

ROYCE A. SINGLETON, JR. • BRUCE C. STRAITS

Approaches to  
SOCIAL  
RESEARCH

FIFTH EDITION



- c. To carry out an internally valid experimental test of DARE, what features of the study would you need to have control over? Explain.
5. Suppose in the audit study reported in Box 8.1 that the researchers found no interaction effect but that both race and resume quality had statistically significant effects on call-backs for interviews. Graph this outcome in a figure like the one in Box 8.1.

### Notes

1. This section and the next three sections draw heavily on Donald Campbell and Julian Stanley's classic treatment of these subjects (1963).
2. Campbell and Stanley (1963) originally used the term "mortality" to describe this source of invalidity. However, "attrition," which is now commonly used, seems better to capture the intended meaning.
3. However, if the teacher initiated the experiment one day because of a sharp increase in talking out of turn, the incidence of this behavior may regress toward a more typical classroom level on the posttest.
4. However, this design does not by itself control for effects of unintended events that might occur *within* a treatment group. For a discussion of this problem and suggestions for dealing with it, see Campbell and Stanley (1963:13–14).
5. Notice how the principle of "doing only one thing at a time" applies in interpreting factorial results. The only difference between the first-row and the second-row groups is the factor-A manipulation. The two groups are equivalent on factor B (each row has an equal number of pretested and nonpretested subjects) and should be approximately equivalent on extraneous variables as a result of subject randomization.
6. For a more complete discussion, see Campbell (1969b) and Cook and Campbell (1976, 1979).

# 9



## Survey Research

Survey research in its many forms has become a very common activity in today's world, and most of us have had some experience with it in one form or another. Perhaps you have been stopped on the street by a news reporter and asked your opinion on some issue of local or national importance. You may have responded to a reader survey found in a popular magazine. Or perhaps you have replied to an Internet questionnaire when registering a newly purchased product. You or someone else in your household very likely responded to the last U.S. Census, which attempted to enumerate and gather confidential information about every person living in the United States. You have been a consumer of survey research if you have read in the newspaper the results of Gallup or Roper public-opinion surveys. You have done a little "survey research" of your own if you moved to a new community and asked a number of residents about local restaurants or where to obtain various services.

### General Features of Survey Research

These survey examples differ in their degree of formality and in the extent to which they conform to these typical features of professional survey research:

1. A large number of respondents are chosen through probability sampling procedures to represent the population of interest.
2. Systematic questionnaire or interview procedures are used to ask prescribed questions of respondents and record their answers.
3. Answers are numerically coded and analyzed.

As we now elaborate on these three features through reference to actual studies, we will also describe exceptions to the general rule for each feature.

### Large-Scale Probability Sampling

Professional surveys make use of large samples chosen through scientific sampling procedures to ensure precise estimates of population characteristics. National opinion polls typically number around 1000 respondents, but surveys of national samples can be much larger. In the National Educational Longitudinal Study (NELS), for example, the probability sample interviewed in 1988 consisted of 24,599 8th-graders from 1052 public and private schools. Participating students were asked

about a wide range of school, family, and home experiences; and these data were supplemented by information obtained from parents and teachers as well as achievement test scores and school characteristics such as teacher and student composition. Beckett Broh (2002) analyzed the NELS data in her study of the impact of extracurricular involvement on academic achievement. Using data from follow-up surveys of NELS respondents, Thomas Deleire and Ariel Kalil (2002) examined the effect of family structure on four developmental outcomes among teenagers: whether they graduate from high school, attend college, smoke or drink, and initiate sexual activity. As previous researchers had found, youths from nonmarried families did not fare as well as youths from married families on all four outcomes. A major exception to this pattern, however, was youths from single-parent multigenerational families, who did just as well as youths from married families.

The NELS was conducted by the National Opinion Research Center on behalf of the National Center for Educational Statistics of the U.S. Department of Education. Such national studies require considerable resources—time, money, and personnel—that are beyond the capacity of independent researchers or small research teams. Many surveys, therefore, involve somewhat smaller samples drawn from state or local populations. Kim Fridkin, Patrick Kenney, and Jack Crittenden (2006), for example, surveyed the political attitudes of 166 Anglo, 37 African American, 125 Latino, and 58 Native American 8th-grade students in six integrated public schools randomly drawn from metropolitan Phoenix. Anglo adolescents were more likely to discuss politics and watch news programs at home, practice democratic skills at school, and hold positive views of government. Among minority adolescents, Native Americans had the fewest opportunities to develop democratic skills at home and school and expressed the greatest distrust of the political system.

Useful surveys also are conducted with nonprobability samples. Patrick Carr, Laura Napolitano, and Jessica Keating (2007), for example, interviewed a purposive sample of delinquent and nondelinquent young men and women living in three high-crime neighborhoods in Philadelphia. Although most of the youth were negatively disposed toward the police because of their personal experiences with law enforcement, they nevertheless favored a tougher and fairer police force as a means of reducing crime in their neighborhoods.

Thus, although large-scale probability samples are the ideal, surveys vary considerably in sample size and sampling design. There are legitimate reasons for doing a small-scale survey, particularly if you have a low budget or some specialized or applied research purpose. Indeed, students can often conduct their own low-budget research if they have a research problem that can be studied appropriately with a brief questionnaire survey of the home campus or a telephone survey of the local dialing area.

At the other end of the spectrum from surveys of local populations are surveys that draw samples from more than one nation. In comparative or **cross-national surveys**, equivalent surveys are conducted in different countries. For example, based on interviews with 9529 married women from twenty-two different nations, Carrie Yodanis (2005) analyzed whether tolerance of divorce on the national level is associated with enhanced or diminished gender equality within marriage. Her results suggest that wives experience a more equal division of household labor when living

in a country in which divorce is accepted and practiced. The ease of divorce in a society was associated with a more equal division of household labor.

### Units of analysis

While in most surveys the units of analysis are individuals, this is not always the case. An example of a survey treating employers as units is the Current Employment Statistics (CES) program, a monthly sample of about 150,000 businesses and government agencies, conducted by the U.S Bureau of Labor Statistics to measure changes in employment, wages, and hours (see Current Employment Statistics). Another example utilized cities as units to study community decisions about whether to fluoridate water supplies to reduce tooth decay. A controversial issue in many communities in the 1950s, the proposal to fluoridate provided social scientists with an excellent opportunity to study the political decision-making process in cities. To learn how cities had dealt with the issue of fluoridation, Robert Crain, Elihu Katz, and Donald Rosenthal (1969) sent questionnaires to three informants in each of 1181 cities: the public-health officer, the publisher of the largest city newspaper, and the city clerk. One finding was that when the city decided the issue by referendum, the odds against fluoridation were five to one.

### *Systematic Procedures: Interviews and Questionnaires*

Surveys obtain information through interviews and/or self-administered questionnaires. The survey of political attitudes among 8th-graders was based on interviews; the study of youth in high-crime neighborhoods used a combination of interviews and self-administered questionnaires; the study of political decision making in cities made use of questionnaires only.

An additional example of the use of questionnaires is the College Alcohol Study (see Wechsler and Nelson, 2008), a widely cited national survey of American college students' drinking habits. The mail questionnaire, completed by more than 50,000 students at 120 colleges, asked numerous questions about drinking behavior and other health issues. The most controversial findings concerned binge drinking, operationally defined as the consumption of five or more drinks in a row for men and four or more drinks in a row for women during the 2 weeks prior to the survey. In each of the 1993, 1997, 1999, and 2001 surveys, the researchers found that two in five students who responded were binge drinkers. Binge drinkers, especially those who frequently binged, were far more likely than nonbinge drinkers to experience a variety of alcohol-related and other health problems, such as engaging in unprotected and unplanned sex, getting in trouble with campus police, damaging property, and getting hurt or injured.

### Unstructured versus structured interviewing

Regardless of whether the survey researcher makes use of interviews only, questionnaires only, or some combination of the two, the procedures tend to be standardized for all respondents in order to enhance the reliability of the data. This

epitomizes a **structured interview**. The objectives are very specific, all questions are written beforehand and asked in the same order for all respondents, and the interviewer is highly restricted in such matters as the use of introductory and closing remarks, transitions or "bridges" from topic to topic, and supplementary questions to gain a more complete response (called "probes"). There are exceptions to this rule, however, as informative and scientifically useful interviewing is sometimes carried out in a less formal or structured manner.

In an **unstructured interview**, the objectives may be very general, the discussion may be wide-ranging, and individual questions will be developed spontaneously in the course of the interview. The interviewer is free to adapt the interview to capitalize on the special knowledge, experience, or insights of respondents. An everyday example of an unstructured interview might be a journalist's interviewing a celebrity to learn more about his or her personal background, interests, and lifestyle. Between the two extremes, the **semistructured interview** would have specific objectives, but the interviewer would be permitted some freedom in meeting them. The scope of the interview would be limited to certain subtopics, and key questions probably would be developed in advance.

The choice of a highly structured, semistructured, or unstructured approach depends on the researcher's objectives. For example, in their ongoing study of undocumented migration among rural Mexicans, Douglas Massey and associates (1987) wanted to gather quantitatively rigorous data but ruled out a highly structured approach because of the sensitive nature of the study and the respondents' limited background (many were poorly educated or illiterate). Their semistructured interview form listed subtopics or data to collect for each household member. The interviewers were given considerable discretion as to question wording and timing to ensure that the interview was informal and not threatening to the respondents. To make sure that the flexible, semistructured interviews produced comparable or standardized information from each respondent, the interviewers were given extensive training and field supervision. This approach, when combined with probability sampling of households in four carefully chosen communities, yielded precise quantitative information in 1987 on Mexican migration to the United States. The four communities were resurveyed in 2007 (see Mexican Migration Project).

When the research purpose is not to derive facts or precise quantitative descriptions but to understand the meaning of respondents' experiences (see Warren, 2002), unstructured interviews often are adopted. This approach allows maximum flexibility in the development of hypotheses and theory; however, it seldom is associated with survey research because it invariably involves small nonprobability samples and qualitative analysis. The Carr, Napolitano, and Keating (2007) study of attitudes toward the police relied primarily on unstructured interviews to explore adolescents' complex views of police and law enforcement. To identify emerging themes, the researchers tape-recorded, transcribed, and then coded the interviews using qualitative data-analysis software. Two emerging themes were "the police" and "how to reduce crime." Some responses with these themes were the product of direct questioning; however, many of the youth talked about the police in response to other questions, such as when asked to describe their neighborhood and the dangers they encountered in their daily lives. As these answers were probed, it became

clear that the attitudes of youth in the high-crime neighborhoods were shaped by their direct experiences with law enforcement.

For some research purposes, a social scientist might utilize two or three sets of interviews, beginning with very loosely structured interviews and progressing to a final set of highly structured interviews. A freer interviewing style in the preliminary stages would yield rich and varied information. This would assist the researcher in formulating or refining hypotheses, clarifying objectives, and specifying subtopics for subsequent semistructured interviews. Findings from a second set of interviews might be applied to the development of a highly structured questionnaire.

### **Quantitative Data Analysis**

Data-analysis techniques depend on whether the survey's purpose is descriptive, explanatory, or a combination of the two. Surveys that are primarily **descriptive** seek to describe the distribution within a population of certain characteristics, attitudes, or experiences and make use of simpler forms of analysis. **Explanatory** surveys, on the other hand, investigate relationships between two or more variables and attempt to explain these in cause-and-effect terms. Sorting out the relationships between the variables in an explanatory survey requires the use of more sophisticated data-analysis techniques. (Data analysis will be dealt with further in chapters 15 and 16.)

An illustration of a primarily descriptive study is the National Health and Social Life Survey (NHSLS), in which a national probability sample of adults between the ages of 18 and 59 were interviewed about their sexual attitudes and practices (Laumann et al., 1994). Originally prompted by the lack of national statistics on the prevalence of various sexual behaviors relevant to the AIDS epidemic, the NHSLS was transformed by actions of federal agencies and hostile politicians into a study with a somewhat broader focus, a much smaller than planned sample (3432 people were interviewed rather than 20,000), and a sponsorship by private rather than by public funding.

Among the interesting NHSLS findings are estimates of same-gender sexual activity and forced sexual behavior. Only 1.4 percent of the women and 2.8 percent of the men during the interview identified themselves as "homosexual" or "bisexual." Yet in a short self-administered questionnaire that the respondent placed in a sealed envelope, 4.3 percent of the women and 9.1 percent of the men reported having engaged in a sexual act with someone of the same gender since puberty. These and other NHSLS estimates of homosexuality are substantially lower than reported in the Kinsey report and other studies that did not use probability sampling (Laumann et al., 1994:286–90). Perceptions of unwanted sexual activity varied by gender. Although only 2.8 percent of the men reported ever having forced a woman to do something sexual that she did not want to do, 21.6 percent of the women reported forced sex by a man.

The aforementioned NELS represents explanatory survey research. The study began with a baseline survey of students' school and school-related experiences in the 8th grade and continued with follow-up surveys 2, 4, 6, and 12 years later. The general purpose of the NELS was to relate the 8th-grade experiences to high school performance and, in turn, to later achievements in life. The wealth of data enables

investigators to examine the impact of numerous factors—for example, students' school experiences and activities; their relationships with peers, parents, and teachers; and characteristics of their schools—on students' academic achievement, persistence in school, participation in postsecondary education, and career choices.

### **Secondary Analysis of Surveys**

In most of the examples presented thus far, each survey had a central topic and the investigators were responsible for all phases of the research project, from the formulation of ideas and research design to the collection, analysis, and presentation of the data. Although this was once the dominant approach, researchers today are more apt to analyze survey data collected by some other person or agency than to conduct an original survey themselves (Presser, 1984). The analysis of survey data by analysts other than the primary investigator who collected the data is called **secondary analysis**. Analyses of the NELS data by Beckett Broh and others represent this form of research.

One of the many advantages of this research strategy is that sample size can be increased greatly by combining data from several surveys. For example, in their study of the long-term effects of education, Herbert Hyman, Charles Wright, and John Shelton Reed (1975) drew together fifty-four surveys conducted between 1949 and 1971, involving 76,671 respondents, from three sources: the Gallup poll, the National Opinion Research Center (NORC) of the University of Chicago, and the Survey Research Center at the University of Michigan. Not surprisingly, a very strong positive relationship was found between amount of education and correct answers to knowledge questions. However, this was true not only for "academic"-type questions taken from the arts and sciences but for questions relating to current affairs and popular culture as well. Furthermore, the relationship persisted over time, suggesting that one of the lasting effects of education is a lifelong openness to learning or tendency to seek information.

The earliest application of secondary analysis was to census data and then, beginning in the 1950s, to opinion-poll data. These data, however, were collected primarily for administrative and journalistic purposes. The major impetus for secondary analysis came in the 1970s with the advent of surveys designed expressly for the purpose of making high-quality data available to the social science research community (Glenn, 1978). The first large-scale survey of this type, called the General Social Survey (GSS), began in 1972 and was conducted annually (except for 1979, 1981, and 1992) until 1994. Each GSS involved personal interviews with about 1500 respondents, drawn from a probability sample of the adult population of the United States. Starting in 1994 the GSS shifted to biennial surveys with two 1500-person samples in even-numbered years. In 2006 a third sample was added for a total sample size of 4510 respondents.

In contrast to most surveys, which have a central topic, the GSS is eclectic, with questions pertaining to a broad range of attitudes and behavior. A portion of each GSS includes questions that are replicated across years to facilitate research on social trends; time trends are available for 269 variables for 20 or more years between 1972 and 2006. The remainder of each survey is devoted largely to topical and cross-

national modules, which comprise blocks of questions relating to special topics. The cross-national modules, which are sponsored by the International Social Survey Program (ISSP), are included in national surveys in over thirty collaborating countries. The Yodanis comparative study of divorce culture and marital gender equality used the 1994 ISSP module on Family and Changing Gender Roles.

The 2008 GSS contains topical modules on self-employment, preparedness for terrorism, global economics, sexual orientation, Jewish identity, firearms, clergy/congregant contacts, knowledge about and attitudes toward science, religious trends, social inequality, and sexual behavior, as well as two ISSP modules, one on religion and the other on sports/leisure. The religion module, which replicates questions asked in 1991 and 1998, includes questions on sexual morality, abortion, science, tolerance, beliefs, and religious upbringing. The ISSP sports/leisure module includes measures on leisure-time activities and preferences, time pressures, evaluation of the value of sports, TV viewing of sports, and group memberships.

Data from the GSS are often used in the teaching of sociology and other social science courses, especially statistics (see Babbie, Halley, and Zaino, 2007; Frankfort-Nachmias and Leon-Guerrero, 2006). Moreover, analyses of GSS data can be found in over 14,000 books, articles, chapters, and dissertations (National Opinion Research Center [NORC], 2008). We make extensive use of GSS data in chapters 15 and 16. In this chapter and the next, we will report some of the survey methods and materials used in producing these data so that the reader may become more familiar with this important database.

### **The Uses and Limitations of Surveys**

Now that we have examined a wide range of surveys, we are in a position to consider the uses as well as the strengths and weaknesses of this approach. With the exception of the NELS and CES, conducted by a federal government agency, all of the studies we have cited were carried out by social scientists for social scientific purposes. In fact, "surveys are the most widely used method of collecting data in the social sciences, especially in sociology and political science" (Bradburn and Sudman, 1988:61). Yet the use of surveys as a scientific tool outside the scientific community is even more extensive. The ubiquitous opinion polls, oft reported in the news media and eyed warily by politicians, monitor public reactions to people, events, and policies. Marketing research by businesses, advertising agencies, and other organizations tests consumer reactions to new products and services, assesses customer satisfaction, and compiles audience profiles for various media. And the single largest user of surveys, the federal government, conducts or commissions scores of surveys every year to help in planning, decision making, and policy assessment. Given the importance of surveys in shaping major decisions by politicians, businesspeople, and government officials, we all need to know something about surveys. What can they tell us better than other methods of social research? And what are their major limitations?

Whereas experiments are used almost exclusively for explanatory, hypothesis-testing research, surveys are used extensively for both descriptive and explanatory

purposes. Among all approaches to social research, in fact, surveys offer the most effective means of social description; they can provide extraordinarily detailed and precise information about large, heterogeneous populations. By using probability sampling, one can be certain, within known limits of sampling error, whether the responses to a sample survey accurately describe the larger target population. Furthermore, the topics covered and the questions that may be included in surveys are wide-ranging. Topics of the studies we have cited ranged from academic achievement to alcohol consumption and from sexual activity to attitudes toward the police. The scope of possible survey questions is suggested by the following classification (Schuman and Kalton, 1985):

1. Social background information (e.g., What is your religious preference? What is your date of birth?)
2. Reports of past behavior (e.g., Did you vote in the last presidential election? Have you ever been the victim of a crime? On an average day, about how many hours do you personally watch television?)
3. Attitudes, beliefs, and values (e.g., Do you believe that there is a life after death? Do you think homosexual couples should have the right to marry one another?)
4. Behavior intentions (e.g., If the presidential election were held today, whom would you vote for? Would you yourself have an abortion if there were a strong chance of serious defect in the baby?)
5. Sensitive questions (e.g., Have you ever been arrested for a crime? Have you used cocaine in the past month?)

For categories 1, 2, and 5, which pertain to behavior and personal characteristics, the information may be verifiable from records or observer reports; but it is often impractical, unethical, or even illegal to obtain it from sources other than the individuals themselves. For subjective phenomena such as categories 3 and 4, the information can be directly known, if at all, only by asking the individuals themselves.

As this listing suggests, surveys can address a much broader range of research topics than experiments can. Ethical considerations preclude studying some topics experimentally—for example, the effect of emotional traumas on mental health—while practical considerations rule out many others; for instance, one normally cannot experimentally manipulate organizations or nations. Besides this flexibility, surveys can be a very efficient data-gathering technique. While an experiment usually will address only one research hypothesis, numerous research questions can be jammed into a single large-scale survey. Furthermore, the wealth of data typically contained in a completed survey may yield unanticipated findings or lead to new hypotheses.

The secondary analysis of surveys also affords many unique advantages. The cost of obtaining the data for analysis is usually a small fraction of the cost of collecting and coding the data. Survey data made available for secondary analysis tend to come from professional polling and research centers with the resources to obtain high-quality information from large, national samples. In addition, secondary analysis may enable one to (1) assess social trends by examining questions repeated over time and (2) increase sample size by combining data from several surveys.

The major disadvantage of surveys relates to their use in explanatory research. Beyond association between variables, the criteria for inferring cause-and-effect relationships cannot be established as easily in surveys as in experiments. For example, the criterion of directionality—that a cause must influence its effect—is predetermined in experiments by first manipulating the independent (or causal) variable and then observing variation in the dependent (or effect) variable. But in most surveys this is often a matter of interpretation since variables are measured at a single point in time. Consider also the criterion of eliminating plausible rival explanations. Experiments do this effectively through randomization and other direct control procedures that hold extraneous variables constant. In contrast, surveys must first anticipate and measure relevant extraneous variables in the interviews or questionnaires and then exercise statistical control over these variables in the data analysis. Thus, the causal inferences from survey research generally are made with less confidence than inferences from experimental research.

Although surveys are quite flexible with respect to the topics and purposes of research, they also tend to be highly standardized. This makes them less adaptable than experiments and other approaches in the sense that it is difficult to change the course of research after the study has begun. That is, once the survey instrument is in the field, it is too late to make changes. The experimenter, in contrast, can modify the research design after running a few subjects with the loss of only those subjects.

A more serious weakness of surveys is one they share with laboratory experiments: They are susceptible to reactivity, which introduces systematic measurement error. A good example of this, noted in chapter 5, is the tendency of respondents to give socially desirable answers to sensitive questions. Another inherent weakness is that surveys rely almost exclusively on reports of behavior rather than observations

of behavior. As a consequence, measurement error may be produced by respondents' lack of truthfulness, misunderstanding of questions, inability to recall past events accurately, and instability of opinions and attitudes. Finally, a brief encounter for the purpose of administering a survey does not provide a very good understanding of the context within which behavior may be interpreted over an extended period of time. For this kind of understanding, the best approach is field research, discussed in chapter 11.

## Survey Research Designs

*Research design* generally refers to the overall structure or plan of a study. The crucial design features of experiments, with their focus on hypothesis testing, reveal how a given study will test a specific hypothesis. Surveys place much less emphasis, however, on this aspect of research design. The basic idea of a survey is to measure variables by asking people questions and then to examine the relationships among

the measures. The major design option is to ask the questions once or to repeat the questions over time.

### **Cross-Sectional Designs**

The most commonly used survey design by far is the **cross-sectional design**, in which data on a sample or "cross section" of respondents chosen to represent a particular target population are gathered at essentially one point in time. By "one point in time" we do not mean that respondents are interviewed or that self-administered questionnaires are collected simultaneously (although questionnaires might be in some studies). Rather, the data are collected in as short a time as is feasible. For the 2006 GSS, it took over 4 months, between March and July, for about 200 interviewers to complete interviews with 4510 respondents (T. W. Smith, personal communication, 2008). Most of the studies cited earlier are cross-sectional designs.

Two variations on the cross-sectional survey design have been developed out of sociological interest in studying the influence of social contexts and interpersonal relations on individual behavior. Cross-sectional surveys are limited by the amount and accuracy of the information that individual respondents can capably report about the groups and milieus to which they belong. Contextual designs and social network designs address this problem by studying individuals and relationships found within the same social context.

**Contextual designs** sample enough cases within particular groups or contexts to describe accurately certain characteristics of those contexts. Doris Entwistle, Karl Alexander, and Linda Olson (1994) employed this design to study if the gender gap in math achievement favoring boys over girls might be explained by the contextual resources of their schools and their neighborhoods. The investigators first sampled twenty Baltimore elementary schools and then sampled children from each 1st-grade classroom in the selected schools. Student improvement in math skills during the first 2 years of school was predicted from individual-level variables (including initial math reasoning score, race, sex, parents' education, family economic standing) and contextual-level variables (average number of years of schooling for parents of sampled children in each of the twenty schools, racial composition of each school, and median neighborhood household income). The analysis revealed that the boys' gains in math skills were more affected by resources outside the home (school and neighborhood contexts) than were the girls'.

**Social network designs** focus on the relationships or connections among social actors (people, organizations, countries, etc.) and the transaction flows (processes) occurring along the connecting links.<sup>1</sup> Network adherents reject analytical approaches that isolate actors from their social environment (network linkages) and that interpret behavior solely in terms of actor attributes (characteristics). For example, social psychologists may hypothesize that the likelihood of two people becoming friends increases (1) if they share common interests and values (the traditional attribute approach) or (2) if they share a friend in common (the network approach).

Social network designs typically require the interviewing of every person in the group under study. This makes it possible to delineate networks of personal rela-

tionships by asking respondents to provide such information as who their best friends are, whom they most like to work with on a certain project, or to whom they would go for advice. In a social network study of a male-dominated advertising firm, Herminia Ibarra (1992) found that women had to form costly, differentiated interactional networks by choosing (1) other women as friends and sources of support and (2) men mostly as instrumental (work-related) resources. The men, in contrast, chose other men for both expressive and instrumental purposes.

