provides information about ease of use, but it does not provide information about whether or not the data collection protocol is correct. The key area for concern is the "skip" instructions.

A great strength of computer assistance is to help respondents and interviews correctly navigate contingencies: when which questions are asked, or how they are worded, is tied to the answers given to preceding questions. Of course, the accuracy of the "skip" instructions requires careful proofreading of the final versions of paper instruments. However, the challenges of checking the accuracy of computer-assisted instruments are much greater than for paper instruments. The problem is that testers cannot see which questions are skipped and, hence, they may miss the fact that a question is skipped that should have been asked. Proofreading a printout of the program and extensive testing are valuable steps. However, if an instrument is long and contains complex contingencies, those steps may be inadequate.

For this reason, once a survey begins, it should be standard practice to tabulate the distributions of answers to the early returns. It is only by checking such output that a researcher can be sure that the contingency instructions are working as intended.

SURVEY INSTRUMENT LENGTH

One outcome of a good pretest is to find out how long it takes to complete a survey instrument. The criteria for interview length should include cost, effect on response rate, and the limits of respondent ability and willingness to answer questions. The extent to which the length of a self-administered questionnaire affects costs and response rates varies with the population being studied and the topic; generalizations are difficult. It also is hard to generalize about how long people can be interviewed.

When researchers find they have more questions to ask than they feel they can ask, there are two choices available. Of course, the researcher simply may cut questions. An alternative approach is to assign subsets of questions to representative subsamples of respondents. Such an approach increases the complexity of the survey and reduces the precision of estimates of those variables, but this may be preferable to leaving out questions altogether. A clear advantage of computer-assisted data collection is the ease with which such designs can be implemented.

CONCLUSION

There was a time when one might have thought that evaluation of questions was largely a subjective process, contingent on the taste and preference of interviewers and researchers. We now know we can move beyond that. Survey questions should mean the same thing to all respondents, answering the questions should be a task that most or almost all respondents can perform, and the words in an interview schedule should be an adequate script that interviewers can follow as worded in order to conduct an interview.

Obviously, no matter how clear the question, some respondents will have difficulty with it, and some interviewers will misread it. There are judgments to be made about how bad a question can be before it must be changed. A critical part of the design and evaluation process of survey instruments, however, is to gather information about comprehension, the task of answering questions, and how interviewers and respondents use the protocols in order that judgments can be made about whether or not the questions and instruments need to be changed. Good question and instrument evaluation prior to actually doing a survey is a critical part of good survey practice. It is one of the least expensive ways to reduce error in survey estimates. Although there is work to be done to define the most efficient and effective ways of evaluating questions, the procedures outlined on the preceding pages constitute a useful array of techniques that, when used, will have a major positive impact on the quality of survey data.

EXERCISES

Take the questions generated in the exercise for Chapter 6 and transform them into a set of questions that an interviewer could administer in a standardized way.

1. Cognitively test the questions and revise as needed.
2. Pretest the resulting questions. Revise as needed.
3. Now put the same questions in a form for self-administration. Pretest that.

Further Readings

Presser, S., et al. (2004). *Methods for testing and evaluating survey questionnaires.* Hoboken, NJ: John Wiley.

Willis, G. (2005). *Cognitive interviewing.* Thousand Oaks, CA: Sage.

8

Survey Interviewing

Interviewers affect survey estimates in three ways: They play a major role in the response rate that is achieved, they are responsible for training and motivating respondents, and they must handle their part of the interview interaction and question-and-answer process in a standardized, nonbiasing way. This chapter discusses the significance of interviewer selection, training, and supervision, plus the procedures interviewers are given, for minimizing interviewer-related error in surveys.

OVERVIEW OF INTERVIEWER JOB

Although many surveys are carried out using self-administered methods, using interviewers to ask questions and record answers is certainly a common part of survey measurement procedures, both face-to-face and over the telephone. Because of the central role they play in data collection, interviewers have a great deal of potential for influencing the quality of the data they collect. The management of interviewers is a difficult task, particularly in personal interviewer studies. The goal of this chapter is to provide an understanding of what an interviewer is supposed to do, appropriate procedures for managing interviewers, and the significance of interviewer management and performance for the quality of survey-based estimates.

Interviewers have three primary roles to play in the collection of survey data:

- to locate and enlist the cooperation of selected respondents
- to train and motivate respondents to do a good job of being a respondent
- to ask questions, record answers, and probe incomplete answers to ensure that answers meet the question objectives

Gaining Cooperation

Interviewers have to get in touch with respondents in order to enlist cooperation. The difficulty of this part of the job differs greatly with the sample.

Interviewers have to be available when respondents want to be interviewed, they have to be available (and persistent) enough to make contact with hard-to-reach respondents, and for in-person interviews in people's homes, they have to be able and willing to go where the respondents are.