### **Longitudinal Designs**

Because cross-sectional designs call for collection of data at one point in time, they do not always show clearly the direction of causal relationships and they are not well suited to the study of process and change. Of course, investigators can make inferences about the logical relations among variables, and respondents can be asked about both past and present events. But both of these sources of evidence are highly fallible. To provide stronger inferences about causal direction and more accurate studies of patterns of change, survey researchers have developed **longitudinal designs**, in which the same questions are asked at two or more points in time. The questions may be asked repeatedly either of independently selected samples of the same general population or of the same individuals. This results in two main types of longitudinal designs: trend studies and panel studies. A **trend study** consists of a repeated cross-sectional design in which each survey collects data on the same items or variables with a new, independent sample of the same target population. This allows for the study of trends or changes in the population as a whole. Trend studies may be illustrated by the monthly government surveys used to estimate unemployment in the United States (target population) as well as by repeated public-opinion polls of candidate preferences among registered voters (target population) as an election approaches.

Tom Smith's (2008) analysis of GSS family-related trends, for example, shows the tremendous changes in American family structure and values between 1972 and 2006. Among married couples with children under age 18, the traditional gender roles of an employed husband and a wife working at home rapidly fell from 60 percent in 1972 to 46 percent in 1980 and to 37 percent in 1985 and then leveled off at under 30 percent in 1993–2004. Accompanying these changes in gender roles were shifts in family values; for example, disagreement with the statement "A preschool child is likely to suffer if his or her mother works" rose from 32 percent in 1977 to 46 percent in 1985 to 59 percent in 2006.

Ideally, all trend information would be obtained through measures repeated frequently at regular intervals, as in the GSS. However, much of our trend-survey data come from infrequent replications of studies. For example, in 1993 Danching Ruan and associates (1997) replicated a 1986 social-network survey in urban China to see if structural changes in Chinese society would be reflected in changes in personal relationships. Both surveys asked respondents to name others with whom they have discussed important matters over the last 6 months. The discussion partners named in 1986 were primarily co-workers (44 percent) and relatives (39 percent). The 1993 replication revealed a dramatic shift in personal connections: More

friends and associates and fewer co-workers and relatives were named as discussants. The 1986–93 changes in people's discussion networks were consistent with concurrent large-scale social changes, including a diminished role of the traditional workplace in people's lives and the new opportunities provided by an emerging market economy.

Most trend studies measure changes in the general population over time. To study the developmental effects of aging as well as chronological changes, it is also possible to focus on a specific cohort of persons. A *cohort* consists of persons (or other units such as organizations and neighborhoods) who experience the same significant life event within a specified period of time (Glenn, 1977). Most often the life event that defines a cohort is birth, but it also might be marriage, completion of high school, entry into medical school, and so forth. Demographers long have analyzed birth cohorts from census data to predict population trends. In the past half-century, social researchers also have begun to do *cohort studies* of various attitudes and behaviors by tracing changes across cohorts in repeated cross-sectional surveys. Cohorts are identified (or tracked) by their birth date; for example, the 1972 cohort would be 18 years old in a 1990 survey, 28 years old in a 2000 survey, and 38 years old in a 2010 survey.

Cohort trend studies enable one to study three different influences associated with the passage of time. To get a sense of these influences and of the difficulty of studying them at a single point in time, consider a cross-sectional survey containing measures of age and attitudes toward premarital sex. The GSS, for example, includes the following question: "There's been a lot of discussion about the way morals and attitudes about sex are changing in this country. If a man and woman have sex relations before marriage, do you think it is always wrong, almost always wrong, wrong only sometimes, or not wrong at all?" If we found that responses to this question became more conservative (premarital sex is always or almost always wrong) with age or time, this could be due to one of three kinds of influences: *life course* (as people grow older, they become more conservative), *cohort* (older generations are more conservative than younger generations), or *historical period* (the prevailing culture has become more conservative over time, making it less socially acceptable to hold liberal views toward premarital sex).

The problem with cross-sectional data is that one cannot begin to disentangle these various effects. Longitudinal data in general and cohort analyses in particular are generally superior for this purpose, although the reader should be aware that these techniques seldom provide clear causal inferences (Glenn, 1977). An example of a cohort study is David Harding and Christopher Jencks' (2003) analysis of age, cohort, and period effects on changing attitudes toward premarital sex. Using GSS data from 1972 to 1998, data from a 1965 NORC survey, and Gallup surveys in 1962, 1969, and 1973, Harding and Jencks broke the samples down into 10-year age cohorts according to the decade when respondents turned 18 years of age, from the 1920s to the 1990s. In general, they found a sharp increase in liberal attitudes toward premarital sex since 1962, with much of the change occurring in the early 1970s, which they attributed to a period effect. Within each cohort, they also found an increase in conservatism after people turned age 30, which they attributed to an aging effect.

Whereas trend studies identify which *variables* are changing over time, panel studies can reveal which *individuals* are changing over time because the same respondents are surveyed again and again. Paul Lazarsfeld, Bernard Berelson, and Hazel Gaudet's 1948 classic study of voter behavior, *The People's Choice*, exemplifies the panel method. Before the 1940 presidential election, 600 persons were interviewed repeatedly between May and November. The analysis revealed that persons who expressed a clear preference for the Democratic or the Republican candidates at the first interview were unlikely at the second interview 1 month later to remember having seen or heard any campaign propaganda from the party of the opposing candidate. Because of this phenomenon of selective attention, few voters changed their preferences over the course of the study.

#### KEY POINT

*Panel surveys* measure changes in individuals over time, *trend surveys* track general social changes, and *cohort surveys* gauge changes in age groups.

By repeatedly surveying persons over an extended period of time, a panel study can record life histories—schooling experiences, labor force participation, income changes, health conditions, and so forth—much better than retrospective questions from a cross-sectional survey. Some of the early panel studies analyzed data from small,

highly selective samples over very long periods of time. Glen Elder's (1999) 30-year follow-up study of the impact of the Great Depression on the life course analyzed data from a panel study that began in 1932 with the selection of 167 5th-grade children from five elementary schools in Oakland, California. These individuals were observed, tested, and questioned more than a hundred times before they graduated from high school in 1939. They then completed a short questionnaire in 1941, were interviewed extensively in 1953–54 and 1957–58, and completed a final follow-up mailed questionnaire in 1964. In a study of attitude change among Bennington College students, women who were surveyed initially in the late 1930s either were interviewed or completed a mailed questionnaire in 1960 and then again in 1984 (Alwin, Cohen, and Newcomb, 1991). A major finding of this study was the persistence of political attitudes over the adult years, as the Bennington women, who were liberalized in the college years, remained politically liberal in later life.

Panel studies of any duration were a rarity in the social sciences until the late 1960s, when the federal government began conducting large-scale longitudinal studies for secondary data analysis. In addition to the NELS, which surveyed a sample of 8th-graders in 1988 and then did follow-up surveys of these same respondents in 1990, 1992, 1994, and 2000, other examples include the Panel Study of Income Dynamics (PSID) and the National Longitudinal Survey of Youth (NLSY). The PSID has collected data on a sample of U.S. households annually from 1968 to 1997 and biennially beginning in 1999; by the end of 2003, over 65,000 individuals had been surveyed. Starting in 1979 with interviews of over 12,000 individuals 14–22 years of age, the NLSY reinterviewed this sample annually through 1994 and biennially since then.

Two drawbacks to studies of this magnitude are that they are very expensive and that they take considerable time. In addition, panel surveys have two problems not found in cross-sectional designs: respondent attrition and reactivity stemming

from repeated measurement (Duncan, 2001). Panel attrition occurs when respondents interviewed in the initial wave do not respond in later waves. The more respondents who drop out, the less likely that follow-up samples will be representative of the original population from which the initial sample was drawn. Therefore, researchers typically make great efforts and devote sizeable resources, including incentive payments, to retain a high percentage of the initial sample in follow-up surveys. The NLSY, for example, offered respondents \$10 for their time. In the early years, over 95 percent of the initial respondents were reinterviewed; by 1992, 13 years after the initial 1979 interview, the NLSY was still able to interview 90 percent of the original sample (Brandon, Gritz, and Pergamit, 1995). No matter what the rate of follow-up response, however, a panel survey is like the experimental pretest-posttest design in its ability to assess and control for the effects of attrition. That is, initial information on respondents who fail to participate in later waves can be used to assess nonresponse bias and to make necessary statistical adjustments.

Interviewing respondents repeatedly also may affect how individuals respond to the survey; two problems are “time-in-sample bias” and “panel conditioning” (Sikkel and Hoogendoorn, 2008). The first bias may occur when panel participation becomes routinized to the point that respondents merely repeat what they have said before, rather than carefully consider their current feelings and actions. Or realizing from previous interviews that a “yes” response leads to follow-up questions, respondents may avoid answering the extra questions by answering “no” (Kalton and Citro, 1993). Panel conditioning may stimulate real attitude or behavioral changes so that respondents’ attitudes and behavior are no longer representative of the larger population. In national election studies, for example, Michael Traugott and John Katosch (1979) found that voter registration and turnout increased with the number of times participants were interviewed. On the other hand, G. J. Duncan (2001:1101) argues that “pervasive behavioral effects” are unlikely in panel surveys, “especially when changes in the behavior under investigation require more effort than making a trip to the polls.” He also contends that data from subsequent waves may be less biased as respondents understand better the purpose of the study and become more motivated to give accurate answers.

### Steps in Survey Research: Planning

The activities in doing surveys fall into three broad categories: (1) planning, (2) field administration, and (3) data processing and analysis. The remainder of this chapter is devoted to planning and field administration. Chapters 15 and 16 deal with data processing and analysis.

In planning a survey, a few activities may be thought of as key decision points. These are represented in the flowchart shown in Figure 9.1. Under each major activity we have indicated the chapters that deal with it.<sup>2</sup>

The initial stages in planning a survey are essentially the same as in other forms of research. The first step is to select a topic and formulate a problem in researchable terms. One then reviews relevant journal articles, books, and other materials to determine what is known about the topic and what work remains to be done. Dur-

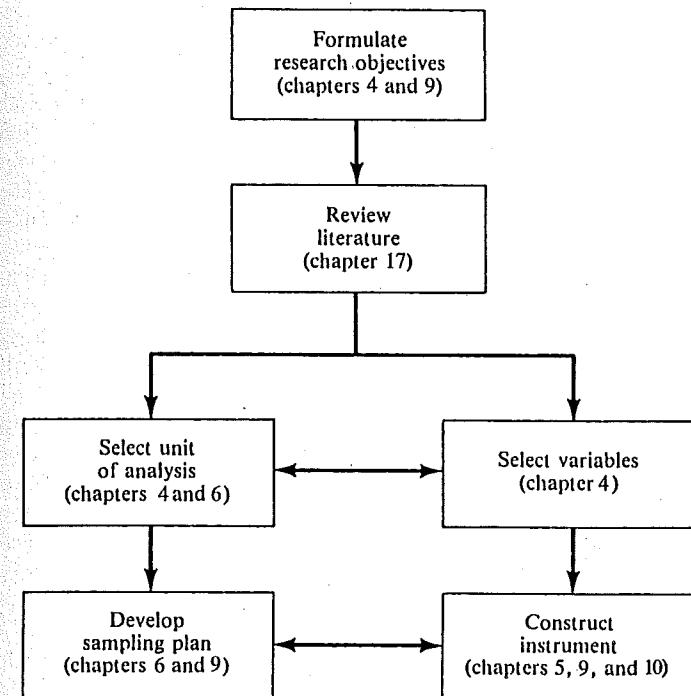


FIGURE 9.1. Key decision points in planning a survey.

ing the course of this review, the researcher inevitably will refine and further specify his or her objectives; also, he or she may become aware of existing measures or scales that may be incorporated into the prospective study. The selection of units of analysis and variables occurs in light of research objectives. In survey research, the units of analysis are either individuals or groups of individuals (including families, organizations, cities); the particular variables selected depend on which characteristics should be studied to meet research objectives.

A very effective planning device when thinking about variables is to work backward mentally from the final steps in the study to the earlier steps in the planning phase. Serious mistakes can be avoided by anticipating the data analysis. To do this, one works out the actual data analysis along with the steps likely to be taken in the event of various outcomes. One might ask questions such as these: If the data support my hypothesis, what will I do next? If the data are the opposite of my hypothesis, what will I do next? (I might decide that I need to control additional variables to test the hypothesis properly.) Have I included all the variables I might want to use? By anticipating the data analysis in this way, the survey content can be planned better. (But, then again, this is not as easy as it sounds, as James Davis reveals in his report of a social survey that we recount in Box 9.1.)

Having selected the unit of analysis and identified key variables, the survey researcher turns his or her attention to the development of the sampling plan and sur-

### BOX 9.1 An Informal Account of a Large-Scale Survey

Researchers do not, as a rule, report the false starts, unsolvable roadblocks, embarrassing problems, and mistakes that are a normal part of the process of doing surveys. James Davis (1964), however, has provided a humor-filled account of a national survey, the Great Books study, from its inception to the publication of a book. Davis describes the events as an illustration of how survey research is carried out in a large research organization.

The parties involved in the Great Books study included the National Opinion Research Center (NORC), with James Davis as study director; the Fund for Adult Education, as client; and the object of the study, the Great Books Foundation. The Fund for Adult Education, a subsidiary of the Ford Foundation, had been supporting, through grants, various educational activities and programs, including Great Books. The Fund hoped by commissioning an evaluation study to gain concrete evidence of the value of the Great Books program and to determine whether the continuation or expansion of support was merited. The Great Books Foundation, a nonprofit corporation, believed in its program, according to Davis, but seriously doubted the usefulness of a survey in evaluating it. The parties in the study agreed at the outset that the study would be concerned with the effects on participants of participation in Great Books.

The Great Books program in 1957–58 involved approximately 1960 discussion groups throughout the United States, with additional groups in Canada and elsewhere, which met every 2 weeks from September to June. At each meeting, the members discussed a book that they had been assigned to read. Book selections were organized into blocks of 1 year. The size of the groups varied, with a sample average of eleven. The groups were sponsored chiefly by public libraries but also by churches, businesses, and individuals. One or two persons served as leaders of each group. No tuition was charged, and no certificates of completion were given. Readings could be bought from the Foundation inexpensively if desired.

The design of the study fell naturally into two parts: sampling and questionnaire construction. In regard to sampling, the "ideal" design would have been a field experiment in which a large probability sample of persons were randomly assigned to a control group or to a Great Books discussion group. After a specified period, both groups would be measured on the dependent variables. However, such a design clearly was not feasible. Among other problems the effects were expected to take a long time to appear, and a report had to be delivered in about a year. The compromise design would use beginning participants as a control group and compare them with advanced-year members on a number of variables. The sample was stratified to overrepresent the advanced-year groups, where the effects might be expected to show. Groups were asked to distribute and complete self-administered questionnaires at a regular meeting. Thus, although it would have been ideal to have had a sample of randomly assigned *individuals* in and out of the Great Books program, practical considerations resulted in a stratified sample of Great Books discussion groups.

Constructing the questionnaire presented more difficulties than devising the sampling plan. The Fund for Adult Education, the Great Books Foundation, and the study director Davis had different ideas about the program's effects. The Fund, interested in community participation, argued that participants would become more involved in community affairs as a result of their exposure to great literature. Davis and the Great Books Foundation doubted that this was the case. The Great Books Foun-

dation insisted that its program did not have any specific purposes that could be evaluated by a survey. Under pressure from Davis, the Foundation was able to make some very general statements regarding the program's objectives: Participants should become more open-minded, develop more critical thinking skills, become more intellectually sophisticated, and so forth, effects that would be difficult to measure. Davis was interested in members' philosophies, ideologies, and tastes. He also wanted to include some social network questions to see how a person's acceptance or nonacceptance by the group affected his or her reactions to the program. However, the Great Books Foundation ruled out the social network questions as possibly offensive to participants. Instead, questions were asked about roles played by the respondent and others in the group. The final questionnaire contained items reflecting the Fund's interest in community participation; a few items related to aesthetics, open-mindedness, and critical thinking; measures of tastes, ideologies, and philosophies; and the role questions.

Over 90 percent of the sampled groups returned the questionnaires. Analysis of the survey results showed marked differences in knowledge between beginner's and advanced-year members, slight differences in tolerance and open-mindedness, and few behavioral differences.

A report for the sponsor was completed on time in August 1958. The project, however, was not finished. Davis immediately conceived the idea of doing a follow-up study to find out which respondents had dropped out of the groups during the year so that data from these individuals could be subtracted. The remaining data would be reanalyzed to see whether omission of the dropouts affected the findings. The Fund agreed to sponsor the follow-up study. It turned out that controlling for dropouts had little effect.

A new study then was conceived to analyze, with the budget money remaining, the factors associated with dropping out. The idea was not only to learn why Great Books members drop out but to view the data more broadly as "why some small-scale social systems lose the commitment of their members" (p. 228). Davis admitted he hoped by this study to impress his colleagues that he was capable of theory-building research, as opposed to purely empirical work. The study was to lead to the publication of *Great Books and Small Groups*.

The study showed that retention of participants was greatest where a large percentage of members played active roles. Theories that the content of the roles is important were not supported, nor was Davis's hunch that retention was related to a balance of power in the group.

At the end of Davis's account (1964:233–34) of the Great Books study, he gives a beautiful description of the satisfactions of doing research:

There is a lot of misery in surveys, most of the time and money going into monotonous clerical and statistical routines, with interruptions only for squabbles with the client, budget crises, petty machinations for a place in the academic sun, and social case work with neurotic graduate students. And nobody ever reads the final report. Those few moments, however, when a new set of tables comes up from the machine room and questions begin to be answered; when relationships actually hold under controls; when the pile of tables on the desk suddenly meshes to yield a coherent chapter; when in a flash you see a neat test for an interpretation; when you realize you have found out something important that nobody ever knew before—these are the moments that justify research.

vey instrument. As the flowchart in Figure 9.1 shows, these plans are made concurrently as each affects the other and both depend on the objectives and resources of the researcher. Constructing the survey instrument or questionnaire requires a great deal of time and thought and the making of many decisions. We devote the entire next chapter to this aspect of survey design.

A critical aspect of survey instrumentation is deciding on the mode of asking questions—interviewer-administered (face-to-face or telephone surveys) or self-administered (paper-and-pencil or computer-assisted questionnaires) or some combination of these modes. This choice depends partly on other planning decisions such as the research objectives, units of analysis, and sampling plan. For example, in the College Alcohol Study, the sampling plan required contact with a very large number of respondents at over 100 geographically dispersed colleges and universities. These requirements ruled out both face-to-face and telephone interviews as too expensive, time-consuming, and impractical; and questionnaires were used. In the study of young people's attitudes toward law enforcement, personal interviews were necessitated by the sensitivity and complexity of the topic and by the researcher's need for an approach flexible enough to explore how attitudes were formed.

The choice of survey mode, in addition to the overriding factor of cost, will determine the optimum sampling design. The most expensive and time-consuming mode of survey research is face-to-face interviewing, the major costs of which are incurred from direct interviewing time and travel to reach respondents. If respondents are widely dispersed geographically, this method also will require an efficient sampling procedure for locating respondents. Under these circumstances, the most cost-efficient procedure usually is multistage cluster sampling. Almost all large-scale surveys are multistage, with stratification at one or more stages. If respondents are reached easily by mail or phone, there is no reason for use of clustering. Simple random or systematic sampling may be implemented, with or without stratification, provided that an adequate sampling frame can be obtained.

As you can see, survey planning is not a linear series of decisions. Rather, it requires the simultaneous consideration of the likely costs and benefits of a number of choices. Also, as planning progresses, revisions in previous decisions and plans frequently are needed. Survey costs and benefits are best viewed from Robert Groves's (1989) "total error perspective," which identifies four types of errors that threaten the accuracy of survey results. We discussed each of these error sources in the chapters on sampling and measurement:

1. *Coverage error*: Differences between the target population and the sampling frame; this is produced when the sampling frame does not include all members of the population, as when a telephone survey omits people without telephones.
2. *Sampling error*: The difference between a population value and a sample estimate of the value that occurs because a sample rather than a complete census of the population is surveyed.
3. *Nonresponse error*: Differences in the characteristics of those who choose to respond to a survey and those who refuse or cannot be contacted be-

cause of an insufficient address, of a wrong telephone number, they are never at home, or they are on vacation.

4. *Measurement error*: Inaccurate responses associated with the respondent, the interviewer, the survey instrument, and the postsurvey data processing.

Ideally, the survey planner effectively allocates available funds and resources to minimize the impact of these potential errors. This is difficult because the errors may be interrelated (reducing one may increase another), certain types of error may be too costly to reduce appreciably, and information on the extent of the errors may

be limited or lacking. The choice of a stratified random sampling plan, for example, might reduce sampling error but increase coverage error compared with a simple random sampling design. Reducing nonresponse error by efforts to convert reluctant respondents may inadvertently increase measurement error (Lavrakas, 1993:78).

Conversely, efforts to decrease measurement error may increase nonresponse error. Survey error can never be totally eliminated. When it is too expensive to reduce a source of error, it may be possible to estimate the magnitude and direction of the error and make suitable adjustments in the data analysis.

We now consider various survey modes. Figure 9.2 lists the four basic data-collection modes along with typical computer-based variations. The modes may be conceptualized as falling along a continuum from the most to the least interactive. At one end of the continuum, involving all channels of communication, is the face-to-face interview; this is followed, in turn, by telephone interviews, various computer-assisted self-interviews, and autonomous self-administered questionnaires. Although we focus primarily on the four basic modes, recent technological developments have rapidly expanded the options and encouraged the use of multiple modes of data collection. Therefore, we discuss mixed-mode data collection strategies after comparing the modes shown in Figure 9.2. Each mode, as you will see, has its distinctive advantages and disadvantages, especially with respect to error reduction. Because sampling error is primarily a function of sampling design, sample size, and population heterogeneity, and not a function of the data-collection mode, our concern here is with the other three sources of error, especially coverage error and nonresponse error.

### Face-to-Face and Telephone Interviewing

Both the face-to-face interview and the telephone interview require trained interviewers proficient in the use of the survey's *interview schedule*. The interview schedule consists of instructions to the interviewer together with the questions to be asked and, if they are used, response options. Interviewers and an interview schedule permit a great deal more flexibility than is possible with a self-administered questionnaire. For example, when research objectives necessitate the use of

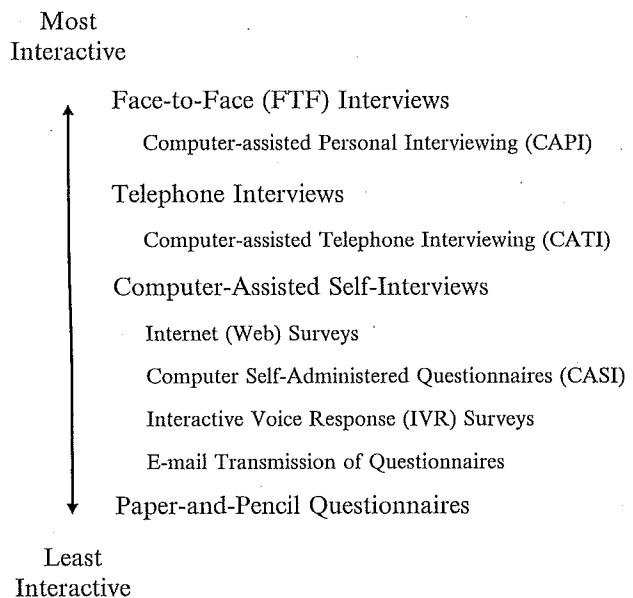


FIGURE 9.2. Major data-collection modes.

*open-ended questions*, which require respondents to answer in their own words, in contrast to *closed-ended questions*, for which specific response options are provided, an interviewer usually will be able to elicit a fuller, more complete response than will a questionnaire requiring respondents to write out answers. This is particularly true with respondents whose writing skills are weak or who are less motivated to make the effort to respond fully. In addition, interviewers can easily utilize question formats in which certain questions are skipped when they do not apply to a particular respondent, while such a format may be confusing for respondents completing a questionnaire. Furthermore, in cases where it is important that questions be considered in a certain order, the self-administered questionnaire presents problems because the respondent may look over the entire form before beginning to answer.

Other advantages of interviewing include the ability of an interviewer to clarify or restate questions that the respondent does not at first understand. An interviewer may also help respondents clarify their answers by using *probes*, such as "I'm not sure exactly what you mean" or "Can you tell me more about that?" Interviewers help to ensure that every relevant item is answered; tedious or sensitive items cannot be passed over easily as in self-administered questionnaires. Even when a respondent initially balks at answering an item, a tactful explanation by the interviewer of the item's meaning or purpose frequently results in an adequate response.

In addition to these characteristics that both modes of interviewing share, each has its own set of advantages and disadvantages.

### Face-to-Face Interviewing

The oldest and most highly regarded method of survey research, **face-to-face (FTF)** or **in-person interviewing** has a number of advantages in addition to the ones already mentioned. The **response rate**, the proportion of people in the sample from whom completed interviews (or questionnaires) are obtained, is typically higher than in comparable telephone or mail surveys (de Leeuw, 2008:128–29). Nonresponses, however, have been increasing in Western government and academic FTF surveys (de Leeuw and de Heer, 2002). In recent GSSs, the overall response rate has been around 70 percent; it was 71.2 percent in 2006 (Davis, Smith, and Marsden, 2007:Table A-6, Appendix A). Reasons for the high response rate probably include the intrinsic attractiveness of being interviewed (having someone's attention, being asked to talk about oneself, the novelty of the experience); the difficulty of saying "no" to someone asking for something in person; and possibly the fact that the importance and credibility of the research are conveyed best by an FTF interviewer who can show identification and credentials.