Although many sampled individuals agree readily to be interviewed, enlisting the cooperation of uninformed or initially reluctant respondents is undoubtedly one of the hardest and one of the most important tasks interviewers must perform. More interviewers probably fail in this area than any other.

There is no doubt that some interviewers are much better than others at enlisting cooperation. It also is clear that different personal styles will work. Some effective interviewers are very businesslike, whereas others are more casual and personable. Experience suggests that there are two characteristics that interviewers who are good at enlisting cooperation seem to share. First, they have a kind of confident assertiveness. They present the study as if there is no question that the respondent will want to cooperate. The tone and content of their conversation does not hint at doubt that an interview will result. Second, they have a knack of instantly engaging people personally, so that the interaction is focused on and tailored very individually to the respondent. It may be very task oriented, but it is responsive to the individual's needs, concerns, and situation. Reading a predesigned script is not an effective way to enlist cooperation.

Although these interviewer skills are important for all surveys, they are challenged particularly by telephone surveys for which respondents receive no advance notice (as in the case when RDD is used) or when the subject matter does not readily engage respondent interest.

Training and Motivating Respondents

Respondents' performance, such as the accuracy of reporting, has been linked to their orientation to the interview. Interviewers have been shown to play an important role in setting respondent goals (Cannell & Fowler, 1964; Cannell, Oksenberg, & Converse, 1977; Fowler & Mangione, 1990). For example, interviewers who rush through interviews encourage respondents to answer questions quickly. Interviewers who read questions slowly indicate to respondents, in a nonverbal way, their willingness to take the time to obtain thoughtful, accurate answers; consequently, they do obtain more accurate answers. Studies also show that the way interviewers provide encouragement to respondents affects their sense of what they are supposed to do and how well they report (Cannell et al., 1987; Cannell, Oksenberg, & Converse, 1977; Fowler & Mangione, 1990; Marquis, Cannell, & Laurent, 1972).

There is no doubt that most respondents have little idea of what they are expected to do and how they are to perform their roles. Interviewers both explicitly and implicitly teach respondents how to behave; this is an often unappreciated but critical part of the interviewer's job.

Being a Standardized Interviewer

Survey researchers would like to assume that differences in answers can be attributed to differences in what respondents have to say (i.e., their views and their experiences) rather than to differences in the stimulus to which they were exposed (i.e., the question wording, the context in which it was asked, and the way it was asked). The majority of interviewer training is aimed at teaching trainees to be standardized interviewers who do not affect the answers they obtain. There are five aspects of interviewer behavior that researchers attempt to standardize: the way interviewers present the study and the task; the way questions are asked, the way inadequate answers (i.e., answers that do not meet question objectives) are probed, the way answers are recorded, and the way the interpersonal aspects of the interview are handled. Each of these is discussed below in greater detail.

1. *Presenting the study.* Respondents should have a common understanding of the purposes of the study, because this sense of purpose may have a bearing on the way they answer questions. Assumptions about such things as confidentiality, the voluntary nature of a project, and who will use the results also potentially can have some effect on answers. A good interviewing staff will give all respondents a similar orientation to the project so that the context of the interview is as constant as possible.

2. *Asking the questions.* Survey questions are supposed to be asked exactly the way they are written, with no variation or wording changes. Even small changes in the way questions are worded have been shown, in some instances, to have significant effects on the way questions are answered.

3. *Probing.* If a respondent does not answer a question fully, the interviewer must ask some kind of follow-up question to elicit a better answer; this is called *probing.* Interviewers are supposed to probe incomplete answers in nondirective ways—ways that do not increase the likelihood of any one answer over another. For fixed response questions, repeating the question and all the response alternatives is the most commonly needed probe. For open-ended questions, repeating the question, or asking "Anything else?"; "Tell me more?"; or "How do you mean that?" will handle most situations if the survey instrument is designed well.

4. *Recording the answers.* The recording of answers should be standardized so that no interviewer-induced variation occurs at that stage. When an open-ended question is asked, interviewers are expected to record answers verbatim; that is, exactly in the words that the respondent uses, without paraphrasing, summarizing, or leaving anything out. In fixed-response questions, when respondents are given a choice of answers, interviewers are required only to record an answer when the respondent actually chooses one. There is potential for inconsistency if interviewers code respondent words into categories that the respondent did not choose.

5. *Interpersonal relations.* The interpersonal aspects of an interview are to be managed in a standardized way. Inevitably, an interviewer brings some obvious demographic characteristics into an interview, such as gender, age, and education. By emphasizing the professional aspects of the interaction and focusing on the task, however, the personal side of the relationship can be minimized. Interviewers generally are instructed not to tell stories about themselves or to express views or opinions related to the subject matter of the interview. Interviewers are not to communicate any judgments on answers that respondents give. In short, behaviors that communicate the personal, idiosyncratic characteristics of the interviewer are to be avoided because they will vary across interviewers. To behave as a professional, not a friend, helps to standardize the relationship across interviewers and respondents. There is no evidence that having a friendly interpersonal style *per se* improves the accuracy of reporting; it probably tends to have a negative effect on accuracy (Fowler & Mangione, 1990).

A special complexity is introduced when the interviewer and respondent come from different backgrounds in society. In this instance, communication may not be as free and easy as when backgrounds are similar. There is some evidence that interviewers who take steps to ease communication in such situations (e.g., by introducing a bit of humor) may be able to produce a more effective interview (Fowler & Mangione, 1990). Efforts to relax the respondent, however, should not detract from a basically professional interaction, focused on good task performance.

Significance of Interviewer's Job

It should be clear from the above that interviewing is a difficult job. Moreover, failure to perform the job may produce three different kinds of error in survey data:

- Samples lose credibility and are potentially biased if interviewers do not do a good job of enlisting respondent cooperation.

- The precision of survey estimates will be reduced, there will be more error around estimates, to the extent that interviewers are inconsistent in ways that influence the data.
- Answers may be systematically inaccurate or biased to the extent that interviewers fail to train and motivate respondents appropriately or fail to establish an appropriate interpersonal setting for reporting what is required.

Given all this potential to produce error, researchers should be motivated to use good interviewers. There are several avenues for affecting the quality of an interviewer's work: recruitment and selection, training, supervision, designing good questions, and using effective procedures. The next five sections will discuss the potential of each of these to influence interviewer performance.

INTERVIEWER RECRUITMENT AND SELECTION

Some of the characteristics of interviewers are dictated by requirements of the survey interviewer's job that have nothing to do with the quality of data per se:

1. Interviewers must have reasonably good reading and writing skills. Many, if not most, interviewers now work with computers, so that typing skills and general familiarity with computers are usually needed, too. Most survey research organizations require high school graduation, and many require or prefer interviewers to have at least some college experience.

2. Interviewing is primarily part-time work. It is difficult to work 40 hours a week every week on general population surveys; survey organizations almost always have some ebbs and flows of work for interviewers. As a result, potential interviewers usually are people who can tolerate intermittent income or are between more permanent jobs. Interviewer pay is usually not high for a college-educated person. Often, there are no benefits, such as health insurance, to the interviewer job. It is unusual for a survey interviewer to be able to rely on interviewing as a sole source of income and support over a long period of time.

3. Personal household interviewers must have some flexibility of hours; surveys require interviewers to be available when respondents are available. One advantage of telephone interviewing is that individual interviewers can work more predictable shifts, although evening and weekend work is prime time for almost all general-population survey work.

4. Personal household interviewers must be mobile, which often excludes people with some physical disabilities and those without the use of a car. Neither of these restrictions is salient to telephone interviewers.

Beyond these practical job requirements, there is little research basis for preferring one set of interviewer candidates over others. For example, experienced interviewers are likely to be better at enlisting cooperation simply because those for whom it is a problem will not continue to work as interviewers; however, there is no documented positive effect of experience on data quality. There is some evidence that interviewers become careless and collect poorer data over time (Bradburn, Sudman, & Associates, 1979; Cannell, Marquis, & Laurent, 1977; Chromy, Eyerman, Odom, McNeeley, & Hughes, 2005; Fowler & Mangione, 1990; Groves et al., 2004).

Likewise, having interviewers who have specialized knowledge about the subject matter is seldom a plus. In fact, because knowledgeable interviewers may assume they know what the respondent is saying when the respondent has not been clear, they may read more into what the individual is saying than people not trained in the area. Unless interviewer observations or ratings requiring an extensive specialized background are needed, a trained interviewer with no special background usually is the best choice.

Age, education, and gender of interviewer seldom have been associated with data quality, though there is some evidence that females may, on average, be more positively rated by cross-section samples (Fowler & Mangione, 1990; Groves, 1989). In general, a researcher would be best advised to send the best interviewer available to interview a respondent, regardless of demographic characteristics. The exception is if the subject matter of the survey directly bears on race or religion (or any demographic characteristic) and the feelings of the respondents about people in the same or different groups. For example, if people are to be interviewed about their own anti-Semitic feelings, the Jewishness of the interviewer will make a difference in the answers (Robinson & Rhode, 1946). In the same way, blacks and whites express different feelings about race depending on the interviewer's skin color (Schuman & Converse, 1971).

It is important to note, however, that matching on ethnicity does not necessarily improve reporting. Two studies of this issue found that black respondents reported income from welfare (Weiss, 1968) and voting (Anderson, Silver, & Abramson, 1988) more accurately to white interviewers than to black interviewers.