This survey mode is appropriate when long interviews are necessary. FTF interviews of 1 hour's length are common, but they sometimes go much longer. (GSS interviews take about 90 minutes for completion of some 400 questions.) When FTF, one can use visual aids such as photographs and drawings in presenting the questions, as well as cards that show response options. The cards may be useful when response options are difficult to remember or when it is face-saving for respondents to select the option or category on the card rather than to say the answer aloud. Finally, FTF interviewing permits unobtrusive observations that may be of interest to the researcher. For example, the interviewer may note the ethnic composition of the neighborhood and the quality of housing.

There are some disadvantages to this method, the greatest of which is cost. The budget for an FTF survey must provide for recruiting, training, and supervising personnel and for interviewer wages and travel expenses, plus lodging and meals in some cases. In one experiment by Herschel Shostack and William Fairweather (1979) in which comparable surveys of physicians were conducted by mailed questionnaires and by personal interviews, the field cost per initial respondent—that is, the data-gathering expenses divided by the number in the sample—amounted to approximately \$63 for personal interviews and \$24 for the mailed questionnaire. These figures excluded the costs of initial planning and data coding, processing, and analysis. Groves et al. (2004:161) estimate that national FTF surveys cost five to ten times as much as telephone surveys. The full cost of the 2006 GSS, from sampling through release of the data to the public, was about \$1100 per respondent (T. W. Smith, personal communication, 2008).

The difficulty of locating respondents not at home when the interviewer first calls is another disadvantage of this survey mode. In more and more households, no adult is at home during the day. This necessitates call-backs in the evening or, as in the GSS, generally restricting contacts to weekdays after 4 PM and on weekends. Even then, many persons in large cities will not open their doors to strangers in the evening, and many interviewers refuse to go into certain areas of cities at night. Fear of strangers and the desire for privacy may be the causes of another dis-

advantage: The response rate for heterogeneous samples in metropolitan areas has been declining for several years. However, this is not true for rural areas or among specialized target groups.

Staff supervision presents special difficulties. Frequently, interviewers, data coders, and the researcher are geographically dispersed. If an interviewer has not recorded responses adequately, effective coding is impossible. Furthermore, many interviews may be completed by an interviewer before feedback gets back to the individual. This is less of a problem when interviewer supervisors go over each interview schedule soon after it is completed. Some of the problems associated with interviewer mistakes may be minimized with **computer-assisted personal interviewing (CAPI)**. A computer program, usually on a portable laptop computer, prompts the interviewer with instructions and question wording in the proper order, skips questions not relevant to particular respondents, ensures that the interviewer enters appropriate response codes for each question, may indicate when respondents are giving inconsistent responses to related questions, and may even digitally record random interview segments with a laptop microphone for quality control. CAPI has become the standard for large-scale survey research in the United States; the GSS switched to CAPI in 2002 and most U.S. government in-person surveys now use CAPI (Couper and Hansen, 2002).

Finally, with the personal interview method, interviewers may introduce bias into the data in a number of ways. For example, they may fail to follow the interview schedule in the prescribed manner or may suggest answers to respondents. Bias also may be introduced through a respondent's reaction to the interviewer's gender, race, manner of dress, or personality. In chapter 7 we spoke of the experiment as a social occasion and the related possibilities of bias. An FTF interview is no less a "social occasion" than an experiment; consequently, interviewers must be carefully trained to be sensitive to the ways in which they may wittingly or unwittingly affect their interviewees' responses.

### **Telephone Interviewing**

Like FTF interviewing, **telephone interviewing** has its advantages and disadvantages. Substantial savings in time and money are two of the reasons survey researchers choose to use this method. Large survey research organizations that have a permanent staff can complete a telephone survey very rapidly, and even those researchers who must hire and train interviewers can complete a telephone survey faster than one requiring FTF interviews or mailed questionnaires. The costs for sampling and data collection in telephone surveys have been estimated to be 10 to 15 percent of those for FTF interview surveys (Groves et al., 2004:161). However, telephone survey costs will exceed those for mailed questionnaires, even with several follow-up mailings included.

Another major advantage of telephone interviewing is the opportunity for centralized quality control over all aspects of data collection (Lavrakas, 1993), including question development and pretesting, interviewer training and supervision, sampling and call-backs, and data coding and entry. Administration and staff supervision for a telephone survey are much simpler than for a personal interview sur-

vey. No field staff is necessary; in fact, it is possible to have the researcher, interviewers, and coders working in the same office. This arrangement permits supervisors to monitor ongoing interviews, allowing immediate feedback on performance and helping to minimize interviewer error or bias. Coders may be eliminated and the interviewers can enter numbers corresponding to respondent answers directly into a computer terminal. If they are used, coders may provide immediate feedback to interviewers and their supervisors.

In terms of sampling quality, the telephone survey mode falls between the FTF interview and the mailed questionnaire. In the past, lists of telephone subscribers were used in the sampling process, creating a problem of coverage error by excluding those who had unlisted telephone numbers or moved too recently to be directory listed (Steeh, 2008:226). The problem of missing those with unlisted numbers was resolved through **random-digit dialing (RDD)** procedures in which telephone numbers are chosen randomly. In the simplest RDD design, telephone prefixes (exchanges) within the target geographic area are sampled and then the last four digits of the telephone number are randomly selected. In actual practice, more complex "list-assisted" RDD procedures are used to screen out nonresidential and other ineligible numbers, and "dual-frame" procedures are being developed to sample both mobile and landline phones (Groves et al., 2004:128–30; Steeh, 2008:227–28).

Besides RDD, the introduction of **computer-assisted telephone interviewing (CATI)** in the 1970s greatly enhanced the efficacy of this survey mode (Nathan, 2001). CATI refers to a set of computerized tools that aid interviewers and supervisors by automating various data-collection tasks. The uses of CATI include sampling and dialing phone numbers, scheduling call-backs, screening and selecting the person to be interviewed at each sampled phone number, prompting the interviewer with appropriate introductions and questions in the proper sequence, skipping irrelevant questions, identifying when responses are inconsistent with replies to earlier questions, ensuring that the interviewer enters legitimate response codes for each question, recording the responses into a computer data file, and producing immediate sampling and interviewing updates for supervisors (Lavrakas, 1993). Finally, an added benefit of CATI systems is that complex factorial designs may be seamlessly incorporated into surveys (see chapter 8).

Still, telephone surveys have their limitations. The declining response rates for telephone interviews are well documented by the monthly University of Michigan's Survey of Consumer Attitudes (SCA), in which overall response rates gradually declined from a high of 72 percent in 1979 to 60 percent in 1996 and then dropped more steeply to 48 percent in 2003 (Curtin, Presser, and Singer, 2005). The substantial 1979–2003 drop in response to the SCA reflects both an increasing inability to reach eligible households and an unwillingness of contacted persons to be interviewed. Factors contributing to the plunging telephone response rates include the growth of caller ID call-screening and call-blocking technologies, the increase in phone numbers dedicated to computer modems or faxes, the proliferation of new area codes and nonhousehold telephone numbers, heightened privacy concerns in the face of increased telemarketing calls and the subsequent National Do-Not-Call Registry, and the increase in cell phone-only households (Curtin, Presser, and Singer, 2005; Link et al., 2008).

Other limitations of telephone surveys are interview duration and the complexity of the questions asked. Conducting a telephone interview longer than 20 to 30 minutes increases the risk of nonresponse and mid-interview termination (de Leeuw, 2008). While the interviewer may repeat a question, it is desirable to develop questions simple enough to be understood and retained by respondents while they formulate an answer. Research also has shown that open-ended questions yield shorter, less complete answers in telephone interviews than in FTF interviews (Groves and Kahn, 1979). Furthermore, questions with multiple response options may present difficulties in that the interviewer cannot present the options on cards but must read and, if necessary, repeat them to respondents at the risk of boring them. For these reasons, the telephone survey mode lacks the advantages of the FTF mode in regard to the types of questions that are used.

Another disadvantage of the telephone interview is that it is more difficult for interviewers to establish trust and rapport with respondents than it is in FTF interviews; this may lead to higher rates of nonresponse for some questions and under-reporting of sensitive or socially undesirable behavior. Robert Groves (1979) compared the results of two identical telephone surveys based on separate samples with the results of an FTF survey asking the same questions. At the end of the questionnaire were items about respondents' reactions to the interview. Among other questions, respondents were asked if they felt uncomfortable talking about certain topics, such as income, their income tax refund, political opinions, or racial attitudes. For each of the sensitive topics, more telephone respondents felt uncomfortable, with the largest differences for the income and income tax questions. Not surprisingly, the telephone surveys showed lower response rates to the income questions. Other studies confirm that respondents are less likely to divulge illegal or socially undesirable behavior to an interviewer by telephone than FTF (Aquilino, 1994:211, 214; Holbrook, Green, and Krosnick, 2003).

Despite these disadvantages, telephone surveys became the most popular survey method in the United States and Western Europe in the last quarter of the twentieth century. Reduced time and cost are major advantages. Furthermore, closer supervision, developments in RDD and CATI, and the high percentage of households with telephones made the quality of telephone surveys only slightly inferior to that of FTF interviewing. Now, however, unless a solution can be found for rapidly declining telephone response rates, some survey re-

#### KEY POINT

Compared to telephone interviewing, face-to-face interviewing has higher response rates, allows visual aids, permits long and complex questionnaires, but is more expensive.

searchers offer a grim prognosis for the long-term future of this survey mode (Kalton, 2000; Curtin, Presser, and Singer, 2005:97). Perhaps the biggest challenge is the rapid proliferation of mobile telephones and the growth of the cell phone-only population. John Ehlen and Patrick Ehlen (2007) forecast that 40 percent of U.S. adults under age 30 will have a cell phone-only lifestyle by the end of 2009. Complex methodological, statistical, legal, and ethical issues must be resolved to develop effective dual-frame sampling of mobile and landline telephones

(Lavrakas et al., 2007). In light of these mounting obstacles to telephone surveys, some foresee an increasing reliance on self-administered questionnaires (Dillman, 2007).

## Paper-and-Pencil Questionnaires

Occasionally, the site of a paper-and-pencil-questionnaire is a school or organization, where the questionnaire may be hand-delivered and filled out in a group or individually. Most often, however, the setting is the home or a workplace. To get to this setting, almost all self-administered questionnaires are mailed to respondents. Therefore, we will discuss the pros and cons of this method as a *mail survey*.

This is less expensive than FTF or telephone interviews, even though the budget for printing and postage must be sufficiently high to permit follow-up mailings. No interviewers or interviewer supervisors are needed, there are no travel or telephone expenses, and very little office space is required. In some surveys, the staff may consist of just one or two persons in addition to the researcher. Groves et al. (2004:161) estimate that mail surveys generally cost 59 to 83 percent less than telephone surveys.

The time required to complete the data-collection phase of the survey is greater than that for telephone surveys but usually less than that for FTF surveys. The sample size may be very large, and geographic dispersion is not a problem. Furthermore, there is greater accessibility to respondents with this method since those who cannot be reached by telephone or who are infrequently at home usually receive mail.

Compared with interviews, self-administered questionnaires offer several advantages to motivated respondents (Mangione, 1995). Respondents are free to select a convenient time to respond and to spend sufficient time to think about each answer. The absence of an interviewer also assures privacy, which may explain why respondents are less willing to reveal illegal or socially undesirable behaviors or other sensitive information to an interviewer than in a self-administered questionnaire (Tourangeau and Smith, 1996:277–79).

On the other hand, coverage and nonresponse error may be magnified with this survey mode. The researcher must sample from a mailing list, which may have some incorrect or out-of-date addresses and which may omit some eligible respondents.<sup>3</sup> Also, the response rate tends to be much lower with mailed questionnaires than with other survey modes. A meta-analysis of forty-five research studies in which two or three data-collection modes were compared found that, on the average, FTF surveys have the highest response rate (70.3 percent), telephone surveys the next highest (67.2 percent), and mail surveys the lowest (61.3 percent) (Hox and de Leeuw, 1994). Still, even though rates of 50 percent or lower are fairly common in mail surveys, it is possible to obtain higher response rates by following the highly detailed "tailored design" procedures developed by Don Dillman (2007).

The most important factors in generating high return rates are reducing the costs for the respondent and increasing the perceived importance and rewards of survey participation (Heberlein and Baumgartner, 1978; Yammarino, Skinner, and Childers, 1991). Costs are reduced by including postpaid return envelopes, enclosing small cash

prepayments (rather than monetary incentives conditional upon survey completion), and making the questionnaire shorter, visually appealing, and easier to complete (Dillman, Sinclair, and Clark, 1993; Warringer et al., 1996). The importance of the survey is impressed on respondents by using stamped rather than metered mail, by making special appeals within the cover letter, by personalizing correspondence (e.g., with real signatures and salutations with respondents' first names), and by making repeated contacts in the form of preliminary notification and follow-ups emphasizing different appeals (Dillman, 2007; Mangione, 1995).

In addition to a generally lower overall response rate, self-administered questionnaires may introduce nonresponse bias due to response selectivity. Certain groups of persons, such as those with little writing ability and those not interested in the topic, would be less likely to respond to a mailed questionnaire than to a personal interview request. Also, more questions are left unanswered with self-administered questionnaires than with interview methods. The problem of item nonresponse may be alleviated to some extent by instructions explaining the need for every item to be answered, by assurances of confidentiality, and by making items easy to understand.

While interviewer bias is eliminated, so are the advantages of an interviewer. There is no opportunity to clarify questions, probe for more adequate answers, or control the conditions under which the questionnaire is completed or even who completes it. A mailed questionnaire usually yields the most reliable information when closed questions are used, when the order in which questions are answered is unimportant, and when the questions and format are simple and straightforward.

The questionnaire may serve the research purposes well under the following conditions: with specialized target groups who are likely to have high response rates,<sup>4</sup> when very large samples are desired, when costs must be kept low, when ease of administration is necessary, and when moderate response rates are considered satisfactory.

### Computer-Assisted Self-Interviews

The advent of computers and the Internet has spawned several new data-collection technologies. We have already mentioned computer-assisted personal and telephone interviewing (CAPI and CATI). By automating many tasks and simplifying others, these methods reduce interviewer errors and facilitate the interview process. However, they do not displace the interviewer; they simply make his or her job easier. More recently, researchers have developed a variety of computer-mediated surveys that are self-administered. These include e-mailed questionnaires, interactive voice response (IVR) surveys, computerized self-administered questionnaires, and Internet (Web) surveys (Figure 9.2).

E-mail and Web surveys are conducted over the Internet. Both involve computer-to-computer transmission of a questionnaire; in e-mail surveys the questions are sent as the text of an e-mail message or in an attached file, whereas in Web surveys the questionnaire is on specially designed Web pages. IVR surveys are conducted by telephone as respondents listen to prerecorded, voice-read questions and then use Touch-Tone data entry or give verbal answers, which are recorded (see

Steiger and Conroy, 2008). In **computer-assisted self-administered interviewing (CASI)**, the questionnaire is transmitted on a program that may be sent to respondents or on a laptop provided by the researcher. Of these innovations, Web surveys have had the broadest application. IVR surveys, which may not be suitable for long or complex questionnaires, have been used extensively in consumer marketing research but have had limited application in general social surveys. CASI often has been used in conjunction with interview-administered surveys. We focus here on the advantages and disadvantages of Web surveys.

One of the greatest advantages of Web surveys is reduced cost. Compared to self-administered questionnaires, the cheapest of the traditional modes, Internet surveys eliminate the costs of paper, postage, assembly of the mailout package, and data entry (Dillman, 2007). The principal costs are computer equipment and programming support, questionnaire development and testing, and Internet service provider fees. A related advantage is time savings. Web surveys require much less time to implement than other survey modes; compared to mail surveys, which may take weeks or months for questionnaires to be delivered and returned, Web surveys may be completed in only a few days (Kwak and Radler, 2002). Finally, Web surveys can substantially reduce the cost of increasing sample size because once the electronic questionnaire has been developed, the cost of surveying each additional person is far less than in an interview or mail survey (Dillman, 2007).

Another advantage of Web surveys, one they share with other computer-mediated methods, is flexibility in the questionnaire design. As Don Dillman (2007:354) points out, the questionnaire can be designed "to provide a more dynamic interaction between respondent and questionnaire" than is possible in a paper-and-pencil survey. Web questionnaires can incorporate extensive and difficult skip patterns, pop-up instructions for individual questions, drop-down boxes with lists of answer choices, feedback on possibly incorrect answers (e.g., birth date "1839"), pop-up word definition screens, and automatic fill-ins for later answers. They can use a great variety of shapes and colors and can add pictures, animation, video clips, and sound (Dillman, 2007:458). When designed carefully, Web survey options and features may be used to motivate and assist respondents and otherwise substitute for the role that an interviewer plays (Couper, Traugott, and Lamias, 2001; Manfreda and Vehovar, 2008:276–81).

At this point, the great practical advantages and enormous design potential of Web surveys for social research are offset by some major weaknesses. Perhaps the greatest of these is coverage error. This error derives from two related problems: the proportion of the general population who are Internet users and the lack of a sampling frame to sample users. By 2007, 71 percent of U.S. households were using the Internet at home or elsewhere. However, a "digital divide" remains, with nonusers being more likely to be black or Hispanic, poorly educated, and older and with less income than those with Internet access (National Telecommunications and Information Administration [NTIA], 2007). The second problem, the absence of a good frame for sampling Internet users, is currently handled by limiting professional Web surveys to special populations having membership lists and Web access, such as college students or employees of an organization (Couper, 2000). Another approach is to use telephone RDD or some other mode to generate a probability sam-

ple and then persuade sampled individuals to complete an online questionnaire (Fricker et al., 2005). In the absence of a well-defined probability sampling plan, consumers should be wary of most Web polls and surveys as they are likely to entail self-selected samples that merely reflect the views of those who choose to respond.

Nonresponse error poses another problem, as studies comparing response rates for mail, telephone, and Web surveys generally have found the lowest response rates for the Internet mode (Fricker et al., 2005; de Leeuw, 2008:129). Furthermore, not only is the response rate low (one meta-analysis of fifty-six Internet surveys found an average response rate of 35 percent [Cook, Heath, and Thompson, 2000]), but Internet surveys face problems similar to those underlying declining telephone response rates, including privacy issues, spam and “phishing” scams, and the proliferation of nonprofessional Web surveys.

#### KEY POINT

Compared to mail questionnaires, Web surveys reduce costs, require less time, may employ complex interactive questionnaires, but reach more limited populations and have lower response rates.

On the other hand, studies have shown that Web response rates comparable to mail rates can be achieved with special populations (college students with free e-mail access) and motivational efforts (an advance cover letter explaining the study's purpose) (Kaplowitz, Hadlock, and Levine, 2004). Also, Web respondents are less likely than

mail or telephone respondents to leave specific questions unanswered and tend to write longer answers to open-ended questions (Kwak and Radler, 2002; Kaplowitz, Hadlock, and Levine 2004).

Despite these coverage and nonresponse problems, Web surveys have developed so rapidly in recent years that some have argued that they eventually will replace mail and interview survey modes. Solutions to some of the problems are being explored, but it will take some time before the advantages of the Internet can be harnessed for use in large-scale, national surveys.<sup>5</sup>

### Mixed-Mode Surveys

Choosing a data-collection mode is difficult when none of the primary modes seems optimum for the intended research. An alternative solution is to design a **mixed-mode survey**, which uses more than one mode, either sequentially or concurrently, to sample and/or collect the data (Lavrakas, 1993). In this way the weaknesses of one mode may be offset by the strengths of another mode. For example, relatively inexpensive phone surveys may be used to screen and locate specialized populations, such as people with a rare disease, for a study requiring expensive FTF interviews. Mixed-mode designs are increasing in popularity as a way to decrease costs, shorten data-collection time, and increase response rates (Groves et al., 2004).

Different modes may be used sequentially for sampling, recruiting, screening, collecting data, and following up on nonrespondents. To increase response rates in their Web survey, Michael Kaplowitz, Timothy Hadlock, and Ralph Levine (2004) sent out an advance recruitment letter by U.S. mail. In another study mentioned

earlier, telephone RDD sampling was used to screen for Internet access for the purpose of recruiting respondents for the Web survey (Fricker et al., 2005). In IVR surveys, a telephone interviewer may initially contact and recruit the respondent and then switch the respondent to the self-administered IVR mode (Steeh, 2008). For the Current Population Survey, in which respondents remain in the sample for several months, initially expensive FTF interviewing is used to recruit and motivate respondents in order to obtain a high response rate; then less expensive telephone interviewing is employed in subsequent months. Follow-ups on reluctant respondents, in the form of simple reminders or a longer persuasive message, may involve a different mode (e-mail, telephone, in-person) than the main data-collection mode (de Leeuw, Hox, and Dillman, 2008).

A mode shift to self-administered questionnaires—paper-and-pencil or CASI—often is used in FTF surveys to increase privacy in the collection of sensitive information. Typically, an interviewer administers the largest part of the interview but then provides the respondent with either a paper questionnaire to be sealed in an envelope or a CASI laptop to complete the self-administered portion that requests the most sensitive information. In this way, respondents are less susceptible to social desirability biases, which are more likely when questions are administered by an interviewer. An early application of this strategy occurred in the 1992 National Health Interview Survey–Youth Risk Behavior Supplement (NHIS-YRBS), which used an audio questionnaire to collect sensitive information about drug use, sexual intercourse, cigarette smoking, and other unhealthy behaviors from adolescents (Willard and Schoenborn, 1995). Teens listened to the questions on a Walkman-type portable audio headset and recorded their answers on an answer sheet that did not contain any information by which parents or other household members would know the questions being answered. Besides providing privacy to the young respondents, this data-collection mode was more effective than a written questionnaire with younger teens and those with poor reading skills. In addition, the pre-recorded questions provided greater standardization (wording, tone of voice, etc.) than when questions are asked by interviewers.

A concurrent or respondent-specific design uses different modes to survey different respondents. One solution to the coverage problem in Web surveys, for example, is a respondent-specific approach whereby those without Web access are surveyed in-person or by mail (Dillman, 2007). Another concurrent example would be increasing the response rate of the sampled population by conducting telephone or FTF interviews with those who did not respond to an initial mail questionnaire (Biemer and Lyberg, 2003). Respondent-specific designs may start with a relatively inexpensive mail survey, then follow up on the nonresponses with telephone interviews, and finally attempt to reach the remaining nonrespondents with expensive FTF interviews (de Leeuw, Hox, and Dillman, 2008). Similar respondent-specific designs occur in multiple frame sampling, such as separate but overlapping frames for mobile and landline phones (Groves et al., 2004:87–90). While concurrent designs may increase the response rate and keep costs down, a major weakness is the uncertain comparability of respondents surveyed by different modes. As the various modes may differ in coverage, sampling, nonresponse, and measurement

errors, merging the mode subsamples to statistically estimate the target population is a difficult and uncertain undertaking.

## Field Administration

Once the planning is completed, fieldwork can begin. This phase of the research begins with the recruitment and training of interviewers, continues with the field interviews, and concludes when follow-up efforts have been completed on initially unresponsive persons in the sample. Figure 9.3 is a flowchart that summarizes the various aspects of a survey's fieldwork phase.<sup>6</sup> Although we focus on FTF interviewing, the discussion includes some aspects of telephone surveys and other modes.

### Interviewer Selection

Although there are no universally agreed-on criteria for the selection of interviewers, experience and common sense suggest that certain qualities are desirable. These would include articulateness; a pleasant personality that inspires cooperation and trust; a neat, businesslike appearance; freedom from prejudices or stereotypes to-

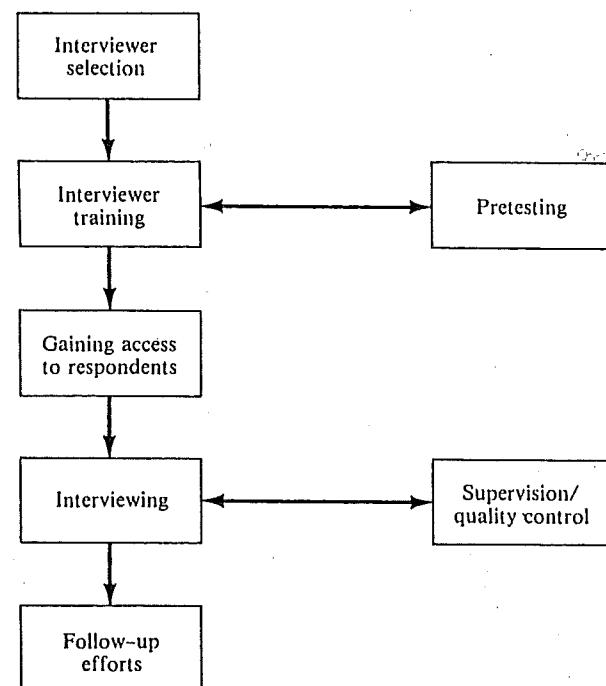


FIGURE 9.3. Flowchart illustrating a survey's fieldwork phase.

ward the population being interviewed; interest in the survey topic; familiarity with computers; and the ability to listen, use neutral probes when needed, and record responses accurately. We suspect that the presence of these qualities is evaluated to some degree during the selection and training processes. This should create reasonably well-qualified interview staffs and would explain why research has found no consistent correlates between interviewer characteristics and the quality of interviewing (Fowler, 1991; Groves et al., 2004).