There is no question that a researcher should consider the interaction between the subject matter of a survey and the demographic characteristics of the interviewers and respondents. If ethnicity (or some other characteristic) is extremely salient to the answers to be given, controlling the relationship of interviewer and respondent characteristics should be considered so that the effect of the interviewer on the data can be measured (Groves, 1989). For most surveys, however, the practical difficulties and costs of controlling

interviewer assignments and the lack of predictable effects will argue against trying to control the demographic characteristics of respondents and interviewers.

Finally, volunteer interviewing staffs are almost always unsuccessful at carrying out probability sample surveys. There are several reasons for the failure of volunteers. Because it is hard to require attendance at lengthy training sessions, volunteers usually are trained poorly. Because it is hard to terminate poor volunteer interviewers, response rates are usually low. Moreover, volunteer attrition is usually high.

The above discussion offers few guidelines for researchers in the selection of interviewers. In some rather specialized circumstances, the interviewer's ethnic background, age, or gender may affect answers; for example, teenagers may respond differently to older female interviewers (Erlich & Riesman, 1961). For most surveys, however, the particular job requirements largely will dictate the pool of interviewers. There is little basis for ruling out people because of their background or personality characteristics. Rather, the key to building a good interviewing staff is good training and careful supervision. In addition, because of the difficulty of identifying good interviewers in advance, attrition of less able interviewers is probably a critical and necessary part of building a good staff of interviewers.

TRAINING INTERVIEWERS

There is great diversity in the kinds of training experiences to which survey interviewers are exposed. The exact amount of time that will be devoted to training, the kind of training session, and the content of the program obviously will depend on the particular organizational setting and what interviewers are going to be doing. There is some disagreement, in addition, on the extent to which effort should be devoted to an initial training session, prior to the onset of field experience, versus continuous learning and retraining after interviewers have begun. Nonetheless, all professional survey organizations concerned about data quality have at least some kind of (usually face-to-face) training of all new interviewers. The following is a general summary of what reasonable interviewer training might entail.

Content of Training

The content of training includes both general information about interviewing that applies to all surveys and information specific to the particular study

on which interviewers are to work. The general topics to be covered will include the following:

- procedures for contacting respondents and introducing the study
- the conventions that are used in the design of the survey instrument with respect to wording and skip instructions, so that interviewers can ask the questions in a consistent and standardized way
- procedures for probing inadequate answers in a nondirective way
- procedures for recording answers to open-ended and closed questions
- rules and guidelines for handling the interpersonal aspects of the interview in a nonbiasing way
- how to use the computer-assisted interviewing programs

In addition, many research organizations feel that it is a good idea to give interviewers a sense of the way that interviewing fits into the total research process. For that reason, they often attempt to give interviewers some familiarity with sampling procedures, coding, and the kinds of analyses and reports that result from surveys. Such information may be helpful to interviewers in answering respondent questions and may play a positive role in motivating the interviewer and helping him or her to understand the job.

With respect to any specific project, interviewers also need to know the following:

- Specific purposes of the project, including the sponsorship, the general research goals, and anticipated uses of the research. This information is basic to providing respondents with appropriate answers to questions and helping to enlist cooperation.

- The specific approach that was used for sampling, again to provide a basis for answering respondent questions. In addition, there may be some training required in how to implement the basic sample design.

- Details regarding the purposes of specific questions—not necessarily their roles in analyses, but at least the kind of information they are designed to elicit.

- The specific steps that will be taken with respect to confidentiality, and the kinds of assurances that are appropriate to give to respondents.

Procedures for Training

There are six basic ways to teach interviewers: written materials, lectures and presentations, computer-based tutorials, planned exercises, practice

role-playing, and observation of early interviews. Written materials are usually of two types. First, it is a very good idea to have a general interviewer manual that provides a complete written description of interviewing procedures. In addition, for each particular study, there normally should be some project-specific instructions in writing. It is tempting when interviewers are being trained in person and a project is being done in a local site to skimp on the preparation of written materials. Newly trained interviewers, however, say that there is an overwhelming amount of material and information to absorb during training. Having the procedures in writing enables interviewers to review material at a more leisurely pace; it also increases the odds that messages are stated clearly and accurately.

Lectures and demonstrations obviously have a role to play in any interviewer training, whether only a single interviewer or a large group of interviewers is being trained. In addition to the general presentation of required procedures and skills, most trainers find that demonstrating a standardized interview is a quick and efficient way to give interviewers a sense of how to administer an interview. Videotapes are often used to supplement lectures. Making videotapes of practice interviews or other interviewer activities is a good tool for interviewer training.

The widespread use of computer-assisted interviewing means that interviewer training must include teaching interviewers to use computer-based instruments. The most widely used survey systems have computer-based tutorials that can be integrated into general interviewer training.

Because these are new skills, supervised structured practice is one of the most important parts of interviewer training. Having interviewers take turns playing the respondent and interviewer roles is common practice. Practice should include enlisting cooperation and handling the question-and-answer process. There also is great value in monitoring some practice interviews with respondents who are not role playing and whom interviewers do not know. For personal interviews, supervisors can accompany and observe new interviewers doing practice interviews or review tape-recorded interviews. On the telephone, interviews may be monitored directly or tape-recorded for later review.

Two studies (Billiet & Loosveldt, 1988; Fowler & Mangione, 1990) concluded that interviewer training of less than 1 day produces unsatisfactory interviewers; they are not able to perform their jobs as instructed, and the resulting data are affected adversely. Training programs lasting from 2 to 5 days are the norm in professional survey organizations. The length of training depends on numerous factors, including the number of interviewers to be trained and the complexity of the project for which they are being trained. The critical key to the quality of training, however, is probably the amount of supervised practice interviewing.

SUPERVISION

The keys to good supervision are to have the information needed to evaluate interviewer performance and to invest the time and resources required to evaluate the information and provide timely feedback. There are four main aspects of interviewer performance to supervise: costs, rate of response, quality of completed questionnaires, and quality of interviewing. It is considerably easier to supervise interviewers who are doing telephone interviewing from a centralized facility than those interviewing in households.

Costs

Supervising costs for interviewers requires timely information about time spent, productivity (usually interviews completed), and mileage charges for interviewers using cars. High-cost telephone interviewers are likely to be those who work at less productive times, who have high refusal rates (a refusal takes almost as much time as an interview), or who simply find ways (e.g., editing interviews, sharpening pencils) to make fewer calls per hour. High-cost personal household interviewers are likely to live far from their sample addresses, to make trips that are too short or at the wrong times (evenings and weekends are clearly the most productive), or to have poor response rates.

Response Rates

It is critical to monitor response rates (particularly rates of refusals) by interviewers on a timely basis; however, this is not easy to do. There are three main problems:

1. For personal interviews, but not telephone surveys from a computerized central facility, it can be hard to maintain timely information about interviewer results.

2. Interviewers can understate their refusals by assigning unsuccessful results to other categories.

3. Assignments to in-person interviewers may not be comparable, so that differences in rates of refusals per interviewer may not be consistent indicators of interviewer performance. This issue applies much less to telephone interviewers working in centralized facilities.

Response rates cannot be calculated accurately until a study is over, but special efforts to identify refusals by interviewer during data collection can alert supervisors to problems and are a very important part of interviewer supervision. It is difficult to help an interviewer who has response rate problems. On telephone studies, a supervisor can listen to introductions and provide feedback immediately after the interview (or noninterview) about how the interviewer might be more effective. For household in-person interviewers, the task is more difficult because the supervisor cannot observe the interviewer's approach unless the supervisor accompanies the interviewer on a trip. Thus the supervisor often must be content with listening to the interviewer give a sample introduction.

Supervisors can give helpful hints to interviewers. It is important to make sure interviewers are fully informed about a survey. Having interviewers practice giving concise, clear answers to common questions may be useful. In addition to working on the details of introductions, supervisors may need to address an interviewer's general feeling about approaching people or about the survey project and its value. There are limits, however, to how much retraining will help; there are people who never can attain good response rates. Although it is stressful, one of the most effective ways to keep response rates high is to take ineffective interviewers off a study.

Review of Completed Survey Instruments

When interviewers are using paper-and-pencil instruments, a sample of completed survey instruments should be reviewed to assess the quality of data interviewers are collecting. When reviewing a completed interview, one obviously can look for whether the recording is legible, the skip instructions are followed appropriately, and the answers obtained are complete enough to permit coding. In addition, looking at a completed interview can give a pretty good idea of the extent to which an interviewer is recording respondent answers verbatim, as compared to recording summaries or paraphrases. For computer-assisted interviews, these issues—except for the recording and probing associated with narrative answers—are not relevant.

The Question-and-Answer Process

The quality of interviewing cannot be supervised by reviewing completed survey instruments; they do not tell the supervisor anything at all about the way the interviewer conducted the interview and how those answers were obtained. In order to learn this, a supervisor must directly observe the interviewing process.

A telephone survey from a central facility permits direct supervision of how the interviewer collects the data. A supervisor can and should be available to monitor interviewers at all times. Some centralized systems include the capability of recording all or a sample of interviews. Supervisors should listen systematically to all or parts of a sample of the interviews that each interviewer takes, evaluating (among other things) appropriate introduction of the study, asking questions exactly as written, probing appropriately and nondirectively, and appropriate handling of the interpersonal aspects of the interview. This process works best if a rating form covering these and other aspects of an interviewer's work is completed routinely by a monitor (Cannell & Oksenberg, 1988).