Unless the survey is being done by a large research organization that has a permanent staff, the researcher must recruit all the interviewers for a given survey. The process of recruiting interviewers is basically the same as hiring for any job (Weinberg, 1983). That is, positions are advertised and applicants are screened and selected. Beyond minimum reading and writing skills, availability and readiness to meet job requirements are the principal selection criteria. Other interviewer attributes are largely dependent on market forces. The majority of interviewers, according to data from a sample of interviewers at U.S. government statistical agencies, do not regard their job as a primary source of income or career, perhaps because of the intermittent nature of the work (Groves and Couper, 1998:198). This work feature as well as job requisites also may account for the composition of the interviewer workforce, which is predominantly female and young to middle-aged, with above-average education. Since such attributes are largely beyond the control of the researcher, it is fortunate that they appear to be much less important in determining the quality of a survey than the interviewer's ability achieved through careful training and experience.

### Interviewer Training

Interviewers receive training in general interviewing skills and techniques as well as in specific procedures required for a particular survey project. In practice, these two aspects of training often are combined, with survey-specific materials (e.g., questionnaire or sampling procedures) used for practical application (Weinberg, 1983; Lessler, Eyeman, and Wang, 2008). More specifically, the training process must accomplish several goals:

1. Provide interviewers with information regarding the study's general purpose, sponsor, sampling plan, and uses or publication plans.
2. Train interviewers in locating households and eligible respondents as prescribed by the sampling design.
3. Teach basic interviewing techniques and rules, such as how to gain respondents' cooperation, establish rapport without becoming overly friendly, ask questions and probe in a manner that will not bias the response, deal with interruptions and digressions, and so forth.
4. Acquaint interviewers with the survey instrument and instructions for CAPI or CATI.
5. Provide demonstrations and supervised practice with the interview schedule.

6. Weed out those trainees who do not possess the motivation and ability to do an acceptable job.

Studies of the effects of training on interviewer errors have shown that too little training (less than a day) is inadequate and that better-quality data are associated with more extensive training (Groves et al., 2004:294). A training program for FTF interviewing may combine home study with a series of classroom sessions. The first session might begin with a general introduction to the study, followed by a presentation and instruction in basic interviewing skills and responsibilities. In the second session, the researcher would thoroughly familiarize interviewers with the survey questionnaire, going over the entire instrument item by item, explaining the importance of each item, and giving instructions for recording responses and examples of problematic responses and ways to deal with them. Next, the researcher would conduct a demonstration interview and then divide interviewers into pairs for supervised practice interviewing. Third and subsequent sessions would involve further practice, possibly including field experience, and further evaluation. Experienced interviewers generally receive survey-specific training through home study of the project's special interviewing procedures and survey instrument followed by discussions or mock interviews with a field manager. For an excellent discussion of what trainees should expect in telephone survey training, see Patricia Gwartney's *The Telephone Interviewer's Handbook* (2007).

### **Pretesting**

A pretest consists of trying out the survey instrument on a small sample of persons having characteristics similar to those of the target group of respondents. The basic reason for conducting a pretest is to determine whether the instrument serves the purposes for which it was designed or whether further revision is needed. **Pretesting** may be carried out before, at the same time as, or after the interviewers are trained. An advantage to completing the pretest before interviewer training is that the final instrument may be used during training. An advantage to delaying pretesting is that the interviewers can assist with this step, either during the field practice part of interviewer training or after the formal training is completed. A disadvantage to delaying pretesting is that there may be a time gap between the completion of training and the start of the "real" interviews while the instrument is being revised. The subject of pretesting is discussed in more depth in chapter 10.

### **Gaining Access**

Gaining access to respondents involves three steps: getting "official" permission or endorsement when needed or useful, mailing a cover letter introducing the study to persons or households in the sample, and securing the cooperation of the respondent.

When doing a community interview survey, it is usually a good idea to write a letter to or visit a local official to describe the general purpose of the study, its importance, the organization sponsoring it, the uses to which the data will be put, the

time frame, and so forth. In addition, endorsement letters from relevant local organizations may be sought, such as the county medical society if doctors will be interviewed, or the chamber of commerce if businesses are being sampled.

Respondent cooperation also will be enhanced by a good **cover letter**. In interview surveys, the cover letter is usually mailed a few days before the interviewer is to call on the respondent. In surveys using mailed questionnaires, the cover letter is sent with the questionnaire either as a separate sheet or attached to the questionnaire. Dillman (2007) also recommends mailing a brief prenotice letter just prior to a questionnaire mailing. Occasionally, cover letters are mailed out in advance of phone surveys when mailing addresses are known. In an analysis of twenty-nine research studies, sending a letter in advance improved telephone response rates by an average of about 8 percent (de Leeuw et al., 2007). The objective of the cover letter, to persuade the respondent to cooperate with the survey, may be met by (1) identifying the researcher and survey sponsor and the phone number of a contact person, (2) communicating the general purpose and importance of the study, (3) showing how the findings may benefit the individual or others (e.g., the results will be used to improve health care or to increase understanding of marriage relationships), (4) explaining how the sample was drawn and the importance of each respondent's cooperation to the study, (5) assuring individuals that they will not be identified and that their responses will be kept confidential and will be combined with those of others for data-analysis purposes, (6) explaining that the questionnaire will take only a few minutes to fill out or that the interview will be enjoyable and will be held at the respondent's convenience, and (7) promising to send respondents a summary of the study's findings. (See Box 9.2 for an example of a cover letter.) Also, response rates may be increased by including a token incentive, such as \$1 or \$2, with the cover letter (Dillman, 2007).

A third step in interview surveys is gaining the cooperation of the respondent. Interviewers must contact or reach the sample person and then persuade him or her to cooperate by completing the survey. Making contact with designated households is primarily a matter of persistence and overcoming barriers. Interviewers vary the time of attempted contacts and may make repeated call-backs, typically at least six for FTF surveys and ten or more for telephone surveys (Fowler, 1993). We will discuss follow-up efforts further as a final phase of fieldwork.

Avoiding refusals is a more difficult problem that requires special interviewer skills. According to Robert Groves and Mick Couper's (1996, 1998) theory of survey participation, which applies mainly to in-person interviews, in the initial moments of the survey encounter, the sample person is actively trying to comprehend the purpose of the interviewer's visit. He or she uses cues from the words, behavior, and physical appearance of the interviewer to arrive at an explanation (or identify a "script") and then evaluates the costs of continuing the conversation. Whether the person eventually agrees or refuses to participate depends on the interviewer's ability to quickly and accurately judge the particular script reflected in the householder's initial response and to react accordingly.

The theory is consistent with analyses of interviewer-householder interactions. For example, Robert Groves, Robert Cialdini, and Mick Couper (1992) found that experienced interviewers use two related strategies to convince respondents to

**BOX 9.2 Sample Cover Letter for Household Survey, with Key Features Identified in Left-Hand Margin**



Washington State University  
Social and Economic Sciences Research Center

Wilson Hall 133  
PO Box 444014  
Pullman, WA 99164-4014  
509-335-1611  
FAX 509-335-0110

- Date** → July 8, 1999
- Inside address** → L. T. Hansen  
2121 Lincoln Way East  
Uniontown, WA 99962-2056
- The request** → I am writing to ask your help in a study of new residents being conducted for the state of Washington. This study is part of an effort to learn what draws people to the state, and whether they are happy or unhappy with what they find here.
- Why you were selected** → It's my understanding that you may have moved to Washington state sometime in the last few years. We are contacting a random sample of new residents from every county in the state to ask why they moved, what their employment experience has been, and whether services are meeting their needs.
- Usefulness of survey** → Results from the survey will be used to help state and local government make Washington a better place for new residents like you. By understanding what people want when they move here, public officials can do a better job providing services and improving the state's quality of life. And by knowing more about the job skills of new residents, public agencies and private businesses can help make the most of what new residents contribute to the state's economy.
- Confidentiality** → Your answers are completely confidential and will be released only as summaries in which no individual's answers can be identified. When you return your completed questionnaire, your name will be deleted from the mailing list and never connected to your answers in any way. This survey is voluntary. However, you can help us very much by taking a few minutes to share your experiences and opinions about Washington state. If for some reason you prefer not to respond, please let us know by returning the blank questionnaire in the enclosed stamped envelope.
- Token of appreciation** → We have enclosed a small token of appreciation as a way of saying thanks for your help.
- Willingness to answer questions** → If you have any questions or comments about this study, we would be happy to talk with you. Our toll-free number is 1-800-833-0867, or you can write to us at the address on the letterhead.
- Thank-you** → Thank you very much for helping with this important study.
- Real signature** → Sincerely,  
  
Don A. Dillman  
Professor and Deputy Director
- P.S. If by some chance we made a mistake and you have not moved to Washington (or back to Washington after living somewhere else) since January 1990, please answer only the first question in the questionnaire and return the rest of it blank. Many thanks.

Source: Page 162 of Dillman (2007). Copyright 2007 by John Wiley and Sons, Inc. Used by permission.

for tailoring the conversation to present the most effective arguments. In a subsequent analysis, Groves and Couper (1996) concluded that the most effective interviewer adaptations were those that addressed the real concerns of the householder. Moreover, experimental tests of a training program to help interviewers anticipate and address such concerns were shown to increase cooperation (Groves and McGonagle, 2001).

In contrast, tailoring is difficult in telephone interviews due to the absence of visual cues and inadequate time—often a few seconds—to maintain interaction and build up rapport (for an exploratory study, see Couper and Groves, 2002). Nevertheless, phone interviewers are trained to instantly make and maintain a positive impression upon the respondent through their voice (tone, pitch, speed, and enunciation) and pre-memorized scripts for introducing the survey and addressing respondents' reluctance to participate (e.g., "I'm not good [at] answering surveys," "I'm too busy," "You are invading my privacy") or other concerns (Gwartney, 2007).

### Interviewing

Although we noted earlier that interviews may be more or less structured, nearly all surveys, especially large-scale surveys, use structured or "standardized" interviewing. The goal of standardization is to expose each respondent to the same interview experience so that any differences in recorded answers are due to differences among respondents "rather than differences in the process that produced the answer" (Fowler and Mangione, 1990:14). To standardize the measurement process, survey instruments consist almost entirely of closed-ended questions that are presented in the same order for all respondents, and interviewers are instructed to carefully follow prescribed procedures. Ideally, interviewers should be perfectly consistent, neutral intermediaries of the survey researcher. In short, "the goal [of standardized interviewing] is nothing less than the elimination of the interviewer as a source of measurement error" (Groves, 1989:358).

Specific practices vary somewhat from one survey organization to another, but the following four rules of standardized interviewing, according to Floyd Fowler (1991:264; also see Fowler and Mangione, 1990), are given almost universally to interviewers:

1. Read the questions exactly as written.
2. If a respondent does not answer a question fully, use nondirective follow-up probes to elicit a better answer. Standard probes include repeating the question, asking "Anything else?" "Tell me more," and "How do you mean that?"
3. Record answers to questions without interpretation or editing. When a question is open-ended, this means recording the answer verbatim.
4. Maintain a professional, neutral relationship with the respondent. Do not give personal information, express opinions about the subject matter of the interview, or give feedback that implies a judgment about the content of an answer.

participate. First, they tailor their approach to the sample unit, adjusting their dress, mannerisms, language, and arguments according to their observations of the neighborhood, housing unit, and immediate reactions of the householder. Second, they maintain interaction, which maximizes the possibility of identifying relevant cues

(Further, more detailed interviewer guidelines from an FTF community survey are provided in Box 9.3.) For a description of standardized interviewing in CATI surveys, see Gwartney (2007:172–233).

Considerable empirical evidence and sound theoretical arguments justify standardization principles (Fowler, 1991; Fowler and Mangione, 1990). For example, if questions are not asked as worded, one cannot know what question was posed; and numerous experiments have shown that small changes in the wording of questions can alter the distribution of answers (Schuman and Presser, 1981). Also, experiments have demonstrated that suggestive questioning (presenting only a subset of answer alternatives that are presumed to be relevant) and suggestive probing can affect response distributions and relationships with other variables (Smit, Dijkstra, and van der Zouwen, 1997).

While there is little doubt that standardization reduces the interviewer's contribution to measurement error, a more serious, long-standing concern is that pre-

### BOX 9.3 Guidelines for Survey Interviewers

Organizations sponsoring surveys, such as the National Opinion Research Center, which oversees the GSS, usually produce an interviewers' manual that explains rules and procedures to follow to minimize the effects of error. The rules and guidelines outlined below have been adapted from an interviewers' manual used in a community survey conducted in a medium-sized city; however, they apply to any well-conducted personal interview survey.

#### *Rules*

Three basic rules are suggested.

1. *Courtesy, tact, and acceptance.* It is of utmost importance that your manner be at all times courteous, tactful, and nonjudgmental. Under no circumstances are you to argue or debate anything that is said. The primary function of the interviewer is to learn what the respondent believes about the items on the schedule without judging or influencing that response in any way. No matter what the respondent says, the interviewer should accept it without showing surprise, approval, or disapproval. Respondents will be less likely to share confidential data about their personal lives, for example, with an interviewer who appears to disapprove of them.

2. *Dress.* Three things should be kept in mind in selecting clothing for interviewing: the expectations that surround the role of the interviewer, the persons being interviewed and their probable response to one's dress, and comfort. Within the limits of good taste, the individual interviewer should make adjustments for the neighborhood in which he or she is working.

3. *Confidentiality.* Under no circumstances is the interviewer to give out to anyone except the supervisor any information gathered in the course of interviews. Incidents that occur or any information gained while interviewing are strictly confidential and should not be discussed with anyone who is not part of the project. Project work-

ers should never discuss interview data, even among themselves, in a public place where the conversation could be overheard.

#### *Procedures in Conducting an Interview*

The following procedures have been found to be helpful in conducting interviews and obtaining accurate, honest responses. In many cases, deviating from these procedures will influence the respondent and introduce bias.

1. *Initiate the interview.* When the respondent comes to the door, introduce yourself by name. Show identification. Explain briefly what the study is about and whom it is you wish to interview. Be prepared to answer briefly questions regarding who is sponsoring the study and how or why the respondent was chosen. Letters of endorsement and newspaper clippings may be presented.

Proper groundwork by the researcher and a positive approach on the part of the interviewer will help to minimize the problem of refusals. In addition, the interviewer can frequently overcome an initial refusal by listening to the respondent's concerns and then addressing those concerns. For example, is the respondent "too busy"? Stress the brief and enjoyable nature of the interview or, when necessary, make an appointment to call back. Does the respondent appear suspicious about the uses to which the data will be put? If so, explain again the study's purpose, provide an assurance of confidentiality, and explain that the data will be combined so that no individual's responses can be identified or linked to that person. Of course, firm refusals must be respected.

2. *Put the respondent at ease.* This is a major part of successful interviewing. A conversational, convivial attitude may help put the respondent at ease. Try to be relaxed and "natural."

3. *Be businesslike.* While it should be relaxed, an interview should not be long-winded. Remember that you and the respondent are busy people. If the respondent strays far afield from the point of a question, politely pull him or her back on the track.

4. *Keep the interview situation as private as possible.* If you are in a room with other people, do not let your attention wander to other parts of the room. Direct your questions to the respondent, and maintain eye contact. This will help both you and the respondent focus on the task.

5. *Avoid stereotyping.* Do not try to "categorize" the respondent as your preconceptions may interfere with your objectivity and may influence the respondent. You can also help to prevent the respondent from stereotyping you by not identifying yourself with any particular group or ideology.

6. *Be thoroughly familiar with the survey instrument.* Know the instrument so well that you can look at the respondent while asking questions.

7. *Ask every question in its proper sequence and exactly as written.* The interview schedule should have been carefully constructed. Questions are in the order presented because it is easier for interviewers to ask them in this order, because there is a logical flow of the topics, because the order helps respondents think through or recall material, or for similar reasons. Remember that very slight changes in the wording of items have been shown to affect the results.

8. *Do not assume the answer to any question.* A respondent may imply the answer to a question in answering a previous question but may respond differently when asked the question formally.

9. Speak slowly in a clearly understood, well-modulated voice. If respondents are to give reliable answers, they must understand the questions.

10. Do not put answers in the respondent's mouth. This is one of the most common mistakes of interviewers. If a respondent seems unsure of an answer, pause and then repeat the question exactly as worded. Do not suggest an answer or a series of answers. Not all interviewers would suggest the same responses, and therefore, respondents would not be choosing from the same suggestions. This would result in biased data.

11. Use an appropriate, neutral probe when needed. Probing is used when the initial response is incomplete, ambiguous, or irrelevant. A variety of probes are possible, but they must be neutral; that is, they must stimulate a more valid response without suggesting an answer. Sometimes, as suggested in point 10, pausing or repeating the question may be sufficient to motivate the respondent to add to or clarify the response. At other times, a neutral question such as one of the following may be needed: Is there anything else? Can you tell me more? In what ways? The survey instrument designer may include on the interview schedule certain probes to use with particular items when needed.

12. Record responses as you go along. Do not try to recreate the interview later. Before leaving the residence, skim the instrument to be sure all questions have been answered. Later, check to be sure responses will be understandable to the coders. Add notes in parentheses if necessary.

senting a standard stimulus in and of itself can produce measurement error. According to this view, standardized interviewing stifles interviewer–respondent communication in two ways: (1) it inhibits the ability to establish rapport, which motivates respondents to cooperate and give complete and accurate answers, and (2) it ignores the detection and correction of communication problems (Beatty, 1995; also see Maynard and Schaeffer, 2002).

Regarding motivation, some respondents feel irritated by the unilateral nature of a structured survey. They cannot converse with the researcher or interviewer, they cannot qualify or expand answers, and they may be forced to choose among alternative answers that they find unsatisfactory. Fowler (1995:99–102) believes that respondents' resistance to standardized survey interviewing can be overcome by initially orienting or training them to play the role of a respondent. Since interviewer–respondent interaction in a highly structured survey is quite different from everyday conversation, respondents should be given an explanation of the rationale and the rules of standardized surveys to prepare and motivate them for the interview task.

Standardization also may reduce validity if respondents' misinterpretations of questions are ignored or uncorrected. In a widely cited article, Lucy Suchman and Brigitte Jordan (1990) argued that standardization suppresses elements of ordinary conversation that are crucial for establishing the relevance and meaning of questions. Interviewers who are trained to read questions as written and to discourage elaboration are not prepared to listen carefully for misunderstandings and correct them. From videotapes of standardized interviews, Suchman and Jordan gave sev-

eral examples of miscommunication, such as an interviewer failing to correct a respondent who interprets "alcoholic beverages" to include hard liquor but exclude wine, which led to invalid responses. Such problems could be resolved, they claimed, if interviewers were granted the freedom and responsibility to negotiate the intended meaning of questions through ordinary conversational conventions.

While acknowledging the communication problems identified by Suchman and Jordan, advocates of standardized interviewing tend to disagree with them about causes and remedies. Some advocates contend that problems arise chiefly because of poorly worded questions and that whether respondents interpret questions consistently and accurately depends on adequate question pretesting (Kovar and Royston, 1990). Detractors of standardization maintain that efforts to improve question wording are rarely, if ever, sufficient. Researchers need interviewers to help ensure consistent interpretations, but this is unlikely unless interviewers are given a freer hand in communicating with respondents than standardization allows.

Paul Beatty (1995) notes that advocates on both sides of this debate tend to take extreme positions that seem incompatible. He argues, following Nora Cate Schaeffer (1991), that while the problems of standardization should not be overlooked, it

#### KEY POINT

Although standardized interviewing is intended to reduce interviewer error, it may dampen respondents' interest and inhibit interviewers from clarifying questions and correcting respondents' misinterpretations.

also is foolish to ignore its merits. The real issue for him is how researchers can solve communication problems while harnessing the full benefits of standardization. One means is the development of better questions; another may involve adapting the role of the interviewer. For certain types of questions, such as requests for factual information, permitting interviewers to stray from standardized scripts to correct respondents' misunderstandings may improve data quality; but allowing interviewers too much flexibility will greatly increase interviewer error as well as interview length (Groves et al., 2004).

The standardization controversy aside, a serious weakness of surveys, alluded to earlier, is their reactivity. This is especially problematic in interviews, where biases may be produced not only by the wording, order, and format of the questions but also by the interaction between interviewer and respondent. Like the subject in an experiment, the respondent's chief concern often is with gaining the interviewer's social approval, or at least with avoiding his or her disapproval (Phillips, 1971). And even though the interviewer's main goal is to obtain valid responses, interviewers may unknowingly affect responses in myriad ways. In survey research, the general term for such effects, or measurement errors, is **response effects**. More precisely, a response effect is "the amount of the error in response to a question that is associated with a particular factor" (Sudman and Bradburn, 1974). Chapter 10, on survey instrument design, suggests ways in which to minimize response effects related to the interview schedule (or questionnaire), such as question wording or position. Here, we focus on effects due to the interaction of interviewer and respondent.

One source of response effects is the interviewer's physical characteristics. For example, the race of the interviewer has been shown to have a considerable impact on certain types of responses. In general, blacks express fewer antiwhite sentiments to white than to black interviewers and whites give fewer antiblack answers to black than to white interviewers (Schuman and Kalton, 1985). A 1989 pre-election survey also found a significant race-of-interviewer effect on expressed voter preferences for a black versus white gubernatorial candidate (Finkel, Guterbock, and Borg, 1991). Respondents in the state of Virginia were more likely to support the black candidate, L. Douglas Wilder (subsequently elected as the nation's first black governor), when the interviewer was black and were more likely to favor the white candidate, J. Marshall Coleman, when the interviewer was white.

Less is known about whether other demographic variables such as age, gender, and class have the same effect as race, although this seems likely when a specific interviewer trait is particularly relevant. For instance, Emily Kane and Laura Macaulay (1993) found gender-of-interviewer effects on several items measuring gender-role attitudes, with both male and female respondents tending to give more egalitarian responses to female than to male interviewers.

In a fashion similar to experimenter effects, interviewers also may inadvertently communicate their expectations to respondents about how they should respond. To illustrate, if an interviewer believes a respondent to be of limited intelligence and inarticulate, he or she may expect shorter, less articulate responses and may communicate this indirectly by short pauses. Since the respondent is looking to the interviewer for clues to the appropriateness of his or her behavior, he or she will likely provide short responses, thus fulfilling the interviewer's expectations.

On the other side, the respondent's reports to an interviewer may easily be distorted by such things as poor memory, desire to impress the interviewer, dislike for the interviewer, or embarrassment. Similarly, a respondent's feelings about the topic of the study or toward the organization sponsoring it may also affect the quality of data obtained. Finally, settings for interviews may present problems. A housewife who is being interviewed while supervising children may not be able to focus on the tasks of the interview sufficiently to provide as full and accurate responses as she might in another situation. In the British Household Survey, the presence of a spouse during an interview led to greater agreement between husbands and wives on several attitudinal and behavioral items (Zipp and Toth, 2002). A study by Nancy Brener and associates (2006) showed that high school students were more likely to report drinking alcohol, smoking marijuana, sexual activity, and other risk behaviors when questioned at school than when questioned at home.

### ***Supervision and Quality Control***

Once the interviewing phase of the survey begins, the researcher or an interviewer supervisor oversees all aspects of the data collection. The supervisor's role involves three interrelated sets of activities: managing the work of the interviewers, monitoring their performance, and administering quality-control procedures. In their management role, supervisors provide materials, collect completed survey records,

pay for work done, make themselves available to answer questions and provide help, identify and resolve problems, hold regular meetings with interviewers, and review and give feedback on the interviewers' work.

To detect problems and review interviewers' progress, supervisors monitor their performance. In virtually all surveys, part of performance monitoring involves careful record keeping and evaluation of completed interviews. Records kept on the number of hours worked, amount paid, number of eligible contacts, and number of refusals provide critical information on interviewer productivity, survey costs (e.g., time and dollars per interview), and survey quality (response and refusal rates). By reviewing completed surveys, supervisors also can screen uncoded responses and check to see if interviewers are following instructions and recording answers appropriately. To tell if interviewers ask questions exactly as written, use neutral probes, and record answers correctly, many survey organizations also monitor or observe the interview process. In FTF interview surveys, supervisors may observe interviewers by either accompanying them on interviews or digitally recording random interview segments on the CAPI laptops. In telephone interview surveys, special monitoring phones allow supervisors to listen to a sample of ongoing interviews without being heard by either the interviewer or the respondent.

Reviewing completed questionnaires and observing interviews are two mechanisms of controlling the quality of interviewers' work and thereby determining the quality of the data. Two other processes are retrieving missing data and validating interviews (Weinberg, 1983). When data are missing, particularly for factual items that are critical to the survey, respondents may be recontacted to retrieve the information. Also, it is standard in all well-conducted surveys to verify that the interview actually took place and that the interviewer did not fabricate some responses or deviate from prescribed procedures (see Groves et al., 2004:366–67). Validation usually is done by reinterviewing a sample of respondents for each interviewer. Respondents may be asked a subset of items, and in the case of in-person surveys, the reinterviews often are carried out by telephone.

Finally, another important aspect of the supervisory role is sustaining interviewers' motivation and enthusiasm and preventing feelings of isolation. It is easy to assume that interviewers will do better work as they gain experience; however, studies have shown that the opposite may be the case. For example, validity studies of reports of hospitalizations and physician visits found that the more interviews an interviewer had done, the greater the problem of underreporting (Cannell, Marquis, and Laurent, 1977). Thus, supervisors should meet with interviewers regularly throughout the interviewing period not only to provide continuing feedback and reinforcement of skills but also to communicate the importance of good interviewing. GSS interviewers report, mostly via e-mail, to their field supervisors at least weekly, but usually more often, to go over the status of their assignments (T. W. Smith, personal communication, 2003).