When interviewers are doing studies in respondents' homes or in other distant places, it is more difficult to supervise the question-and-answer process. There are only two ways to do it: A supervisor can accompany an interviewer as an observer, or interviews can be recorded. Without recording or a program of observation, the researcher has no way to evaluate the quality of interviewing. All the most important aspects of the measurement process are unmonitored. Poor interviewers cannot be identified for retraining, and the researcher cannot report the quality of interviewing beyond saying that the interviewers were told what to do. Indeed, from the interviewer's point of view, it must be difficult to believe that standardized interviewing is important when it is the focus of training but is not attended to further.

Fowler and Mangione (1990) present evidence that personal interviewers are less likely to interview the way they are trained if their work is not monitored directly by tape recording. Both Fowler and Mangione and Billiet and Loosveldt (1988) found that data quality was improved when interviewers were monitored directly in this manner. It is now clear that direct supervision of the interview process should be a part of a well-managed survey. The fact that laptop computers can be set up to audio-record interviews makes it quite feasible to review interviewer behavior.

SURVEY QUESTIONS

Although training and supervision are important to producing good interviewing, perhaps the most important step a researcher can take to produce good interviewing is to design a good survey instrument. Research has shown that certain questions are misread consistently, whereas others consistently are answered inadequately, requiring interviewers to probe to obtain adequate answers (Fowler, 1991; Fowler & Cannell, 1996; Fowler & Mangione, 1990; Oksenberg et al., 1991). These questions can be identified with the kind of pretesting described in Chapter 7.

The more interviewers have to probe, explain, or clarify, the more likely they are to influence answers. The better the survey instrument, the more likely it is that the interviewer will conduct a good, standardized interview. The role of good question design in producing good interviewing is discussed in detail in Fowler and Mangione (1990) and Fowler (1991).

INTERVIEWING PROCEDURES

Training and Motivating Respondents

Studies have demonstrated the value of going beyond good question design to help standardize the interview (Cannell et al., 1987; Cannell, Oksenberg, & Converse, 1977; Miller & Cannell, 1977). For example, the researcher can help the interviewer train the respondent in a consistent way. Before the interview begins, the interviewer might read something like the following:

> Before we start, let me tell you a little bit about the interview process, since most people have not been in a survey like this before. You will be asked two kinds of questions in this survey. In some cases, I will be asking you to answer questions in your own words. In those cases, I will have to write down every word you say, not summarizing anything. For other questions, you will be given a set of answers, and you will be asked to choose the one that is closest to your own view. Even though none of the answers may fit your ideas exactly, choosing the response closest to your views will enable us to compare your answers more easily with those of other people.

Interestingly, interviewers like this instruction a great deal. It explains the respondents' task to them, and it makes the question-and-answer process go more smoothly. In fact, good interviewers give instructions such as these on their own. The value of providing explicit instructions is that it reduces differences among interviewers by having them all do the same thing. In addition, such instructions have a salutary effect on the interviewer's performance. Once the interviewer has read an instruction explaining the job expectations, it is easier to do the job the way it should be done, and it is a little harder to do it wrong, because the respondent now also knows what the interviewer is supposed to do (Fowler & Mangione, 1990).

Standardized instructions to respondents also can be used to set goals and standards for performance:

> It is very important that you answer as accurately as you can. Take your time. Consult records if you want. Ask me to clarify if you have any question about what is wanted.

Such statements ensure that respondents have a common understanding of their priorities. Some interviewers unintentionally promise respondents they will make it easy on respondents if the latter will just give the interview; interviewers who hurry communicate that speed is more important than accuracy. When an instruction such as the above is read, it forces accuracy and data quality to be a central part of the role expectations for both respondent and interviewer. One more source of between-interviewer variability is reduced, and the odds of good performance by both are increased.

In conclusion, there are critical parts of the interviewer's job besides the direct question-and-answer process. In particular, the interviewer is responsible for communicating to the respondent how the interview is to proceed: what the respondent is supposed to do, what the interviewer is going to do, and what their joint goals are. This aspect of the interviewer's job mainly has been left up to the interviewer, and not surprisingly, interviewers differ in how they do it in ways that affect data. By developing standardized instruction programs for respondents, researchers can make the job of the interviewer easier, reduce an important source of between-interviewer variance, and improve the extent to which interviewers and respondents behave in ways that will make the measurement process go better.

Standardized Wording

It was stated previously that asking questions exactly as worded is a foundation of standardized measurement, but not everyone agrees (Tanur, 1991). Critics of standardized interviewing have observed that some questions are not consistently understood by all respondents. When that is the case, they argue that it would produce better data if interviewers were free to clarify or explain the meaning of the question (e.g., Conrad & Schober, 2000; Schober & Conrad, 1997). In a similar vein, critics note that some data collection tasks—for example, when the same information is being gathered about several different people or events—produce very stilted or awkward interactions when interviewers try to use only scripted wording. In these instances, it is argued that giving interviewers more flexibility with wording would result in a more comfortable interviewer–respondent interaction (Schaeffer, 1992).

Some of the criticism of standardized interviewing is primarily the result of poorly designed questions (see Houtkoop-Steenstra, 2000; Suchman & Jordan, 1990). When questions are unclear or provide awkward scripts for interviewers, the solution often is to write better questions, not to have interviewers redesign the questions (Beatty, 1995). There is real basis for concern that when interviewers are given flexibility to reword or explain the questions, they will do it in a way that changes the meaning of questions and makes the resulting

data worse, not better (Fowler & Mangione, 1990). However, there are certain questions—such as repetitive series or when a few respondents need detailed definitions that would be cumbersome to provide to all respondents—that might be better handled by giving interviewers more flexibility. Moreover, when interviewers make changes in question wording it has not consistently been shown to increase interviewer-related error or response error (Dykema, Lepkowski, & Blixt, 1997; Fowler & Mangione, 1990).

There have been some experiments giving interviewers more discretion about how to ask and probe questions (Conrad & Schober, 2000; Schober & Conrad, 1997). To date the results have been mixed: the accuracy of some reports may be improved, but considerably increased interviewer training and sometimes longer interviews are involved. When and how to give interviewers more flexibility is a topic that warrants further experimentation. Meanwhile, for most surveys, designing questions that interviewers can and will ask exactly as worded remains the primary way to conduct a good survey.

VALIDATION OF INTERVIEWS

The possibility that an interviewer will make up an interview is a potential concern. The likelihood of this happening varies with the sample, the interviewing staff, and the field procedures. For the most part, concern about validation is restricted to surveys in which interviewers are conducting interviews in respondents' homes or are doing telephone interviews from their own homes. In such cases, the actual collection of data is not observable by supervisors. The number of hours to be devoted to carrying out an interview is often sufficient to motivate an interviewer to make up an interview rather than take the time and effort to carry it out.

In the long run, probably the best protection against faked interviews is to have a set of interviewers that have some commitment to the quality of the research and the organization. Such problems seem to occur most often with newly hired interviewers. Even organizations with an experienced, professional staff, however, routinely check a sample of interviews to make sure they actually were taken.

There are two approaches to this type of validation. One approach is to mail all respondents a brief, follow-up questionnaire asking about reactions to the interview. Probably a more common procedure is to have interviewers obtain a telephone number from every respondent; a sample is called by a supervisor. Simply knowing in advance that a validation by mail or telephone will be done is likely to be a deterrent to interviewer cheating. In addition, to be able to say that such a check was done may be reassuring to users of the data.

THE ROLE OF INTERVIEWING IN SURVEY ERROR

As noted at the onset of this chapter, interviewers affect response rates, the accuracy of reporting, and the consistency or precision of measurement. Each of these has a central role in the quality of a survey estimate.

One of the most observable effects of good survey management is the response rate. Although this issue is discussed more thoroughly in Chapter 4, it is worth repeating that the quality of an interviewing staff is critical to the rate of response that will be obtained in any particular survey.

It is more difficult to measure the error introduced by interviewers in the question-and-answer process. Often survey error is undetectable. When asking questions about subjective states, objective checks for bias or inaccuracy are generally not meaningful, as was discussed in Chapter 6. There have been studies, however, in which researchers had objective measures of facts respondents were asked to report, permitting evaluation of the accuracy of reporting. In one such study (Cannell, Marquis, & Laurent, 1977), samples of households in which someone had been hospitalized in the year preceding were interviewed. The accuracy of reporting could be evaluated by comparing the health interview reports of hospital stays with hospital records. One measure of reporting accuracy was simply the percentage of known hospitalizations that was reported.

In this study, it was found that the number of interviews assigned to an interviewer correlated very highly ($r = .72$) with the percentage of hospitalizations that were unreported in the interview. Interviewers who had large assignments, with whatever pressures that brought to bear on them, collected much less accurate data than those with small assignments.

A different study using the same criterion (the percentage of hospitalizations reported; Cannell & Fowler, 1964) reached a similar conclusion. In this case, half of an interviewer's respondents reported hospitalizations in an interview, whereas the other half completed a self-administered form regarding hospitalizations after the interviewer had completed the rest of the health interview. It was found that interviewers whose respondents reported with great accuracy when asked to report hospitalizations in the interview also had accuracy when asked to report hospitalizations in the interview also had respondents who reported very well in the self-administered form after the interviewer had left ($r = .65$). This study suggested not only that interviewers had a critical role to play in affecting the error of their respondents reporting, but also that one way in which interviewers affected respondent performance was the degree to which they motivated respondents to perform well. In both cases, the effect of the interviewer on reporting accuracy was clear.