### ***Follow-Up Efforts***

The final phase of fieldwork consists of following up on nonrespondents in an attempt to gain their cooperation. If no contact is made initially, then GSS interview-

ers are instructed "to try several follow-up procedures, such as (a) calling back at different times and days of the week, (b) asking neighbors when people are usually at home or how they might be contacted, (c) leaving notes, and (d) using a reverse directory to get a telephone number for the household" (Davis and Smith, 1992:51). Follow-up efforts help to raise response rates.

Besides follow-ups, response rates may be improved (1) by appropriate efforts to gain access, discussed earlier; (2) in interview surveys, by proper interviewer training and supervision; and (3) for mailed questionnaires, by inclusion of a stamped return envelope and a token prepayment, as well as by attention to the length, difficulty, and appearance of the questionnaire (see chapter 10).

The particular follow-up activities depend on the survey mode. For telephone and FTF surveys, the major problem is dealing with refusals. In many surveys, more experienced interviewers or supervisors are used to try to gain the respondent's cooperation on the second try. The National Health and Social Life Survey (NHSLS) dealt with some initial refusals by employing follow-up interviewers who specialize in converting reluctant respondents and by sending out "conversion letters" that answered special concerns that potential respondents were raising (Laumann et al., 1994). In response to a clear refusal, however, one follow-up call or visit should be the limit to avoid respondent feelings of harassment.

Because the NHSLS investigators were worried about reaching their target response rate (75 percent), they started offering incentive completion fees to reluctant respondents on a selective basis in low-response areas (Laumann et al., 1994:56–57). Although incentive fees ranging from \$10 to \$100 were offered for interviews, the fees were viewed as cost-efficient given the high cost of interviewer wages and travel costs (the cost of completed interviews averaged \$450). The fees and other techniques for converting reluctant respondents eventually paid off with a final 79 percent response rate.

Since response rates are typically lower for mailed questionnaires, follow-up efforts are especially important with this mode. Typically, three follow-up mailings are used. For example, in Henry Wechsler and associates' (1994) initial college alcohol survey, respondents received four separate mailings, approximately 10 days apart: the initial mailing of the questionnaire, a postcard thanking those who had completed the questionnaire and urging those who had not to do so, a mailing with another copy of the questionnaire again appealing for its return, and a second reminder postcard. Given that this survey was truly anonymous, all persons in the sample had to be sent the subsequent mailings. If questionnaires have been coded so that the researcher knows who has responded, there can be savings in postage and paper as only nonrespondents need receive the follow-up mailings.<sup>7</sup>

In the Tailored Design Method, Don Dillman (2007) recommends three more widely spaced follow-ups, each of which entails new and more persuasive appeals. The first follow-up, sent out about 2 weeks after the original mailing, consists of a postcard thank you/reminder as in the Wechsler et al. alcohol survey. The second follow-up, mailed 2 weeks later, is sent only to nonrespondents and contains a replacement questionnaire; the cover letter for this mailing communicates that the respondent is receiving individual attention, emphasizes their importance for the suc-

cess of the survey, and notes that others have responded as a means of encouraging response. The third follow-up, mailed 4 weeks later, invokes special procedures, such as special mail or contact by telephone, to emphasize the importance of the respondent's cooperation. When Dillman (2007:185) used certified mail for this final request in five statewide surveys of the general public, he raised the response rates from an average of 59 to 72 percent. These follow-up efforts should yield adequate response rates, but pushing rates higher is

very expensive and difficult; therefore, additional follow-up mailings are seldom used.

These follow-up efforts are intended to raise response rates. As we note in chapter 6, however, the common wisdom of pursuing high survey response rates in order to decrease nonresponse bias is being challenged (Groves, 2006). Instead, concern should focus on nonresponse bias for key variables within a survey. From this perspective, aggressive efforts to increase survey response rates may actually increase nonresponse bias in individual variables. If men are overrepresented among nonrespondents in a telephone survey, for example, additional call-backs may actually increase the gender bias by reaching and interviewing more women than men. Groves (2006:666) cites another example of an election exit poll experiment in which an incentive did increase response rates but also increased nonresponse bias by attracting more Democratic than Republican voters.

## Summary

Social surveys have three common features: a relatively large number of respondents generally chosen by some form of probability sampling, formal observation procedures involving interviews and/or questionnaires, and computerized statistical analysis of data. The units of analysis in sample surveys are typically individuals selected from a single community or nation; however, surveys of social units other than individuals are also done and samples from different countries are sometimes combined, as in comparative studies. While standardization is the norm, data-collection procedures vary along a continuum from highly formal and structured to less formal and structured, the choice of which depends on the researcher's objectives.

Surveys are used extensively for both scientific and nonscientific purposes, largely because of their ability to describe large populations in terms of a broad range of characteristics, attitudes, and behavior. Relative to other research strategies, surveys generally are more flexible in that they can be used with equal facility for both descriptive and explanatory research and can address a wider range of research topics. On the other hand, surveys present greater problems for inferring causal relationships, are less easily altered once the study has begun, are limited to reports of behavior rather than observations of behavior, and, like laboratory experiments, are subject to reactive effects.

Survey research designs are divided into two broad categories: (1) cross-sectional studies, in which data are gathered from a sample of a community or larger grouping at essentially one point in time, and (2) longitudinal studies, in which data are gathered at two or more points in time from either the same sample of respondents (panel study) or independently selected samples of the same population (trend studies). Two variants of the cross-sectional design have been developed to study social structural influences: (1) contextual studies, in which different groups are sampled to examine the effects of the group or context on individuals, and (2) social network studies, in which every individual in a group is studied to delineate networks of personal relationships. Cohort studies, a special instance of trend studies, trace changes across cohorts to separate the effects of age, cohort or generation, and historical period.

Of the three steps in conducting a survey, planning and field administration were considered in this chapter. Data processing and analysis are covered in chapter 15. The initial stages in planning a survey depart from other research strategies primarily in the construction of the survey instrument (interview schedule or questionnaire), which we discuss at length in the next chapter, and the development of the sampling plan. Survey data may be conducted through personal FTF interviews, telephone interviews, self-administered questionnaires, various computer methods, or a combination of these modes (mixed-mode survey).

Interviews, in comparison with questionnaires, offer greater flexibility in the type and format of questions, provide the opportunity to clarify questions and elicit fuller responses, and tend to have higher response rates. FTF interviews also permit lengthy interviews with complex questions, although they may introduce greater bias through interviewer-respondent interaction, and they tend to be very expensive. By comparison, telephone interviews offer a substantial savings in time and money and are easier to administer, although they require simpler questions and may elicit less complete responses. Mailed self-administered questionnaires are the least expensive survey mode, but they tend to have lower response rates, result in more nonresponse to individual questions, and must contain simpler questions for reliable responses. New electronic self-administered methods, which are rapidly gaining acceptance, may reduce time and cost and allow for more varied and user-friendly questionnaire design; however, Internet surveys, in particular, suffer from coverage error, and their most productive use in the short run may be as a complement to interviews in mixed-mode surveys.

Field administration in survey research begins with interviewer selection and training and instrument pretesting, continues with interviews and staff supervision, and ends with follow-up efforts to reach initial nonrespondents. Interviewers must be able simultaneously to put the interviewee at ease, to ask questions and guide the interview to its completion, and to record responses. To accomplish this requires careful interviewer selection and training. Before data are gathered from respondents, the survey instrument should be pretested and access to respondents should be gained by contacting appropriate community officials and mailing a cover letter. During the course of interviewing, interviewers must (1) be wary of introducing bias; (2) be courteous and tactful, dress appropriately, and guard the confidentiality of respondents' replies; (3) be relaxed but businesslike; and (4) be thoroughly fa-

miliar with the instrument so as to maintain rapport, use probes when necessary, and record answers as they go along. Supervision of interviewees—keeping a record of each interviewer, checking completed interview schedules, and so forth—is essential for quality control, while follow-up efforts to contact initial nonrespondents are essential to obtain an adequate response rate, especially in mail surveys.

## Key Terms

cross-national surveys  
structured interview  
semistructured interview  
unstructured interview  
descriptive and explanatory surveys  
secondary analysis  
survey research designs  
cross-sectional design  
contextual design  
social network design  
longitudinal design  
trend study  
panel study  
cohort study

interview schedule  
response rate  
face-to-face interview  
computer-assisted personal interviewing  
telephone interview  
computer-assisted telephone interviewing  
random-digit dialing  
paper-and-pencil questionnaire  
computer-assisted self-interviewing  
mixed-mode survey  
pretesting  
cover letter  
response effect

## Exercises

1. In December 1994 Henry Wechsler and associates reported the results of their first college alcohol survey in the prestigious *Journal of the American Medical Association* (vol. 272, pp. 1672–77). This study raised awareness of the extent and harmful consequences of binge drinking and stimulated a great deal of research on drinking among college students. Read the article and answer the following questions.
  - a. What was the researchers' target population in this study?
  - b. Fully describe the sampling design.
  - c. To draw the sample, the researchers used two sampling frames. What were they?
  - d. Which school—a 20,000-student public university or a private liberal arts college with 2000 students—had a better chance of being selected for the study? Explain.
  - e. This study consisted of a mailed questionnaire survey. What principal advantages of this survey mode are likely to have prompted its use?
  - f. What inducement was offered to encourage students to return their questionnaires?
  - g. What follow-up procedures were used to increase the response rate?
  - h. How did the researchers attempt to rule out biases due to nonresponse?
2. A week before the 2008 presidential election, with Barack Obama comfortably ahead of John McCain in most polls, an article in the *Washington Post* asked, "Could

the polls be wrong?" The article claimed that there was an undercurrent of concern among polling professionals, in part because of the wide variation in Obama leads: from 2 to 15 percentage points. Suppose someone asks you how there can be so much variation in poll results. Assuming that all the polls were telephone surveys that used probability sampling, identify and briefly describe four possible sources of error.

3. Suppose that you are planning to conduct a campus survey of volunteerism. Besides accurately documenting the extent of formal volunteering, you intend to describe the types of volunteer service that students perform and the relationship between volunteering and background factors such as gender, religion, and academic major. From a "total error perspective," discuss the relative advantages and disadvantages of doing an FTF survey, telephone survey, mail questionnaire survey, and Web-based survey.

### Notes

1. Social network analysis is an interdisciplinary paradigm that developed primarily from earlier work in American sociology and social psychology (sociometry) and British anthropology (Wasserman and Faust, 1994).

2. Whereas the flowchart in Figure 9.1 shows the top-to-bottom and sideways influences, in reality later decisions may result in modifications of earlier decisions. For example, sampling-plan cost considerations may bring about a change in the research objectives.

3. However, a recent mail survey development, sampling addresses from the U.S. Postal Service Delivery Sequence File (DSF), shows promise as a feasible alternative or multi-mode complement to RDD telephone surveys (Link et al., 2008).

4. See Shostack and Fairweather (1979) for a study comparing physician response rates to mailed and personal interview surveys.

5. One approach is to employ probability-based, Web-enabled panels designed to be representative of the U.S. or another target population. In a series of mode experiments, Tom W. Smith and J. Michael Dennis (2004) compared responses to questions asked in the 2000 and 2002 GSS in-person surveys with responses to identically worded questions from the Web-based Knowledge Networks panel. Knowledge Networks' panel members are initially recruited by list-assisted RDD telephone sampling, and then households that agree to participate are given the equipment necessary to access the Internet. While some of the differences between the in-person and Web surveys were small, they concluded that "much more information is needed on differences in measurement across modes before results from Web-based surveys and surveys using other modes can be considered generally comparable" (p. 10).

6. Fieldwork is not always carried out exactly in this order. Pretesting, for example, may occur before, during, or after the training of interviewers. Efforts to gain access to respondents may well begin earlier than depicted.

7. Another strategy is to enclose a postage-paid postcard that identifies the respondent (name or identification number) in the first mailing (Mangione, 1995). Respondents are asked to mail the postcard separately from the anonymous questionnaire, and they are told that the postcard indicates that the questionnaire has been returned and they will receive no reminder notices.

# 10



## Survey Instrumentation

The two most critical features for successful survey research are the sample and the survey instrument. Accurate generalizations about populations of interest depend on the quality of the sample. But no matter how carefully the sample is selected, a sample survey is only as good as the design of the questionnaire or interview schedule.

Survey instrument design is a creative process, partly art and partly science (Payne, 1951). Like an artist, the survey designer selects "raw materials" and combines them creatively within certain principles of design. For the artist, raw materials may consist of paper or canvas; pencil, charcoal, chalks, watercolors, acrylics, or oils; brushes; and so forth. The survey designer's raw materials are such things as free response and fixed-choice questions, direct and indirect questions, question-and-response formats, overall physical layout, and instructions.

However, the survey designer is unlike the artist in at least one important way: An artist is mainly concerned with expressing his or her own personal ideas, emotions, or other subjective experience, whereas the designer of a survey instrument must be concerned ultimately with getting reliable and valid reports about other people. This concern is what makes the survey designer a scientist rather than an artist. The reports may be of subjective experiences, such as values, opinions, fears, and beliefs; or they may be of overt experiences, such as job history, salary, political behavior, place of residence, consumer behavior, or leisure activities. The verbal reports obtained in surveys are, of course, answers to questions; survey instrumentation is thus the science of asking questions.

In this chapter we focus on standardized instruments, in which the wording and order of questions are the same for all respondents. Yet many of the principles discussed here apply to less structured interview approaches. Also, we generally do not treat separately the primary modes of asking questions (face-to-face interviews, telephone interviews, paper-and-pencil questionnaires, and computer-assisted self-interviews), although these may require different questions and formats. We do consider mixed-mode instrument designs, in which different respondents are surveyed by different data-collection modes.

We begin by discussing cognitive and communication theories of how people answer survey questions. Then we examine "materials" available to the survey designer—types of questions and response formats. Next we consider the overall design or "sketch" of the survey instrument. Then we look at practical guides for "filling in the sketch"—writing items to arrive at a first draft. Finally, we discuss the pretesting operations that are necessary to transform a draft instrument into a finished product.

## The Survey as a Social Occasion

Until recently, the science of asking questions was based largely on two sources of information: the experience of professional survey researchers and empirical evidence from experiments on question wording. For example, experience has shown that it is best for question crafters to use simple, plain-spoken language (Converse and Presser, 1986); research has shown that respondents are more likely to give a "no opinion" response if it is explicitly offered as a response alternative (Schuman and Presser, 1981). Such wisdom has resulted in some important guidelines for questionnaire design, but these guidelines are ad hoc and often not clearly understood. What has been needed is a solid theoretical understanding of how people go about answering survey questions. Fortunately, over the past two decades survey methodologists have drawn heavily on work in cognitive and social psychology (for detailed discussions, see Schwarz, 1996; Sudman, Bradburn, and Schwarz, 1996; Sirken et al., 1999; Tourangeau, Rips, and Rasinski, 2000) to develop theoretical models of the response process.

Scientifically, the survey interview is an occasion to obtain valid responses to a series of questions; however, it also is a social occasion subject to the influences of the social world. Being interviewed is an uncommon, sometimes anxiety-provoking experience for many respondents. Unfamiliar with the canons of structured interviewing, they turn to the social and linguistic rules governing everyday conversation for guidance on how they should behave as respondents. In answering each question, they also attend to a sequence of cognitive tasks with certain levels of ability and motivation. What can each of these perspectives tell us about designing a survey instrument?

From the perspective of cognitive processing, answering survey questions requires that respondents (1) comprehend the question, (2) retrieve the information requested from memory, (3) formulate a response in accord with the question and the information retrieved, and (4) communicate a response deemed appropriate (Sudman, Bradburn, and Schwarz, 1996:58–75; see also Tourangeau, Rips, and Rasinski, 2000; Groves et al., 2004:202–207). Each of these steps involves mental tasks that can give rise to response errors. For example, to comprehend a question according to what the researcher had in mind, the respondent must understand both its literal and intended meanings. Therefore, the question-answer task breaks down not only when the question wording is vague but also when the purpose of the question is misunderstood. Consider the question, "What have you done today?" While the literal meaning of the words may be clear, this may not be sufficient to answer the question because the respondent still needs to know what sort of activities are of interest to the researcher. Would this include, for example, mundane activities such as brushing one's teeth and taking a shower (Schwarz, Groves, and Schuman, 1998)? Analyses of the other cognitive steps in the response process indicate that response errors are likely to occur when there is insufficient time to access relevant information from memory, when the accessed information does not fit the response options provided in the question, and when the respondent modifies the information to project a favorable image to the interviewer. We discuss ways to minimize these problems later in this chapter.

Another perspective, conversational analysis, suggests that respondents make sense out of structured interviews and self-administered questionnaires by relying

upon tacit assumptions underlying ordinary conversations. This perspective draws on the work of philosopher of language Paul Grice (1989). According to Grice, conversations generally run smoothly because speakers and listeners follow four maxims:

1. Speakers should not say things that they believe to be false. [Truthfulness]
2. Speakers should make comments that are relevant to the purposes of the conversation. [Relevance]
3. Speakers should make their contributions as informative as possible and not repeat themselves. [Nonredundancy]
4. Speakers should express themselves as clearly as possible. [Clarity]

As respondents follow these conversational maxims to infer the intended meaning of questions, they may be unintentionally influenced by subtle aspects of the survey instrument and administration, such as prior questions, response formats, and other cues irrelevant to the intended purpose of the current question. The principle of relevance, for example, implies that respondents will see everything about the survey as relevant to a given question, including seemingly irrelevant design features such as the numeric values on a rating scale. Norbert Schwarz and associates (1991) demonstrated this by asking respondents to answer the question, "How successful would you say you have been in life?" on an 11-point scale with endpoints labeled "not at all successful" and "extremely successful." The answers, as it turned out, depended on the numeric values assigned to these endpoints.

When the endpoints ranged from 0 (not at all successful) to 10 (extremely successful), 34 percent endorsed a value from 0 to 5, whereas when the endpoints ranged from -5 (not at all successful) to +5 (extremely successful), only 13 percent endorsed a formally equivalent value of -5 to 0. Follow-up questions revealed that the numeric values affected how respondents interpreted the endpoints.

When accompanied by a value of 0, they interpreted "not at all successful" to mean the absence of success, but when accompanied by a value of -5, they interpreted this to mean explicit failure.

Similarly, the experimental inclusion of a picture of either a healthy woman exercising or a sick woman in a hospital bed in a Web survey affected respondents' health self-assessment ratings. Those seeing the picture of the fit woman reported themselves to be in poorer health than those exposed to the picture of the sick woman (Couper, Conrad, and Tourangeau, 2007). Thus, the pictures were relevant to respondents' interpretation of the intended meaning of a health self-rating question.

Another study shows how the methodological strategy of exploring complex topics by asking a series of related questions runs counter to the conversational principle of nonredundancy. Fritz Strack, Norbert Schwarz, and Michaela Wänke (1991) asked German students to answer two questions about their lives: One question

### KEY POINT

Seemingly innocuous features of a survey instrument, such as numerical labeling of rating scales, may serve as cues by which respondents interpret the meaning of questions.

asked them to rate their happiness, and another asked them to rate their satisfaction with life as a whole. Both questions had 11-point rating scales with similar endpoints (11 = very happy and 11 = very satisfied). "Happiness" and "life satisfaction" are nearly identical concepts. Not surprisingly, when both questions were asked in ostensibly unrelated questionnaires, the two measures were highly correlated; however, when the questions were introduced in the same questionnaire with the sentence "Now, we have two questions about your life" and then were asked one after the other, the correlation was much lower. Apparently, asking both questions in the same conversational context implied that they were intended to convey a different meaning; after all, asking the same question twice would not make sense.

Finally, a third theoretical perspective, rational choice theory, focuses on how respondents' *motivation* to provide acceptable answers to survey questions can affect their responses (Krosnick, 1991, 1999). Answering survey questions requires a great deal of mental effort. At the least respondents must go through the four cognitive steps outlined above. Of course, some questions, such as those that require memory recall or thoughtful opinions on controversial issues, require more work than others; and lengthy interviews require more effort than shorter ones. When respondents exert the maximum effort to generate a complete and unbiased answer, they are said to be "optimizing"; when they do not expend the necessary effort but take shortcuts, such as not searching their memories thoroughly, they are said to be "satisficing." Satisficing may explain why some respondents tend to give "no opinion" or "don't know" responses; it requires less effort to express no opinion than to think through one's position. On attitude questions that ask respondents whether they agree or disagree with a series of statements, satisficing also may account for the tendency to agree more often than disagree because it is easier to politely agree with reasonable assertions than to generate reasons why they may not be true.

Theoretically, the likelihood of satisficing behavior increases with question difficulty and decreases with the respondent's motivation and ability to answer the question. One study (Holbrook, Green, and Krosnick, 2003) compared evidence of satisficing responses in telephone versus face-to-face interview surveys. As evidence of their lesser motivation to provide optimal answers, telephone respondents were judged to be less cooperative and engaged in the interview and expressed more dissatisfaction with the length of the interview. Consistent with the theory, they also evidenced more satisficing behavior. For example, they were more likely to give "no opinion" answers on several attitudinal items and were more likely to express agreement on yes/no and agree/disagree questions.

Although these theories are in an early stage of development, their methodological implications are clear. To identify likely problems when designing survey instruments, researchers need to understand the conversational, cognitive, and motivational processes that underlie question asking and answering (Sudman, Bradburn, and Schwarz, 1996:258). As we examine the options available to the question designer, we will consider these perspectives on the response task. Given the complexity of this task, however, decisions about questionnaire design are best worked out in a thorough process of pretesting. Indeed, one of the most important products of this theoretical movement is the development of rigorous diagnostic procedures for pretesting survey instruments, which we discuss at the end of the chapter.

## Materials Available to the Survey Designer

Like the artist, the survey designer has a number of choices about raw materials. We now offer some guides regarding choices among certain types of question and response formats and regarding the use of visual aids and questions from previous research.

### Open-Ended and Closed-Ended Questions

A major choice among "materials" concerns open-ended and closed-ended questions. The **open-ended** (also called "free-response") question requires respondents to answer in their own words (in written form on a questionnaire or aloud to an interviewer). The **closed-ended** (or "fixed-choice") question requires the respondent to choose a response from those provided. Here are examples of two questions written in both open and closed forms.

1. How would you rate the president's performance in office so far? (OPEN)
1. How would you rate the president's performance in office so far? (CLOSED)
  - ( ) Poor
  - ( ) Below average
  - ( ) Average
  - ( ) Above average
  - ( ) Excellent
2. What do you think is the number one domestic issue the president should be concerned with? (OPEN)
2. Which one of the domestic issues listed below should the president be most concerned with? (CLOSED)
  - ( ) Unemployment
  - ( ) Inflation
  - ( ) Reform of the welfare system
  - ( ) Balancing the budget
  - ( ) The energy crisis
  - ( ) Other

The choice between open- and closed-ended questions is a complex one because each has a number of advantages and disadvantages. The greatest advantage of the open question is the freedom the respondent has in answering. The resulting material may be a veritable gold mine of information, revealing respondents' logic or thought processes, the amount of information they possess, and the strength of their opinions or feelings. Frequently, the researcher's understanding of the topic is clarified and even completely changed by unexpected responses to open questions. But, alas, this very quality of open questions, the wealth of information, has a drawback: the "coding" problem of summarizing and analyzing rich and varied (and often irrelevant and vague) responses. Coding such material is a time-consuming and costly process that invariably results in some degree of error (Sudman and Bradburn, 1982). (See chapter 15 for a further discussion of coding.)

Open questions also require interviewers skilled in "recognizing ambiguities of response and in probing and drawing respondents out . . . to make sure they give codable answers" (Sudman and Bradburn, 1982:151). The following interviewer-respondent exchange illustrates the importance of probing to give respondents encouragement and time to think and to clarify their responses (Sudman and Bradburn, 1982:150):

- Interviewer: What are the most important problems facing the nation today?
- Respondent: I don't know, there are so many.
- Interviewer: That's right, I'd just like to know what you think are the most important problems.
- Respondent: Well, there's certainly inflation, and then government spending.
- Interviewer: Government spending . . . how do you mean? Could you explain that a little? What do you have in mind when you say "government spending"?
- Respondent: There's no end to it. We have to cut down federal spending somehow.
- Interviewer: Any others?
- Respondent: No. I think those are the most important ones.

Other problems with the open question include (1) the varying length of responses (some people are unbelievably verbose; others, exceedingly reticent), (2) the difficulty with inarticulate or semiliterate respondents, (3) the difficulty interviewers have in getting it all down accurately, and (4) the reluctance of many persons to reveal detailed information or socially unacceptable opinions or behavior. Open-ended questions also entail more work, not only for the researcher but also for the respondent. Indeed, open questions should be used sparingly, if at all, in self-administered questionnaires or Web surveys, where respondents must write or type rather than speak.

Closed-ended questions are easier on the respondent because they require less effort and less facility with words. The presence of response options also enhances standardization by creating the same frame of reference for all respondents. When used in an interview, closed questions require less work and training to administer, and the interview may be shortened considerably.

On the other hand, good closed questions are difficult to develop. It is easy to omit important responses, leading respondents to choose among alternatives that do not correspond to their true feelings or attitudes. Research shows that respondents tend to confine themselves to the alternatives offered, even if they are explicitly given a choice such as "Other \_\_\_\_ (please explain)" (Schuman and Presser, 1981). This would be consistent with Grice's relevance principle because respondents are likely to view the list of response options as indicative of the researcher's interests. To provide a list of response options that are meaningful to the respondent, the recommended procedure is to use open questions in preliminary interviews or pretests to determine what members of the study population say spontaneously; this information then may be used to construct meaningful closed

alternatives for the final instrument. Unfortunately, this procedure is not always followed; time and financial limitations may prevent pretesting of sufficient scope to yield adequate information on the population's responses.