In the absence of validating data, one cannot assess accuracy. However, it is possible to assess the extent to which interviewers influence the answers of

their respondents. If an interviewing staff operated in a perfectly standardized way, one would be unable to explain any variation in answers by knowing who the interviewer was. To the extent that answers are predictable in part from knowing who did the interview, it can be concluded that the interviewer is inappropriately influencing answers. Groves (1989) thoroughly discusses the techniques for calculating the extent to which interviewers were affecting the answers to questions and summarizes the results of numerous studies in which interviewer effects were calculated. It turns out that for many questions that interviewers ask, one cannot see any effect of the interviewer on the answers. For between one third and one half of the questions in most surveys, however, interviewers significantly affect the answers.

The result of these interviewer effects is to increase the standard errors around survey estimates. The size of the multiplier depends on the size of the intraclass correlation (rho) and on the average size of interviewers' assignments (see Groves, 1989; Kish, 1962). If the intraclass correlation is .01 (which Groves found to be about the average), and the average number of interviews per interviewer is about 31, the standard errors of means will be increased by 14% over those estimated from the sample design alone. When interviewer assignments average closer to 50, for items with an intraclass correlation of .02, the estimates of standard errors will be increased by 41%.

Out of this discussion there are several points to be made about the role of the interviewer in the total error structure of survey data:

1. In addition to their role in response rates, interviewers can be associated with the extent to which respondents give inaccurate answers in surveys and with measurement inconsistency. Existing evidence clearly indicates that interviewers are a significant source of error for many kinds of measures.

2. The training and supervision that interviewers receive can significantly increase the consistency of interviewers, thereby improving the reliability of estimates, and reduce bias. In particular, interviewers who receive minimal training (e.g., less than 1 day) and interviewers who receive minimal or no feedback about the quality of their interviewing are poorer interviewers.

3. Procedures that structure the training and instruction of respondents, minimize inappropriate interviewer feedback, and in general, control more of the interviewer's behavior can reduce interviewer effects on data and increase overall accuracy.

4. Better question design is a key to better interviewing.

5. One design option that has been underappreciated is the size of the average interviewer assignment. Although training and management costs may be

lower if fewer interviewers are used, researchers may pay a price in data reliability for allowing individual interviewers to take large numbers of interviews. Reducing average interviewer assignments often is a cost-effective way to increase the precision of survey estimates.

6. Virtually all reports of the reliability of survey estimates ignore the effects of interviewers on data. In part, this is because researchers cannot sort out interviewer effects from sampling effects when interviewers are assigned samples on a nonrandom basis, such as convenience or geographic proximity. Interviewer effects are a significant source of error, however, for many items in most surveys. Any report of the precision of a survey estimate that ignores interviewer effects is likely to be an underestimate of survey error.

In conclusion, the role of the interviewer in contributing to error in survey data has not been appreciated generally. Although most survey researchers know that some training is necessary for interviewers, procedures for training and supervising interviewers vary widely and often are not adequate. It is unusual for researchers to make any efforts beyond training and supervision to minimize interviewer effects. Yet, these aspects of survey design constitute some of the most cost-effective ways to improve the quality of survey data. The impact of the interviewer on survey estimates deserves a central place in the design and reporting of survey studies that it has not yet achieved.

EXERCISES

1. Tape-record some role-played interviews in which you and/or others use a standardized interview schedule (the questions developed in Chapter 6, or a schedule from another source). Then listen to the tapes and systematically evaluate interviewer performance by noting for each question at least the following errors: did not read question exactly as worded; probed an inadequate answer in a biasing (directive) way; failed to probe an unclear answer; or any other possibly biasing or unstandardized interpersonal behavior. The evaluations are particularly instructive if done by a group, so that interviewer errors can be discussed.

2. Perform a similar exercise role playing an interviewer trying to enlist the cooperation of a potential respondent for an interview.

Further Readings

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

9

Preparing Survey Data for Analysis

Survey answers usually are transformed into data files for computer analysis. This chapter describes options and good practice for data formats, code development, coding procedures and management, data entry, and data checking procedures.

Once data have been collected by a survey, no matter what the methods, they almost invariably must be translated into a form appropriate for analysis by computer. This chapter is about the process of taking completed questionnaires and survey interviews and putting them into a form that can be read and processed by a computer. The process of coding or data reduction involves five separate phases:

- deciding on a format (the way the data will be organized in a file)
- designing the code (the rules by which a respondent's answers will be assigned values that can be processed by machine)
- coding (the process of turning responses into standard categories)
- data entry (putting the data into computer readable form)
- data cleaning (doing a final check on the data file for accuracy, completeness, and consistency prior to the onset of analysis)

There are two kinds of errors that can occur in going from an answer to an entry in a data file. First, there can be transcription or entry errors any time someone records an answer or number. Second, there can be coding decision errors, misapplications of the rules for equating answers and code values. The options for quality control are tied to the particular data collection, data entry, and coding procedures chosen. Those options and various alternative procedures are discussed below.

FORMATTING A DATA FILE

The term *record* as used here refers to all the data that pertain to a single individual case or interview. A record can consist of one or more lines of data.