Even when the range of possible responses is known, they may be too numerous to list in a closed question, especially in a telephone interview. It is better, obviously, to ask the open question, "In what state or foreign country were you born?" than to list all states and foreign countries. Similarly, questions about occupations, medical conditions, favorite television shows, and the like, are best asked open-ended (Fowler, 1995).

Given these advantages and disadvantages, when should one choose open or closed questions? Robert Kahn and Charles Cannell (1957) suggest these five considerations:

#### KEY POINT

Compared to open-ended questions, effective closed-ended questions are easier on the respondent but harder on the survey designer.

(1) the objectives of the survey, (2) the level of information possessed by respondents in regard to the topic, (3) how well respondents' opinions are thought out or structured, (4) motivation of respondents to communicate, and (5) the extent of the researcher's knowledge of respondents' characteristics.

1. Consider first the study's objectives. If you simply want to classify respondents with respect to some well-understood attitude or behavior, the closed question would probably be appropriate and most efficient. However, the open question is usually preferable when the survey objectives are broader and you are seeking such information as the basis on which opinions are founded, the depth of respondent knowledge, or the intensity with which respondents hold opinions.

2. A second consideration is the amount of information that respondents already have on the topics of interest. If you believe that the vast majority will have sufficient information regarding the survey's topics, the closed question may be acceptable. On the other hand, if you are uncertain as to the level of information of the respondents or if you anticipate a wide range in the amount of knowledge, the open question is more appropriate. With closed questions, uninformed respondents may conceal their ignorance by making arbitrary choices, yielding invalid reports. And even adding the response option "don't know" may not resolve the problem since this option is unlikely to be popular with respondents who are sensitive about appearing ill-informed.<sup>1</sup> It is easier, for example, to respond "approve" or "disapprove" of the Supreme Court decision *Roe v. Wade* than to admit not knowing what it is.

3. A related consideration is the structuring of respondent thought or opinion. Are respondents likely to have thought about the issue before? Can they take a position or express a definite attitude? If respondents are likely to have given previous thought to the matter and the range of typical responses is known to the researcher, the closed question may be satisfactory. This might be the case, for example, with a survey designed to measure the attitudes of registered voters toward the performance of the president. However, if respondents' ideas are less likely to be structured, open questions may be preferable. Suppose you wanted to ascertain why college students

chose XYZ University or why couples desired a certain number of children; for such questions, the reasons may be numerous and not always immediately accessible to respondents. A series of open questions would allow respondents time to recall, think through, and talk about various aspects of their decisions, rather than hastily selecting a possibly incomplete or inappropriate response provided by a closed question.

4. Motivation of respondents to communicate their experiences and thoughts is a further consideration. In general, the open question will be successful only when the respondent is highly motivated because this question type is more demanding in terms of effort, time, and self-disclosure. Therefore, with less motivated respondents, closed questions may lead to better-quality data. On the other hand, closed questions sometimes dampen respondent motivation, in that some people prefer to express their views in their own words and find being forced to choose among limited fixed-choice responses very irritating.

5. A fifth important consideration in choosing between open and closed questions is the extent of the researcher's previous knowledge of respondent characteristics. That is, how well does the researcher understand the vocabulary and amount of information possessed by the respondents, the degree of structure of respondents' views, and their level of motivation? Unless the researcher has done similar studies previously or has done extensive preliminary interviewing, the most likely answer is "not very well." If this is the case, open questions should yield more valid (albeit more difficult to summarize and analyze) data.<sup>2</sup>

One approach is to use the different types of questions at different stages of research, first utilizing open questions in preliminary interviewing and using the information provided by these early interviews to develop closed questions or a combination of open and closed questions on the final instrument. Or you may decide from the start upon some combination of open and closed questions appropriate to your purposes. (For a further comparison of open and closed questions, see Box 10.1.)

### ***Direct and Indirect Questions***

Another choice of "materials" concerns the use of direct and indirect questions. A **direct question** is one in which there is a direct, clear relationship between the question that is asked and what the researcher wants to know. The bulk of questions used in survey research are direct. "What is your total family income?" and "What do you think is the ideal number of children for your family?" are examples of direct questions.

With **indirect questions**, the link between the researcher's objectives and the question asked is less obvious. An investigator interested in studying the sex-role attitudes of male factory workers, for example, might ask the indirect question "Do you believe your co-workers would mind having a woman as supervisor?" instead of the direct question "Would you mind having a woman as supervisor?" Although the investigator really wants to determine the respondents' own sex-role attitudes, he or she may suspect that they will be unwilling to admit personal sexist sentiments

### **BOX 10.1 An Experimental Comparison of Open and Closed Questions**

Discussions of the advantages and disadvantages of open and closed questions are based largely on common sense and the unsystematic experiences of survey researchers. One of the few exceptions is Howard Schuman and associates' (Schuman and Presser, 1981; Schuman, Ludwig, and Krosnick, 1986; Schuman and Scott, 1987) use of an experimental design within large-scale sample surveys to compare systematically responses elicited by parallel open and closed questions. The table in this box shows the outcome of one of these experiments (Schuman and Scott, 1987), in which half of the respondents in a national telephone survey were randomly assigned to an open version and half were assigned to a closed version of the same question about the most important problem facing the country.

The open question has been used repeatedly in national surveys. This particular closed version listed four problems, each of which had been mentioned rarely in recent uses of the open question. As expected, only 2.4 percent of the sample mentioned any of these four problems in response to the open question. The categories most frequently mentioned were "unemployment" and "general economic problems." In response to the closed question, however, 60 percent of the respondents chose among the alternatives offered to them, in spite of the fact that the question explicitly instructed them to consider other alternatives. Furthermore, unemployment, mentioned by 17 percent of respondents on the open version, was named by only 6.2 percent of respondents on the closed version. As Schuman and Scott (1987:957) point out, "the issues mentioned on the open question give the better overall picture of American concerns and . . . the findings on the closed question are distorted by the constraint or inertia produced by listing the four problems as part of the question."

<i>Open question</i>	<i>Closed question</i>
What do you think is the most important problem facing this country today?	Which of the following do you think is the most important problem facing this country today—the energy shortage, the quality of public schools, legalized abortion, or pollution—or if you prefer, you may name a different problem as most important?
1. The energy shortage	0.0%
2. The quality of public schools	1.2
3. Legalized abortion	0.0
4. Pollution	1.2
Subtotal	2.4
5. All other responses	93.0
6. Don't know	4.7
	100.0%
Number of respondents	(171)
	(178)

*Source:* Reprinted with permission from Schuman and Scott, 1987. Copyright 1987 American Association for the Advancement of Science.

This experiment clearly demonstrates how differences in the form of a question can produce quite different results. Although these data suggest the superiority of the open version, other evidence indicates that open and closed questions will ordinarily lead to the same conclusion, so long as the closed form includes as alternatives the most frequently mentioned issues in response to the open form (Schuman and Presser, 1981). When in doubt about response alternatives, investigators should be careful not to impose their own frames of reference on respondents but should begin with open questions and then use the "free" responses to construct a set of "fixed" alternatives. In addition, open questions may be necessary as "why" follow-ups to closed questions, when alternatives are too complex or too many to present easily in a closed version, and when rapidly changing events undermine the adequacy of closed alternatives (Schuman and Presser, 1981).

to an interviewer. Knowing, however, that individuals' attitudes and beliefs shape the way they perceive the world about them, the researcher assumes that respondents will impute their own attitudes to their co-workers.

Indirect questions may be appropriate when the researcher is interested in characteristics or experiences that the respondent is unwilling or unable to reveal in direct terms. Respondents may be unwilling because the behavior in question is considered socially undesirable or unacceptable; they may be unable because the characteristics—motives, needs, fears—are below their level of conscious awareness.

Most indirect measurement is based on the notion of projection: that individuals tend to attribute their own inner needs, feelings, opinions, and values to the outer world. Thus, an individual presented with ambiguous stimuli will tend to interpret the material in ways that reflect his or her own needs and values. Some of the more common projective techniques are word association, sentence completion, and storytelling. In word association, respondents say the first thing that comes to mind in response to each of a list of words read by an interviewer. Sentence completion requires respondents to finish incomplete sentences, such as "When I think of cities, I think of \_\_\_\_\_. Storytelling involves asking respondents to interpret such ambiguous stimuli as inkblots and pictures.<sup>3</sup>

Most indirect measures were developed for clinical use and designed to aid in the diagnosis of emotional disorders by revealing an individual's personality and needs. This remains their principal use, although they also have been used in attitude and motivation research, especially in marketing. To illustrate the indirect approach, let us examine a classic study from marketing research, conducted in 1950. (For a more complete discussion of the use of indirect questions and approaches, see Kidder and Campbell, 1970.)

Mason Haire (1950) was interested in attitudes toward instant coffee, one of the early instant-food preparations. Using direct questioning, Haire found that when women who reported not using instant coffee were asked "What do you dislike about it?" most of them said that they did not like the flavor. Suspecting, however, that this was a stereotypic response that concealed other motives, Haire developed an indirect

approach to measure consumers' attitudes. He prepared two grocery shopping lists that were identical, except that one contained "Nescafé instant coffee" and the other contained "1 lb Maxwell House coffee (drip ground)." He then asked subjects to read one or the other shopping list and to describe the personality and character of the woman who made it out. As it turned out, Nescafé and Maxwell House coffee users were perceived quite differently. For example, nearly half of the subjects described the Nescafé user as lazy and failing to plan household purchases, whereas the Maxwell House purchaser was rarely described in these terms. The evidence from the indirect approach thus suggested that the decision to buy instant coffee was influenced as much by prevailing attitudes about what constitutes good housekeeping as by the flavor of instant coffee. Although a 1968 replication of this study produced essentially the same results (Webster and von Pechmann, 1970), it would be interesting to see if the findings would be repeated today, given the apparent changes in attitudes toward gender roles and housework in the past three decades.

While one should be aware of the possibility of indirect questioning, it is used relatively infrequently in survey research today. One likely reason is that its use often requires intensive training in the administration, scoring, and interpretation of responses. Some researchers also question the validity of such measures. On the infrequent occasions that tests of validity and reliability have been done, the results have not been encouraging (Kidder and Campbell, 1970). Furthermore, there are some serious ethical concerns with the use of indirect questions. To what extent are respondents giving their "informed consent"? To what extent is deception being used, and is its use justifiable?

### **Response Formats**

In addition to making decisions about the broad categories of open and closed questions and direct and indirect questions, the creative survey designer will consider the possibilities offered by various **response formats** for closed-ended questions. The simplest response option is a simple "yes" or "no." This would be appropriate for such questions as "Do you belong to a labor union?" and "Have you ever been threatened with a gun?" However, even though many types of information form natural dichotomies, this kind of question appears less frequently than you might think. With many apparently dichotomous items, respondents may prefer to answer "don't know," "uncertain," or "both." About 10 percent of General Social Survey (GSS) national samples, for instance, respond "don't know" to the question "Do you believe in life after death?" (Davis, Smith, and Marsden, 2007).

The decision to encourage or discourage "don't know," "no opinion," or "refuse" responses is a complex one, depending upon the question subject and if the mode is interviewer-assisted or self-administered (Fowler, 1995:164–65; Dillman, 2008). When respondents are truly uninformed or lack firsthand experiences with the subject of the question, the researcher needs to learn this by explicitly providing "don't know" options and, in some circumstances, including a statement that sanctions a know-nothing response. Otherwise, respondents may haphazardly pick a response to conform to normative expectations that they should answer the question (Grice's

relevance principle). When measuring feelings or subjective states, on the other hand, respondents who are ambivalent, indifferent, or even lazy are likely to choose an explicit “don’t know” option rather than attempt to express their feelings (satisficing strategy). In such situations, it may be more appropriate to omit “don’t know” response categories, thus limiting no opinions to volunteered responses. In interviewer surveys, “don’t know” options may be omitted from the question and left to the interviewer’s discretion. If they are omitted from the set of responses in a Web survey, the respondent should be allowed to skip the question and not be forced to choose an answer to continue the survey.<sup>4</sup>

Ordinal response scales are commonly used to measure the strength or intensity of respondents’ feelings. One of the most popular formats, the Likert response scale used in Likert scaling (see chapter 13), consists of a series of responses ranging from “strongly agree” to “strongly disagree.” This is a common way of measuring attitudes. The respondent is presented with a statement and asked to indicate the extent of his or her agreement. For example,

Men have greater sexual needs than women.

- ( ) Strongly agree
- ( ) Agree
- ( ) Uncertain
- ( ) Disagree
- ( ) Strongly disagree

Besides degrees of agreement, a variety of other rating scales exist for assessing attitudes and opinions. To evaluate objects ranging from consumer products to personal attributes to government policies, respondents could be given the categories “excellent,” “very good,” “good,” “fair,” and “poor,” as in the following question.

Compared with the jobs that your friends have, would you say that your job is excellent, very good, good, fair, or poor?

- ( ) Excellent
- ( ) Very good
- ( ) Good
- ( ) Fair
- ( ) Poor

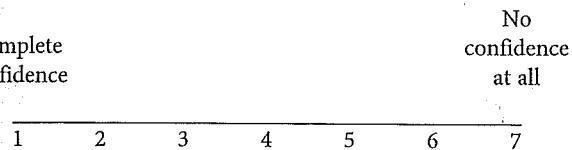
One can also create scales by asking respondents “how” they feel and then using a series of adverbs to modify the intensity of opinion. The following item from the GSS measures job satisfaction in this way.

On the whole, how satisfied are you with the work you do—would you say you are very satisfied, moderately satisfied, a little dissatisfied, or very dissatisfied?

- ( ) Very satisfied
- ( ) Moderately satisfied
- ( ) A little dissatisfied
- ( ) Very dissatisfied

It is also common to use numerical rating scales, with verbal ratings provided for the numerical endpoints, as in the GSS item below (Davis, Smith, and Marsden, 2007).

I am going to name some institutions in this country. Some people have complete confidence in the *people running* these institutions. Suppose these people are at one end of the scale at point number 1. Other people have no confidence at all in the *people running* these institutions. Suppose these people are at the other end, at point 7. And, of course, other people have opinions somewhere in between at point 2, 3, 4, 5, or 6. Where would you place yourself on this scale for banks and financial institutions?



Another popular rating approach measures people’s reactions to stimulus words or statements on a 7-point scale with endpoints anchored by opposing adjectives (e.g., “good” and “bad,” “fast” and “slow”).<sup>5</sup> Respondents are asked to rate the stimulus word or statement as it relates to each pair of adjectives. For example,

President Obama									
Fun	:	:	:	x	:	:	:	:	Boring
Real	:	:	:	:	:	x	:	:	Fake
Strong	x	:	:	:	:	:	:	:	Weak
Cold	:	:	:	x	:	:	:	:	Warm

Here, a hypothetical respondent evaluated President Obama to be slightly fun and fake and extremely strong and was neutral or undecided on the cold-warm dimension.

A particular advantage with this format is that it is easily adapted for people from different backgrounds by using words appropriate to their vocabulary. Robert Gordon and colleagues (1963) adopted this format to test predictions from several delinquency theories regarding the values of gang, nongang lower-class, and non-gang middle-class boys. With use of such adjective pairs as “clean—dirty,” “good—bad,” “brave—cowardly,” and “smart—sucker,” the boys studied were asked to rate various stimulus descriptions (e.g., “someone who works for good grades at school,” “someone who shares his money with his friends,” “someone who knows where to sell what he steals,” “someone who stays cool and keeps to himself”). The gang boys were found to endorse middle-class norms to a much greater extent than predicted.

In creating a rating scale, two important considerations are the number of categories to use and whether the scale should be defined primarily by numerical or adjectival labeling (Fowler and Cosenza, 2008). Should the response format comprise two categories (e.g., “yes,” “no”), five categories (e.g., Likert), seven categories

(e.g., the preceding two examples), or even more? Unless respondents' views are limited to a few categories, seven categories seem best for measuring the full range of their attitudes, beliefs, or feelings. An exception to the principle of using longer scales to obtain more precise information is the telephone survey, in which respondents may not retain all of the scaling instructions because, unlike other survey modes, they cannot be handed a visual aid or questionnaire portraying the scale. When scales are administered by telephone, it is best to keep them short and provide a description that is easy to remember.

Adjectival labeling of each response category conveys less ambiguity to the reader than numerical values alone or numerical values with verbal labeling of endpoints (Fowler, and Cosenza, 2008). Suppose that respondents are asked to rate the honesty of a particular politician, either on scale A from 0 ("Not at all honest") to 6 ("Extremely honest") or on scale B from -3 ("Not at all honest") to +3 ("Extremely honest"). Although both scales have seven categories and the same end-point labels, as we discussed earlier, respondents will actively use the numerical information in determining an appropriate answer (Schwarz et al., 1991). Thus, they are likely to interpret scale B as bipolar, with categories ranging from degrees of dishonesty (negative values) to degrees of honesty (positive values). In contrast, scale A may be construed as unipolar categories representing degrees of honesty ranging from its absence (0) to its highest presence (6). Also, the "0" category of version B better conveys a middle or neutral position than the somewhat ambiguous "3" of version A. If the intended scale was conceptualized as bipolar, the following adjectival labeling format might be used instead of numerical values:

Extremely dishonest	Moderately dishonest	Slightly dishonest	Neither dishonest	Slightly honest	Moderately honest	Extremely honest
			nor honest			

Ranking questions present another possibility to the survey designer, as in this example.

A number of factors influence people's career choices. Rank the factors listed below from 1 to 5, according to their importance to you in choosing a career:

- Making a lot of money
- Being creative
- Being free from supervision
- Having opportunities for advancement
- Avoiding a high-pressure job

Although ranking questions frequently are handy and appropriate to use, we have chosen this example to illustrate several potential problems. First, unless pretesting with an open-ended question has been done, who can say these five factors are the most salient to the respondents? You can probably think of other important factors, especially if you

are interested in working with people and helping others. (This, of course, is a general problem with closed questions.) Even if these factors have been shown to be the most important ones, further questions of validity may be raised. Do people really *know* why they choose certain careers? If so, will they tell you? How many respondents would freely admit that, to them, "making a lot of money" is the most important factor? Would "being creative" tend to get a spuriously high ranking? (For several additional examples of ranking questions, see Sudman and Bradburn, 1982:158–65.)

### Visual and Media Aids

An effective survey instrument also may include miscellaneous aids such as illustrations, photographs, films, and cards that contain written material. The use of illustrations in a self-administered questionnaire may improve clarity and appeal, increasing respondent motivation and perhaps improving the completion rate. In interviewing, photographs and films sometimes are used to acquire data not obtainable through questioning alone. In a study of early socialization practices, for example, photographs of mothers with their young babies might be taken to stimulate the mothers to discuss their child-rearing attitudes and behavior, especially if the mothers have difficulty answering highly abstract questions.

Other aids used in face-to-face interviewing may speed up the interview or make it easier for the respondent to answer accurately. The latter purpose may be served by presenting respondents a card containing the responses to a fixed-choice question so that the respondent can view the choices while answering the question. As a rule, cards are used whenever the response categories are lengthy or more than five or six in number. For example, GSS interviewers hand respondents a card similar to the one shown here when they ask, "Which of the categories on this card come closest to the type of place you were living in when you were 16 years old?" (Davis, Smith, and Marsden, 2007).

1. In open country but not on a farm
  2. On a farm
  3. In a small city or town (under 50,000)
  4. In a medium-size city (50,000–250,000)
  5. In a suburb near a large city
  6. In a large city (over 250,000)

Compared with other survey modes, Web surveys permit richer visual and media presentation, including pop-up menus, sounds, animation, slider-bars, automatic skipping over nonrelevant questions, and so forth. The limited empirical research on Web design indicates that potential Web features are not well understood and may have unintentional outcomes. Just as respondents pick up clues from numerical values assigned to rating scales, they use visual stimuli to

interpret and answer survey questions. Recent research on visual design in Web surveys, however, shows how good design can improve survey effectiveness (Christian, Dillman, and Smyth, 2007; Tourangeau, Couper, and Conrad, 2004). For a discussion of how effective visual design in a questionnaire can guide respondents and partly substitute for the role interviewers play, see Dillman (2007:462–93).

### **Existing Questions**

Of all the raw materials available to the survey researcher, perhaps the most important are questions that have been used in previous research. Most survey instruments contain existing questions, at least in adapted form, and it is easy to understand why. The use of existing questions shortcuts the measurement and testing processes. It also enables researchers to compare results across studies, to estimate trends, and, under certain conditions, to estimate response reliability (Sudman and Bradburn, 1982:14). Lest one be concerned about the ethics of using another person's questions, Seymour Sudman and Norman Bradburn (1982:14) note that "the mores of social science in general and survey research in particular not only permit but encourage the repetition of questions." Unless questionnaire items have been copyrighted, no permission is required.

Many sources of questions are available on most topics. Literature reviews on the research topic should uncover references with pertinent questions. One

#### **KEY POINT**

To the extent possible, survey researchers should use existing questions developed by reputable sources.

may also consult more general sources of questions, such as the CBS-New York Times poll, as indexed in the *The New York Times Index*, and the polls section of *Public Opinion Quarterly*, or data archives, such as the Roper Center and National Opinion Research Center. Finally, every

survey contains some demographic questions on age, gender, marital status, and so forth. An excellent critical discussion of alternative ways to word these basic questions about respondent characteristics may be found in Floyd Fowler (1995:166–76).

### **"Sketches" or Preliminaries**

Having considered likely raw materials, the survey designer, like the artist, will go on to make a "sketch." The sketch for the survey designer essentially is an outline of the question topics to be covered in the interview or questionnaire.

The most general principle to follow is to formulate your research objectives clearly before you begin. We could not agree more with Sudman and Bradburn's recommendation (1982:41), especially for beginning researchers, "that—before you write any questions—you put down on paper the aims of the study, hypotheses, table formats, and proposed analyses." This material then "should

not become a straightjacket for you," but should guide you in writing items, clarifying their meaning to respondents, and organizing them into a meaningful sequence.

Beyond a formal statement of objectives, certain practices in survey design have been shown to enhance the effectiveness of the research instrument. For instance, when a topic is somewhat abstract, using a number of items or questions results in better-quality data. In general, the more abstract the topic, the more items are required; however, this principle must be balanced by the need to keep the instrument reasonably short. The principle may be illustrated by a hypothetical interview schedule investigating job satisfaction. Topics to be included might be wages, hours or shift worked, opportunities for learning and advancement, whether the work is interesting, supervisor-worker relations, and other working conditions. The topic concerning hours or shift worked might well be covered by a couple of questions, whereas "supervisor-worker relations" is a bit more abstract; several distinct questions probably would be needed to get an accurate picture of the worker's satisfaction or dissatisfaction in this area.

Finally, the instrument design should reflect the perspectives of the target population as well as the survey designer. The job satisfaction survey outlined above, for example, could be ineffective if it excludes topics important to workers. **Focus group** discussions, which are unstructured discussions among a small group of participants led by a skilled interviewer, are often used to learn how people think about a survey topic (Fowler, 1995:105–10; Krueger and Casey, 2000). In planning a survey of sexual practices among teenagers, for example, teen focus groups could provide insight into how teens cognitively organize, label (vocabulary), and retrieve sexual information.

### **The Opening**

At this point, one should decide what the opening topic will be and draft the opening questions. It is best to have an interesting and nonthreatening topic at the beginning that will get respondents involved and motivate them to cooperate in completing the interview or questionnaire. The first question should be congruent with respondents' expectations: It should be a question they might reasonably expect to be asked, on the basis of what they have been told by the interviewer about the study. This sometimes involves using a question that has no research purpose other than motivating respondents by conforming to their preconceptions about what should occur in a competent survey. The first question also should be relatively easy to answer, thus preventing respondents from becoming discouraged or feeling inadequate to fulfill their role as respondents.

If both open-ended and closed-ended questions are used, the beginning is a good place to have an open-ended question.<sup>6</sup> Most people like to express their views and to have someone listen and take them seriously. An interesting opening question is a good way to meet this need of respondents and to get them to open up and warm to the respondent role. Here are two examples:

As far as you're concerned, what are the advantages of living in this neighborhood? What do you like about living here?

Let's talk first about medical care. What would you say are the main differences between the services provided by doctors and hospitals nowadays compared with what they were like when you were a child?

### **The Placement of Sensitive and Routine Questions**

It would be prudent to avoid both boring, routine questions and sensitive, personal questions in the beginning; build up interest, trust, and rapport before risking these. Uninteresting routine questions such as background information (e.g., age, gender, marital status) are often placed toward the end of the survey instrument. Asking personal questions (e.g., racial prejudices, income, sexual activity, alcohol or drug use, religious beliefs) prematurely may embarrass or otherwise upset respondents and possibly cause them to terminate the interview or question the researcher's motives.

Some researchers place sensitive or personal topics at the end of an interview, arguing that, if the respondent fails to cooperate at this point, not much information will be lost. However, this may leave respondents with a bad taste in their mouths and may promote negative feelings toward survey research. Probably it is best to introduce such questions after the interview is well under way because the respondent will have invested time and effort by then and possibly will have developed trust toward the research and/or interviewer. In addition, sensitive questions should fit into the question sequence logically; they should be preceded when possible by related but less sensitive questions or topics so that the relationship of the personal questions to the topic and to the research is clear to the respondent. It also may be helpful to precede the most sensitive questions with a direct explanation of their importance to the research and to repeat an assurance of confidentiality.

### **Order, Flow, and Transition**

After decisions have been made regarding the first topic and questions and you have a general idea of how you plan to introduce both sensitive and routine questions, the next task is to put the remaining topics in some reasonable order. What additional considerations go into the organization of topics?

The respondent's point of view must be considered as the researcher attempts to order the topics. The order must seem logical to respondents if their thinking about the questions is to be facilitated and motivation enhanced. Early topics should be easy to answer and of interest to the respondent; subsequent topics should seem to flow naturally from the early ones. It may even be that the survey instrument will begin with topics of little or no interest to the researcher but that will facilitate the introduction at a later point of topics more pertinent to the research objectives. An excellent example of topic flow is provided by the University of Michigan Surveys of Consumer Attitudes (SCA). According to Robert Kahn and Charles Cannell (1957:161-62),

the objectives of these annual surveys [were] to ascertain the respondent's income, his savings patterns and the amount he has in various forms of savings, his buying

intentions and major items purchased over the past year, his indebtedness, and his feeling about his own financial situation, both present and anticipated. The questionnaire starts with broad attitudinal questions on how the respondent feels about economic conditions generally, and moves to questions on his feelings about his own financial position and his expectations for the next few years. The interview then considers the respondent's assets, beginning with home ownership. How much is the house worth? When did he buy it? Does he have a mortgage? How much does he still owe on the mortgage? Similarly, ownership of automobiles is discussed. Then other major purchases are discussed. Next comes the topic of plans to purchase goods in the near future, which leads logically to the problem of sources of funds for such purchases. Will the money come from savings, from current income, or where? This introduces the topics of how much income is available and how much savings the person has. Last, to round out the picture, the amount of money already committed (debts) is discussed.

If there are to be questions that demand hard work on the part of the respondent, these should not be at the beginning but included sometime later, when commitment and momentum have been developed but before the respondent could become tired.

The researcher also must be sensitive to the problem of question-order effects (Schwarz et al., 2008). A topic or question appearing early in an interview may start the respondent thinking in a way that will affect later responses by activating or changing the information readily accessible in memory. Since respondents tend to truncate memory searches as soon as they have enough information for an acceptable answer, the most accessible information is likely to be that used recently to answer previous questions (Sudman, Bradburn, and Schwarz, 1996). For example, questioning cigarette smokers early in an interview about their beliefs in a link between cancer and smoking could influence their responses to later questions about perceived benefits and drawbacks of continued smoking. Even questions appearing later in self-administered questionnaires can influence responses to earlier questions since respondents may skip ahead and then return to the earlier questions (Sudman, Bradburn, and Schwarz, 1996:82).

Research indicates that questions vary in their susceptibility to order effects (McFarland, 1981; Schuman and Presser, 1981). In particular, general questions are more susceptible than questions with more specific content. When, for example, a general question about abortion followed a specific question (abortion in the case of a defect in the unborn child), respondents were less likely to support abortion than when the general question preceded the specific (Schuman and Presser, 1981:37). Responses to the specific question, however, were unaffected by its placement. This pattern of findings is consistent with Grice's nonredundancy principle. After responding to the specific question, respondents interpreted the subsequent general question as excluding birth defects in order to avoid reiterating information already provided.

On the other hand, respondents sometimes include, rather than exclude, information previously given to specific questions in responding to a general question (Sudman, Bradburn, and Schwarz, 1996). That is, they assimilate the old information, which is readily accessible in memory from answering previous questions, into

their response to the general question. For example, in an experiment with college students, the correlation between a specific question on dating frequency and a general question on life satisfaction dramatically increased when the specific question preceded the general question (Strack, Martin, and Schwarz, 1988). When respondents perceive specific and general questions as belonging together, however, they may exclude the information already given to specific questions from the general question to avoid being redundant (Schwarz, Strack, and Mai, 1991). Thus, whether responses from previous questions are included or excluded from subsequent general questions is difficult to predict in advance. Careful instrument pretesting may clarify how respondents are being influenced by question ordering.

#### KEY POINT

Although complex topics are best explored with multiple, overlapping questions, order effects can occur when respondents invoke the conversation principle of nonredundancy to answer similar questions differently.

to include or exclude certain information in answering the later questions (Sudman, Bradburn, and Schwarz, 1996).

After the researcher has made an outline of the topics, considered the location of routine and sensitive questions and the problem of question-order effects, and written the opening questions (and perhaps a few others), transitions between major topics should be considered. Transitions indicate that one topic is completed and another topic is to be discussed; the main objective is to focus the respondent's attention on the new topic. Transitions also may be used to explain briefly why the new topic will be discussed or how it relates to the research purposes. Although they are not needed between every change in topic, transitions can improve the flow of an interview or questionnaire as well as respondent understanding and motivation. The following are examples of topic transitions:

"Now I would like to ask some questions about your family. As you were growing up, let's say when you were around 16, how much influence do you remember having . . . ?"

"Okay, now I'd like to change the subject slightly to one part of campus life that we're particularly interested in. As you may know. . . ."

"I would like to shift the subject slightly and get some of your opinions about . . ."

"Fine. Now we have just a few background questions."

The sketch of the instrument will be completed by drafting an introduction. The introduction should explain briefly the general purpose of the study, assure the respondent of confidentiality, and provide basic instructions for responding to the interview or questionnaire. Of course, the researcher never divulges specific hy-

potheses or relationships of interest during the introduction, for to do so would be an invitation to respondents to give the desired or expected responses.

### Filling in the Sketch: Writing the Items

Drawing on the raw materials outlined earlier, the survey designer is now ready to "fill in the sketch" by writing the individual items. An artist has elements such as line, perspective, light effects, and color to aid in filling in a sketch. Are there principles or "elements of design" by which the survey designer likewise may be guided? We now consider useful principles for wording questions, for grouping questions to achieve greater effectiveness, and for avoiding common problems.

#### Using Language Effectively

In writing question items, using language effectively presents a real challenge. Even slight altering of the wording of a question can greatly affect responses, whether the question is open-ended or closed-ended. For example, a question might be written, "What is your annual income?" or "What is your total annual income from all sources?" A person answering the first item might neglect to consider income from such sources as interest on stocks or savings, sale of stocks, and rental income.

In general, research findings on question wording indicate that it is potentially very important but difficult to predict whether a wording change will have an effect (Converse and Presser, 1986). For this reason, it is essential to pretest new questions adequately. Furthermore, it is generally a good idea not to base conclusions on results from a single item. Approaches to the problem of wording effects include asking multiple questions on a topic; using two forms of the question, each with a different, randomly selected subset of respondents; and using open-ended questions as follow-ups to closed-ended questions to probe for the respondent's meaning (Converse and Presser, 1986). Edward Laumann and associates (1994:292–301), for example, used six different sets of questions in the National Health and Social Life Survey (NHSLS) to examine components of homosexuality (self-identification, homoerotic fantasies, and actual practice).

Additional guidelines for using language to obtain better data have been developed on the basis of survey researchers' experiences and of validity and reliability testing. Although some of these guidelines seem to be self-evident or commonsense suggestions, the plethora of poorly worded items continually being produced would seem to indicate a need for more attention to the language of items. The following questions provide a framework for examining the language of the items.

1. Are the items unambiguous, easily read, and sufficiently brief? Clarity and precision are essential qualities of well-worded items. At times, an item that appears perfectly clear to the designer may be very confusing or carry a different meaning to someone with a different background and point of view. The point is easily illustrated by the question, "How many years have you been living here?" To one respondent, "here" may mean the present house or apartment; to another, the city; and to another, the United States.

Especially troublesome are indefinite words such as "usually," "seldom," "many," "few," "here," "there"; these will have different meanings to different respondents. Following are two alternative items illustrating the problem; the second is an improvement over the first because the responses are specific and thereby have the same precise meaning for both researcher and respondent.

1. How often do you attend religious services?

- ( ) Seldom or never
- ( ) Often
- ( ) Very often
- ( ) Every day

2. How often do you attend religious services?

- ( ) Never
- ( ) Less than once a year
- ( ) About once or twice a year
- ( ) Several times a year
- ( ) About once a month
- ( ) Two or three times a month
- ( ) Nearly every week
- ( ) Every week
- ( ) Several times a week

Items also should be easy to read or hear accurately. Succinct wording of items will enhance the accuracy of responses, as will avoiding the use of negative words such as "not." Some respondents will skip over or tune out the negative word in an item and respond the opposite of the way the question is actually intended. If negative words must be used, it is wise to print them in all capitals, underline them, or verbally emphasize them.

If key terms are subject to multiple meanings, definitions should be provided before they are used in a question. In a self-administered section of the NHSSLS, for example, the respondents were asked about sexual partners after the following definition of "sex" was provided:

People mean different things by sex or sexual activity, but in answering these questions, we need everyone to use the same definition. Here, by "sex" or "sexual activity," we mean any mutually voluntary activity with another person that involves genital contact and sexual excitement or arousal. That is, feeling really turned on, even if intercourse or orgasm did not occur.

... please include all persons or times . . . where you had direct physical contact with the genitals (the sex organs) of someone else and sexual excitement or arousal occurred. Certain activities such as close dancing or kissing without genital contact should NOT be included. (Laumann et al., 1994:622)

If a key term is defined within a question, the definition should precede the actual question. Otherwise, respondents may skip or tune out the definition after

hearing the question (Fowler, 1995). In the following question, for example, female respondents might report only pregnancies that resulted in live births if the definition of "pregnancy" had followed the actual question:

Now I would like to ask you about *any* pregnancies you might have had. I'm interested in all your pregnancies, whether they resulted in a live birth, stillbirth, abortion, or miscarriage, even those which ended very early. How many pregnancies have you ever had, NOT including a current pregnancy? (Laumann et al., 1994:619)

2. Is the instrument's vocabulary appropriate for the respondents you intend to interview? If the intended study population is highly heterogeneous, the vocabulary must be kept extremely simple, and the survey designer should be aware of regional and other group differences in the meanings of words. To exemplify the problem of translating social science concepts into language understandable by nearly everyone, let us assume for a moment that you are studying "socialization" among a random sample of American parents. You decide to start the interview with a broad, open question. While any of the three examples that follow could be used, the third would doubtless be most effective.

What do you consider to be the most important factors in the socialization of children?

What child-rearing practices do you think are most important?

What things do you think are most important for parents to do if they want to bring up their children right?

On the other hand, with some surveys, "talking down" may be a potential problem. When sampling a more homogeneous or specialized group (say, city managers, doctors, or nuclear engineers), use vocabulary that is appropriately sophisticated and technical for that group.

In choosing an appropriate vocabulary to characterize sexual practices, Laumann and associates (1994) steered a middle course between language that would be too technical (e.g., "fellatio") or too colloquial or slangy (e.g., "blow job") to be widely understood in a national survey. Their solution was to use simple, standard English (e.g., "oral sex") and to provide definitions when terms were first introduced.

3. Do the questions contain a single idea, or are any of them addressing two or more issues at once? A **double-barreled question** is one in which two separate ideas are presented together as a unit. An example (perhaps from the "socialization" study) might be, "What factors contributed to your decision to marry and have children?" The researcher seems to assume that marrying and having children is a single act or decision, whereas in fact there are two questions being asked here. It is a good idea for the survey designer to examine all questions with the word "and" or "or" in them to be sure that they are not double-barreled.

Consider the following question that was included in a "questionnaire" distributed by a political interest group:

Do you believe that for every dollar of tax increase there should be two dollars in spending cuts with the savings earmarked for deficit and debt reduction?

- Yes
- No

About four ideas are hidden within this "quadruple-barreled" question. Should taxes be increased? Should every tax increase dollar be matched by two dollars in spending cuts? Should spending cut savings be applied to deficit reduction? Should savings be applied to debt reduction?

The single-idea principle may also be violated in response formats as illustrated by asking workers to describe their boss using such fixed-choice categories as "Very professional and friendly," "Generally professional and friendly," "Somewhat unprofessional and unfriendly," and "Very unprofessional and unfriendly." Even the popular Likert response categories combine two ideas, as in the following example.

A preschool child is likely to suffer if his or her mother works outside the home.

- Strongly agree
- Agree
- Disagree
- Strongly disagree

The respondent is given response choices that combine two different dimensions: agreement and strength of feelings (Fowler, 1995). The first response might be selected because the respondent is in complete agreement or has a very strong emotional conviction about how to raise young children. Floyd Fowler (1995) suggests an alternative wording to avoid such double-barreled response categories:

- Completely agree
- Generally agree
- Generally disagree
- Completely disagree

4. Are the items free of emotionally loaded words and other sources of bias? Emotionally loaded words and phrases, such as "communists," "racial preferences," "pro-life," "cops," or even "the president's statement," may evoke cognitive responses that have little to do with the real attitudes or opinion of the respondent regarding the issue the researcher is attempting to study.<sup>7</sup> In general, try to word questions in a neutral way, and avoid identifying a statement or position with any controversial or prestigious person or group. Notice the loaded words ("union czars," "forcing," "knuckle under") in the following question (Sudman and Bradburn, 1982:2):

Are you in favor of allowing construction union czars the power to shut down an entire construction site because of a dispute with a single contractor, thus forcing even more workers to knuckle under to union agencies?

This question was part of a questionnaire distributed by a political lobbying group for fund-raising purposes, a practice that social scientists consider deceptive and unethical.

Another source of bias is **leading questions**. Leading questions suggest a possible answer or make some responses seem more acceptable than others. The question "How often do you smoke marijuana?" may seem to imply that everyone indulges at least occasionally.<sup>8</sup> A question that begins "Do you agree . . ." may suggest to some persons that they ought to agree. The generally accepted practice is to balance attitudinal questions by using such phrases as "agree or disagree," "favor or oppose," and "satisfied or dissatisfied" (Sudman and Bradburn, 1982).<sup>9</sup>

An incomplete listing of alternatives either in the question or in the response options given is another common error. An example of the former would be the question "When you discipline your children, do you spank them, take away privileges, or what?" It would be better to either give a complete listing of alternatives or give none. The problem of making sure the responses to a closed question are exhaustive was discussed earlier in the chapter; it is also important that the responses be balanced. For example, if the responses represent attitudes toward the United Nations, it would be judicious to have an equal number of positive and negative statements, as well as a neutral statement. Furthermore, you would want to represent both extreme and moderate attitudes in each direction. Otherwise, respondents interpret an unbalanced set of response options as relevant to question intent and indicative of the range of typical or normal responses.

Consider the following question, which was used in a mail questionnaire to gauge University of California employee satisfaction with their health maintenance organizations (HMOs).

Based on the experience of everyone in your family, how satisfied was your family with your HMO plan in 1995?

- |                   |                                       |                                |
|-------------------|---------------------------------------|--------------------------------|
| 1. Very satisfied | 3. Neither satisfied nor dissatisfied | 5. Very dissatisfied           |
| 2. Satisfied      | 4. Dissatisfied                       | 6. Don't know or doesn't apply |

Both the question and the formatting of the response categories skew the responses in the direction of satisfaction. A more neutral wording such as "How would you rate your HMO?" is likely to produce more critical responses than asking "How satisfied . . . ?" Note also that the two "satisfied" categories (1 and 2) are spatially grouped together in such a way as to be seen next after reading the end of the question. The scale endpoint printed first tends to draw more responses than if the categories are reversed and it is printed last (Sudman, Bradburn, and Schwarz, 1996:157). It is not surprising, then, that some HMOs received 80 percent or higher satisfaction ratings (combining "very satisfied" and "satisfied" responses) on this leading question.

5. On personal and sensitive questions, is the wording as tactful, diplomatic, and face-saving as possible? The topic of sensitive questions was discussed earlier in reference to their placement in the instrument. It was pointed out that motivation

could be enhanced not only by logical placement of the item but also by showing the relevance of the question to the research purpose, reassuring the respondent of confidentiality, and including a wide range of response options. In addition to these efforts, careful diplomatic wording of the items may boost motivation and facilitate more candid responses. A housewife may feel irritated or defensive if asked "Do you work?" but perhaps not if asked "Do you work outside the home?"

A sensitive question might also be preceded by a statement that in effect sanctions the less socially desirable response, such as "Some people feel that smoking marijuana is pleasant and harmless, while others feel that it is harmful. What do you think?" or "Many people have taken an item from a store without paying for some reason. Have you ever done this?"

Finally, to make the best use of language, the survey designer will usually need to make, test, and revise several successive drafts of the questions. Early drafts

#### KEY POINT

The survey designer starts with a topic outline, drafts and refines questions, and then pretests the survey instrument.

should be subjected to careful scrutiny by both the researcher and colleagues with the aforementioned wording issues in mind. Then, cognitive interviewing techniques, discussed later in this chapter, should be used to identify further problems by probing the thought processes employed to answer the questions. Finally, a semifinal draft

should be field-pretested under realistic conditions with a population similar to the one for which the survey instrument is designed.

#### *The "Frame of Reference" Problem*

Often, the questions we ask people seem clear in meaning to us but can be answered from several perspectives or frames of reference. For example, suppose a survey of second-semester college students asked, "Generally speaking, how satisfied are you with your decision to attend State University?" Students giving the same response, such as "relatively satisfied," could have very different reasons for doing so. One may feel relatively satisfied because she has a generous scholarship and feels the school is as good as most others from an academic standpoint. Another may be thinking of the social life, and a third of intramural sports activities. Yet, from their answers, the researcher would not know what the respondents' reasons were for their satisfaction or dissatisfaction.

There are a number of ways to determine or to control the respondent's frame of reference. A straightforward way to determine the frame of reference is to follow the question with a probe such as "Can you tell me why you feel that way?" "What things specifically do you feel are satisfactory (or unsatisfactory) about State University?" Another simple means of controlling the frame of reference for individual questions is to specify the frame of reference within the question, such as "Compared with other universities in the state system, how do you feel about the intellectual life at Caulfield State?"

By particular arrangements of questions, the researcher also can direct the respondent to the investigator's frame of reference. A **funnel sequence** (Kahn and Cannell, 1957:158–60) moves from a very general question to progressively more

specific questions. Suppose one wanted to study the impact of inflation on people's attitudes about the performance of the president. Research on question-order effects, reported earlier, suggests that asking specific questions about inflation first might impose this frame of reference on respondents so that later general questions about the president's performance are judged with reference to the president's inflation efforts. Instead, a funnel sequence could start out with general questions about achievements ("What do you think about the president's performance in office?" "Why do you feel this way?"), which will likely disclose the respondents' frame of reference. These questions then could be followed by questions on inflation ("Do you think we have a serious inflation problem?" "Has it had much effect on you?") and, finally, specific questions about the president's activity in this area ("Do you believe the president is doing a good job of fighting inflation?").

The previous example illustrates the effectiveness of a funnel sequence in avoiding the possibility that asking more specific questions first will influence responses to more general questions. This sequence also offers the advantage of beginning with the respondent's ideas and perspectives, which may increase interest and motivation. Funnel sequences may consist entirely of open-ended questions or of an open-ended question (or questions) followed by closed-ended questions.

A common frame of reference also may be established through an **inverted-funnel sequence** of related questions (Kahn and Cannell, 1957:160). Here, one begins with the most specific questions and ends with the most general. While this approach lacks the advantages of the funnel sequence, it is useful in some situations. First, it may be used to ensure that all respondents are considering the same points or circumstances before expressing their general opinions. For example, if we wanted to make sure that respondents were judging the president's performance in office on similar bases, an inverted-funnel sequence would enable us to bring up and question performance in specific areas (inflation, unemployment, foreign policy) before asking for a general evaluation. A second advantage of the inverted funnel sequence is that, whether or not respondents have previously formed an opinion regarding the final question in the sequence, all will have time to think through certain aspects of a complex issue before giving their opinion. Instead of asking respondents to express immediately their attitude toward liberalizing laws on abortions, for example, one might ask about approval of abortion in various specific circumstances (if there is a strong chance of a birth deformity, if the woman became pregnant as a result of rape, if the woman's own health is seriously endangered, if the woman is married and does not want the child, if the family cannot afford any more children), at the end of which the respondents' general opinion would be sought. (See Box 10.2 for additional examples of a funnel sequence and inverted-funnel sequence.)

Will an inverted-funnel sequence be susceptible to order effects, since respondents may avoid being redundant by excluding previously given information from the final general question? There is limited evidence (Schwarz, Strack, and Mai, 1991) that another conversational norm, a request for a summary judgment, may operate when a general question follows a block of specific questions as in the second example of Box 10.2.

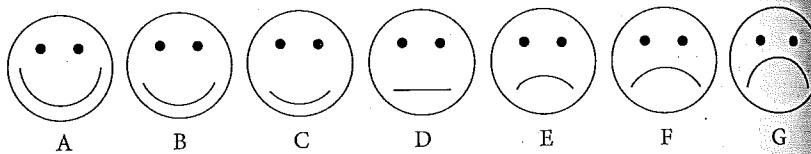
### BOX 10.2 Examples of Funnel and Inverted-Funnel Sequences of Questions

Here is a sample funnel sequence from a study of union printers (Lipset, Trow, and Coleman, 1956:493-94).

7. (a) All things considered, how do you like printing as an occupation?  
Do you dislike it?  
Are you indifferent?  
Do you like it fairly well?  
Do you like it very much?
- (b) Why do you feel this way?
8. (a) Is there any occupation you would like to have other than the one you now have—either in or outside the printing trade?  
(If so) Which one?
9. Let's look at it another way: If you were starting all over again, what occupation would you want to get into?
10. (a) How would you rate printing as an occupation? For example:  
(1) Would you rate the *pay* as excellent, good, fair, or poor?  
(2) How about *job security* in the printing trade? Would you rate it as excellent, good, fair, or poor?  
(3) How about the *prestige* printing receives from people outside the trade?

A brief inverted-funnel sequence used to measure perceptions of well-being (Andrews and Withey, 1976:376) is reproduced here.

13. Here are some faces expressing various feelings. Below each is a letter.



- 13.1 Which face comes closest to expressing how you feel about *your life as a whole*? WRITE LETTER ON LINE
- 13.2 Which comes closest to expressing how you feel about *your house or apartment*?
- 13.3 Which comes closest to expressing how you feel about *what our national government is doing*?
- 13.4 Which comes closest to expressing how you feel about *your sparetime activities*?
- 13.5 Which comes closest to expressing how you feel about *your independence or freedom—the chance you have to do what you want*?
- 13.6 Which comes closest to expressing how you feel about *your standard of living—the things you have like housing, car, furniture, recreation, and the like*?

14. Taking all things together, how would you say things are these days—would you say you're *very happy*, *pretty happy*, or *not too happy* these days? CHECK ONE BOX

- Very happy
- Pretty happy
- Not too happy

### Reason Analysis

The funnel sequence illustrates the benefits of using a sequence of questions to explore complex issues. Similarly, a well-devised series of questions is invariably more effective than the simple question "Why?" in finding out the reasons for people's behavior.<sup>10</sup> Suppose we ask undergraduates why they decided to attend UCLA, and we receive the following responses: Mary—"My parents convinced me"; Sam—"Because I live in LA"; Pascual—"For a scholarship"; Irma—"To be with my boyfriend"; Reuben—"It's a fun school and I wasn't accepted at the other schools I applied to." Not only are these reasons brief and quite diverse, they seem incomplete; surely these students selected UCLA for more than one reason (only Reuben mentioned two reasons). Perhaps Mary and Sam also were influenced by scholarships; maybe Irma was not alone in having a close friend at UCLA, or Reuben was not the only one turned down by other schools. Also, other determinants of the respondents' choice of UCLA may have been overlooked, such as the recommendations of high school teachers, the academic reputation of UCLA, and the climate of southern California.

The simple "why?" question implies that a simple, prompt answer is desired. Consequently, respondents may expend minimum effort (satisficing) to generate quickly a plausible answer that is easy to verbalize. How, then, can we go about discovering the main factors that influenced our respondents to attend UCLA? Fortunately, Hans Zeisel (1968) systematized the process of asking "why?" The key idea in Zeisel's **reason analysis** is the development of an "accounting scheme" outlining the general categories of reasons, or dimensions of the decision, which in turn provides a model or structure for formulating a comprehensive series of questions.

The accounting scheme is based on the research objectives (not all reasons may be of interest), on the researcher's creativity, and usually on informal exploratory interviewing or focus group discussions with members of the group under study. Unless one is reasonably well informed about the various categories of reasons for an action (perhaps from previous research or experience), these are sought through exploratory interviewing. The information from these preliminary interviews is culled for reasons not pertinent to the research objectives,<sup>11</sup> and the remaining responses are creatively grouped into an integrated model of relevant categories to form the accounting scheme.

Often, an accounting scheme is structured in terms of the typical stages in a decision process. To return to our example, for many students the decision to attend UCLA involved (1) a decision to attend college, (2) the selection of schools to apply to for admission, and (3) the final choice of UCLA. At each of these stages the

respondent's behavior may be influenced by sources of information (knowledge, gossip, rumors learned from the media, friends, and strangers), advice from significant others (recommendations of friends, parents, teachers), constraints or limitations (financial needs, health requirements, need to live with or near a friend or relative), the respondent's particular needs (academic interests, vocational objectives, social and recreational interests), perceived characteristics of the schools considered (reputation or image, strengths and weaknesses, location and climate, housing availability, costs), and responses by the schools (offers of aid or scholarships, athletic awards, special recruitment efforts, rejection of admission application).

After a systematic outline of the accounting scheme just sketched out is made the final step in our reason analysis would be to write a series of questions to assist the respondents in reviewing the stages of their decisions to attend UCLA: "When did you first seriously consider going to college?" "How did you reach this decision?" "Did you apply to any other colleges or universities? Which ones?" "What were the characteristics of UCLA that you particularly liked?" "Did your parents, friends, teachers, or any other persons help you come to your decision? Who?" "How much influence did this have on you?" Although the process of reason analysis in this situation obviously calls for intense mental labor, we would end up knowing immensely more about the college decisions of Mary, Sam, Pascual, Irma, and Reuben than had we simply asked, "Why UCLA?"

The sequence of stages of a decision process is not the only basis for accounting schemes. Zeisel (1968) discusses in detail several common schemes. In studies of why people move from place A to place B (or shift jobs or in general, change preferences from A to B), the reason analysis focuses particularly on the negative aspects of A and the positive aspects of B. This is called the "push-pull scheme" in migration studies since the decision to move may be precipitated by factors that push one out of the old location (e.g., an increase in rent, a change in neighborhood character, the loss of a job) and/or that pull or attract one to a new habitat (e.g., the anticipation of employment, a better wage, a better environment). Another common accounting scheme used in studies of consumer behavior has three major dimensions: the perceived attributes of the consumer product (its effects, price, etc.), the motives or needs of the respondent to be satisfied by the product, and the influences that affect the purchase decision (advertising, recommendations of friends, etc.).<sup>12</sup>

Reason analysis thus is a very useful and general technique for guiding questionnaire construction and for avoiding superficial and incomplete answers to important questions. The researcher basically is asking many specific "whys?" rather than a general "why?" question. Compared with a general question, a comprehensive series of specific questions better conveys to the respondent the information sought by the researcher, as well as facilitating information retrieval by triggering different memory associations.

### **Memory Problems**

There are two kinds of memory problems. First, respondents may be unable to retrieve information the researcher is seeking about life events. According to Cannell and Kahn (1968), the ease with which material may be recalled depends on three

factors: how long ago the event occurred, how significant the event was when it occurred, and how relevant the event is to the respondent's life currently. Generally, respondents will have a longer memory for significant rare and costly events in their lives, such as getting married or having major surgery, than for more trivial habitual occurrences, such as losing a small amount at gambling or the content of the previous evening's television programs. Furthermore, they will remember more effectively events of continuing importance in their lives. The presence of an unsightly scar, for example, is a reminder of the accident that produced it.

A second difficulty arises from memory distortion. People do not uniformly recall events objectively; rather, memories seem to be distorted either in the process of organizing one's past and making it consistent or in an unconscious effort to maintain a positive self-image. One common type of memory distortion in responses to survey questions is called "telescoping" (Sudman and Bradburn, 1974). This occurs when a respondent erroneously recalls the timing of an event as having occurred more recently than it did. When respondents are asked about events that occurred within a specific reference period, such as visits to doctors in the past 6 months, telescoping leads to overreporting because some events that happened before the reference period (e.g., a doctor visit 7 months ago) are remembered as occurring during the reference period. Another type of memory distortion involves respondents reconstructing the past by projecting their present experiences and views backward (Groves et al., 2004:217). For example, respondents may draw upon their present attitude toward capital punishment to infer their attitude 10 years ago.

What can be done to stimulate accurate recall? Here are some suggestions (see Cannell and Kahn, 1968; Converse and Presser, 1986; Sudman, Bradburn, and Schwarz, 1996):

1. Shorten the reference period. As a rule, the shorter the recall period, the more accurate the report. Sudman and Bradburn (1982) emphasize the appropriateness of the time period—with memory of a year or more for high-salience events (e.g., major accidents or illnesses) and a few weeks to a month for low-salience events. For example, the NHLS (Laumann et al., 1994) asked about sexual partners during the last 12 months, whereas the National Health Interview Survey, 1992: Youth Risk Behavior Supplement (Willard and Schoenborn, 1995) used a shorter reference period of 30 days to ask youth about episodes of heavy drinking. Jean Converse and Stanley Presser (1986) recommend that, for certain events, the time period be narrowed to the immediate past, such as yesterday or last week. For example, to obtain an estimate of the amount of time students spend studying, it would be better to ask "How much time did you spend studying on weekdays during the past week?" than to ask "How much time per day do you spend studying?"

2. Give respondents more time to search their memories for a response. This may be accomplished by increasing the length of the question, by instructing the respondents that the question may take some time to answer, and by training interviewers to allow adequate time for a response (Sudman, Bradburn, and Schwarz, 1996:225).

3. Use retrieval cues by providing a helpful context and question sequence as in "reason analysis." Rather than asking a respondent, for example, "At what addresses have you resided in the last 10 years?" begin with the present, asking "How long have

you lived in Atlanta?" A second question might ask if the respondent had lived at other addresses in Atlanta, and a later question might ask about residences before moving to Atlanta. Or one might ask the respondent to recall an important past event, such as completing school or leaving the parental home, and then guide the respondent to recall subsequent events (marriage, first job, etc.) in a forward time sequence.

4. Some questions demanding accurate memory may be best asked by means of closed questions utilizing lists. Doctors sometimes present new patients with a list of illnesses and ask them which they have had. Social researchers likewise may present lists for aiding recall of such things as magazines read, television programs watched, and organizational memberships.

5. Conduct a household inventory with the respondent. Studies of consumer behavior have revealed that even the current day's grocery purchases (type, brand, size) are best documented by examining with respondents their refrigerator and pantry provisions.

6. Where appropriate, ask respondents to check their records (birth and marriage certificates, school records, financial information, scrapbooks, photographs, etc.).

The first suggestion as well as suggestions 3 through 5, which involve the use of "aided-recall" procedures, deal directly with the problem of forgetting but also may contribute to some memory distortion. A shorter reference period, for example, decreases omission errors but increases dating errors (telescoping).<sup>13</sup> The use of records deals most effectively with both kinds of memory problems, although this information is often difficult to acquire.

All of these suggestions are geared toward conventional standardized interviewing, which has been the focus of this chapter. By contrast, a relatively new method, the event history calendar (EHC), employs "flexible" semistructured interviewing to enhance recall (Belli et al., 2007). In EHC interviewing, information is collected according to a calendar of salient life events (e.g., residence, schooling, employment, marriage) that extend across a predetermined time period (e.g., 2 years, 30 years) and serve as a common reference point during the interview. As interviewers follow a timeline in collecting information on major life events, questions about these events serve as retrieval cues that enable respondents to recall other events of interest in the survey.

**KEY POINT**  *Reason analysis accounting schemes and event history calendars use cuing techniques to facilitate respondents' information retrieval.*

For example, someone may recall that he or she started smoking a week after he or she graduated from college. Unlike standardized interviewing, interviewers may construct their own wording of questions and may decide how to use information from one life domain to help respondents recall events in another.<sup>14</sup>

### Response Bias Problems

Other instrumentation problems arise from the social situation surrounding the administration of interviews and questionnaires. Just as experimenters must be concerned about how the laboratory setting may produce disingenuous subject behavior, the survey researcher must be alert to how the social situation in surveys may

produce responses irrelevant to the object of measurement. We discussed this problem briefly in chapter 9 in relation to response effects introduced by the interaction between interviewers and respondents and extensively in this chapter from the perspective of cognitive, conversational, and motivational theories. Here, we focus on biases produced by respondent tendencies to answer in certain ways as a function of the content or form of survey questions.

One frequent response bias tendency is to answer in the direction of **social desirability** (see DeMaio, 1984). We all have our private self-image to maintain; in addition, many respondents will want to make a good impression on the researcher by appearing sensible, healthy, happy, mentally sound, free of racial prejudice, and the like. Some individuals and groups demonstrate this tendency more than others. Indeed, Derek Phillips (1971:87–88) suggests that the consistent finding of greater happiness, better mental health, and lower racial prejudice among middle-class compared with lower-class respondents may not reflect true class differences in these variables but instead a greater concern among the middle class to give socially desirable responses. Some common techniques for minimizing social desirability bias have been mentioned previously: use of indirect questions, careful placement and wording of sensitive questions, assurances of anonymity and scientific importance, statements sanctioning less socially desirable responses, building of rapport between interviewer and respondent, and collection of sensitive information within face-to-face interviews with self-administered forms or other modes that provide privacy to the interviewees.

A second response bias is the **acquiescence response set**. This is the tendency for respondents to be very agreeable: Presented with a question having such an option as "yes/no" or "agree/disagree," they are more apt to agree or say yes than to disagree or say no. This tendency also extends to such formats as the Likert scale, where there are gradations of agreeing or disagreeing. One way of circumventing the problem is to give specific content to the response options. For example, one may construct the second item below to avoid a possible acquiescence effect in response to the first item.

1. Rent control is necessary in order for many of the people of River City to obtain adequate housing.

- ( ) Completely agree
- ( ) Generally agree
- ( ) Uncertain
- ( ) Generally disagree
- ( ) Completely disagree

2. From the following statements, select the one that most closely represents your opinion about the need for rent control in River City.

- a. Rent control is necessary in order for many of the people of River City to obtain adequate housing.
- b. Rent control is necessary in order for some of the people of River City to obtain adequate housing.
- c. I am uncertain whether rent control is necessary or not.
- d. Very few people are really helped by rent control.
- e. Rent control is unnecessary.

One problem with this approach, however, is that it requires much greater effort on the part of the survey designer.

Another way to control as well as check for an acquiescence response set when the researcher wishes to retain a yes/no or an agree/disagree format is to include within the instrument two items measuring the same concept that are roughly opposite in meaning. Obviously, the items should be located at different points in the instrument. The following item pair from the F scale, a measure of authoritarianism, illustrates this approach (Bass, 1955:619):

1. Human nature being what it is, there will always be war and conflict.
2. Human nature being what it is, universal peace will come about eventually.

Responses to the two conflicting statements may be compared to see whether a significant number of persons answered both statements affirmatively. The major problem with this approach is the difficulty of constructing truly contradictory items. It would not necessarily be inconsistent, for example, to reject both of the above statements.

Ordinal or position biases represent a third response tendency. In visual modes of data collection, some respondents are more likely to pick response options listed earlier—a primacy effect—such as the first choice in a multiple-choice question. In auditory modes, the opposite—a recency effect—may happen, as when options presented later are chosen more frequently. In 548 telephone experiments, in which two response categories were randomly ordered, recency effects occurred predominantly in questions or answers that were difficult to comprehend and in those that were asked after many prior questions (Holbrook et al., 2007). As these recency biases may reflect satisficing behavior, Allyson Holbrook and associates (2007) recommend simple wording for questions and for response categories, short questionnaires, and rotating the order of response choices in telephone surveys. In visual modes, ways of addressing position biases, as well as the tendency to acquiesce, include varying the arrangement of questions and the manner in which they are asked as well as avoidance of “check-all-that-apply” blocks of questions (Dillman, 2007).

### **Format Considerations**

Finally, the format or physical organization of the survey instrument must be planned. Obviously, there are many possibilities, but a few main points should be considered (for Web and computer-assisted telephone interviewing [CATI] surveys, see Dillman [2007] and Gwartney [2007], respectively).

Probably most important, the physical form of the instrument should be appealing to respondents and interviewers. In addition, the design should facilitate the tasks of reading, completing, and coding. Numbering items consecutively throughout, using clear type, and allowing sufficient space between items and between response options can help greatly. With interview schedules, different type styles (or capitalization and underlining) should be used to differentiate instructions to the interviewer from material to be read aloud to interviewees.

While planning the physical organization, it would be appropriate to consider whether using “filter” and “contingency questions” would improve the instrument. Contingency questions are intended for only a part of the sample of respondents. By addressing only those persons for whom the questions are clearly relevant, contingency questions avoid the waste of time and possible decline in respondent motivation that may occur when the same questions are asked of all respondents. Responses to a filter question determine who is to answer which subsequent contingency question(s). One format is illustrated below from a hypothetical survey of adolescent sexual attitudes and practices.

19. Have you ever had sexual intercourse? (FILTER QUESTION)
- ( ) Yes (Please answer questions 20–29.) (CONTINGENCY QUESTIONS)
- ( ) No (Go to question 30. Please skip questions 20–29.)

(Box 10.3 presents a series of filter and contingency questions from a GSS interview schedule.)

### **Mixed-Mode Instrument Designs**

Instruments optimized for particular modes are not easily mixed in respondent-specific mixed-mode designs in which different modes are used to survey different respondents. Motivating respondents, clarifying questions, probing, and other functions performed by interviewers are difficult to incorporate in paper-and-pencil forms or to program into computer-assisted self-interviewing (CAS) and Web surveys. The absence of visual contact implies that optimal telephone surveys have to be short with simple questions. Questions and/or response choices may be randomly rotated in CATI but not in paper-and-pencil questionnaires. Several strategies are recommended to address these problems. One approach is to optimize the instrument for a preferred or main mode (Biemer and Lyberg, 2003). In a mail survey, for example, telephone interviewing might be used only to follow up with the mail nonrespondents. If the mail survey instrument is then optimized for visual self-administration, it will be less than optimal for CATI follow-up efforts.

When one mode is not preferred over the others, the instrument may be optimized for each mode separately (mode-specific design) or constructed the same for each mode (unified-mode design) (Dillman, 2008). In a mode-specific design, for example, face-to-face interviewers might hand out response cards for lengthy response choices even though visual cards could not be used in telephone interviews. Although the rationale for a mode-specific design is to use the best data-collection approach for each mode, the effort may be undermined by the incomparability of the data. A unified-mode design, on the other hand, aims to achieve comparability by using the same instrument in each mode. For example, if self-administered survey questions include explicit “don’t know” options, they also should be read with the questions in an interviewer-administered mode. (For mixed-mode design principles, see Dillman, 2007:232–40).

### BOX 10.3 Examples of Filter and Contingency Questions

The following is one page from an interview schedule for the General Social Survey. The physical layout is designed to facilitate interviewing and computer processing.

29. Last week were you working full time, part time, going to school, keeping house, or what?

**CIRCLE ONE CODE ONLY. IF MORE THAN ONE RESPONSE, GIVE PREFERENCE TO SMALLEST CODE NUMBER THAT APPLIES.**



Working full time .....	(ASK A) .....	.01	09-10/
Working part time .....	(ASK A) .....	.02	
With a job, but not at work because of temporary illness, vacation, strike .....	(ASK B) .....	.03	
Unemployed, laid off, looking for work .....	(GO TO Q.30) .....	.04	
Retired .....	(ASK C) .....	.05	
In school .....	(ASK C) .....	.06	
Keeping house .....	(ASK C) .....	.07	
OTHER (SPECIFY AND ASK C)		.08	

A. IF WORKING, FULL OR PART TIME: How many hours did you work last week, at all jobs?

Hours:   11-12/

NOW GO TO Q.30

B. IF WITH A JOB, BUT NOT AT WORK: How many hours a week do you usually work, at all jobs?

Hours:   13-14/

NOW GO TO Q.30

C. IF RETIRED, IN SCHOOL, KEEPING HOUSE, OR OTHER: Did you ever work for as long as one year?

Yes ..... (ASK Q.30) ..... 1 15/  
No ..... (SKIP TO INSTRUCTIONS BEFORE Q.33) ..... 2

### Pretesting

Throughout the chapter, we repeatedly have emphasized the importance of pretesting—of evaluating survey questions to determine if respondents clearly understand and are able to answer them. Failure to conduct adequate pretesting can result in a meaningless study. Once the study has been conducted, it is too late to benefit from the information, for example, that in one item 99 percent of the respondents chose the same option or that a large number of respondents misunderstood the meaning of a question. Experience has shown that the amount of effort expended on study planning and pretesting is related directly to the ease with which data may be analyzed and to the quality of results.

Pretesting should begin as soon as the survey instrument, or portions of it, have been drafted. Traditionally, pretesting has been done solely “in the field,” that is, in the homes of respondents drawn from the target population. However, sparked by the recent interest in cognitive aspects of surveys, researchers have begun to test questions in the laboratory. We first review new procedures developed in the laboratory, which have proven to be very effective in

identifying potential wording, ordering, and formatting problems. The information gained through these procedures gives direction to further revision efforts. Often, several pretests and revisions of a questionnaire may be necessary to arrive at a good semifinal draft. Once the questionnaire is in this form, it is routinely tested in the field; therefore, we also discuss various techniques used in field pretesting.

### Cognitive Laboratory Interviews

The goal of cognitive interviewing is to understand the thought processes involved in answering survey questions. Verbal reports from respondents, collected during or after their response to pretest questions, are used to identify problems and revise questions. The cognitive laboratories established at the Census Bureau and other federal statistical agencies typically use paid subjects who are recruited for intensive sessions lasting 1 to 2 hours; however, cognitive interviewing also may take place outside a laboratory facility (e.g., in subjects’ homes) and involve only a tape or a video recorder (Fowler, 1995; DeMaio and Rothgeb, 1996). In either case, the primary methods are “thinkaloud” interviews, probing questions, and paraphrasing follow-ups (for a more complete review, see Presser et al., 2004; Beatty and Willis, 2007).

#### “Thinkaloud” interviews

The respondents are asked to think out loud, reporting everything that comes to mind, while arriving at answers to the questions. They verbalize their thought processes either concurrently as they work out an answer to each question or retrospectively after they answer each question or complete the survey (Dillman, 2007:142–46). Here is an example of an improved question resulting from cognitive interviewing:

One question asked, "How many times during the past twelve months have you stopped smoking for one day or longer?" The intent of this question was to measure attempts to quit smoking. However, the question was not always interpreted this way. Several respondents included instances when they had not smoked for at least one day because of illness, excessive drinking the previous day, or other extraneous circumstances. The revised version makes the intent clearer, asking specifically, "How many times during the past twelve months have you stopped smoking for one day or longer because you were trying to quit smoking?" (DeMaio and Rothgeb, 1996:182–83).

### Probing questions

A weakness of the thinkaloud technique is that some respondents have great difficulty verbally describing their thought processes (Schaeffer and Presser, 2003). Instead, an interviewer can focus on particular aspects of the cognitive processes by special probes, such as "What did you think I meant by 'stopping smoking'?" Probes may directly follow each question (when the information used to answer the question is freshest in memory) or may come at the end of the survey (to avoid reactivity associated with the probing). Here is an example of probes used to pretest a question intended for elderly people (Jobe, Keller, and Smith, 1996:201):

By yourself and without using special equipment, how much difficulty do you have bathing or showering: some, a lot, or are you unable to do it?

Probes: "I said without using special equipment. What sort of things do you think would be special equipment?" "Do you use anything to help you bathe or shower?"

### Paraphrasing follow-ups

Another technique is to ask respondents to summarize or repeat the question in their own words. Not only is this a good test of whether respondents understand the literal and intended meaning of the question, but it may also reveal better ways to word the question.

Cognitive diagnostic procedures are only a partial solution to good instrument development. First, they are better at diagnosing problems than providing solutions. Second, since they are usually based on small, unrepresentative samples of paid subjects, they may not detect problems that will occur in subgroups of larger, more diverse populations. Third, different problems may occur in the "field" under more realistic interviewing conditions. We now examine techniques to field test survey instruments.

### *Field Pretesting*

**Field pretesting** a survey instrument consists of trying it out on a small sample of persons (usually twenty-five to fifty) having characteristics similar to those of the target group of respondents. The pretest group is normally not a probability sample since you are not planning to publish your pretest findings. However, it should be as heterogeneous as the target population. For example, if your target group is a

national sample of college and university students, the pretest group should include college students at all levels (first-year through graduate students) and from different types of institutions (large, small, religious, secular, liberal arts, technical, etc.).

Field pretesting should provide answers to questions such as these:

Does the level of language match the sophistication of the target population? Are instructions to respondents and to interviewers clear? Are transitions smooth and informative?

Are the questions and question format varied enough to retain respondents' interest and attention?

Are responses to open questions so diverse as to be impossible to analyze?

Are the choice options to closed questions clear and exhaustive?

Are interviewing aids such as cards or photographs effective and practical?

Are there questions that respondents resist answering?

How long, generally, does the interview take to complete?

For some time, field pretesting to identify such problems has consisted of having either experienced interviewers or interviewers in training both administer the draft questionnaire and observe the process. The interviewers might take notes during the interviews or file reports afterward, but their observations generally are conveyed in a group oral debriefing (Converse and Presser, 1986). However, this process has several limitations (Fowler, 1995; Fowler and Cannell, 1996). Playing the role of interviewer may interfere with the task of observing the process; each interviewer's observations are based on a small number of interviews, which may not be adequate for reliably assessing question problems; the standards for evaluation may not be well articulated or may be applied inconsistently, resulting in a lack of agreement about problem questions; and the recognition of question-comprehension problems is limited to items in which respondents ask for clarification or give inappropriate answers. Some of these problems may be addressed by cognitive laboratory interviewing. Recently, however, several strategies have been applied to make field pretesting more systematic and reliable. These include behavior coding, respondent debriefings, interviewing ratings, split-ballot tests, and response analysis.

### Behavioral coding

Systematically coding live or taped interviewer–respondent interactions is an effective technique for identifying instrument problems under realistic field conditions. Coding consists of categorizing responses and counting the number in each category (see chapter 15). In **behavioral coding** of pretest interviews, interactions with respect to each question are coded to identify the frequency of problematic respondent and interviewer behaviors, such as interviewers incorrectly reading or skipping questions, respondents interrupting interviewers before the question is completely read, respondents requesting that the question be repeated or clarified, and interviewers probing to follow up on inadequate answers (Fowler, 1995:116–21; Holbrook, Cho, and Johnson, 2006). For example, behavioral coding revealed that a question commonly used in governmental health studies—"When was the last time you had a general physical

examination or checkup?"—elicited a high proportion (87 percent) of inadequate answers because respondents did not understand the meaning of "general physical examination or checkup" and if the "when" requested a date, elapsed time, or their age (Fowler and Cannell, 1996). Behavioral coding is designed to identify potential problems. The next two pretest techniques, which involve the active participation of respondents and interviewers, are better at identifying the nature of the problem.

### Respondent debriefings

In chapter 7 we discussed how debriefing sessions following experiments provide valuable insights into subjects' understanding and motivation during the experiment. Similarly, **respondent debriefings**, which are structured follow-up questions at the end of pretest interviews, may reveal question-and-answer problems as well as the sources of the problems (for procedural details, see DeMaio and Rothgeb, 1996:188–94; Martin, 2004). Since the results of cognitive interviews tend to be limited in generalizability due to very small, unrepresentative samples, respondent debriefing provides a way to assess the external validity of cognitive testing results.

### Interviewer debriefings

Another perspective on question-and-answer problems may be sought from the pretest interviewers. For example, interviewers reported that a relentless sequence of Current Population Survey (CPS) labor force questions angered retirees and women working at home by giving the impression that the "government believes persons should either be working or looking for work" (Campanelli, Martin, and Rothgeb, 1991:255). Although **interviewer debriefings** customarily entail informal group discussions with the interviewers, Fowler (1995:121–24) recommends that more systematic information be collected by having the interviewers fill out standardized ratings of each question prior to the debriefing session.

### Split-ballot tests

Experimental manipulations of question ordering, wording, and formats, which survey designers commonly call "randomized" or **split-ballot** tests, are a costly but effective way to check out suspected problems or weaknesses under field conditions (see discussion of survey experimental designs in chapter 7 and Box 10.1). Split-ballot tests require random assignment of adequate-sized samples of pretest respondents to the versions being tested. For example, the results of cognitive interviewing might suggest a serious problem of question-order effects. This suspicion could be experimentally tested by randomly assigning pretest respondents to two forms of the instrument in which the order of related questions differs and then comparing the responses on the two forms.

### Response analysis

The responses of the pretest respondents also can be tabulated and examined for such problems as a low response rate to sensitive questions, the incidence of "don't know" responses, items where nearly everyone makes the same response, the adequacy of responses to open-ended questions, or confusion as to the meaning of ques-

tions. The generalizability of a **response analysis** is dependent on the pretest sample size and the degree to which the pretest respondents resemble the target population.

Each of these pretesting techniques offers a slightly different window on the question-and-answer process: Behavior coding allows the researcher to see interviewer-respondent interaction under actual field conditions, respondent debriefings and other cognitive interviewing methods reveal problems from the respondent's perspective, interviewer debriefings expose problems from the interviewer's perspective, response analysis reveals problems from the researcher's perspective, and split-ballot tests show differences among instrument versions. This suggests that some pretest methods are better than others at identifying particular types of question problems.

To test this idea, Stanley Presser and Johnny Blair (1994) systematically compared four pretest methods: conventional field pretesting with oral group debriefings, behavior coding of live pretest interviews, cognitive interviewing combining follow-up probes with concurrent and retrospective thinkalouds, and panels of experts. Each method was applied to a single questionnaire in repeated trials. Results indicated that behavior coding was the most reliable and conventional pretesting least reliable, expert panels identified the largest number of problems and were the least costly, experts and cognitive interviewing were the only methods likely to spot problems affecting the data analysis, and conventional pretesting and behavior coding were the only effective methods for spotting problems involving the interviewer.

Noting that the expert panel was the most cost-effective, Presser and Blair argued that questionnaire drafts should be routinely subjected to a peer review process. Their results also imply that questionnaires should be subjected to more than one form of pretesting. After reviewing an evaluation of the CPS, James Esposito and Jennifer Rothgeb (1997) concluded that the use of multiple techniques is critical to any comprehensive program of pretesting because it capitalizes on the strengths of individual techniques while compensating for the unique weaknesses associated with each.

### Summary

Survey design is indeed an art as well as an evolving science that draws upon communication, cognitive, and behavioral theories. The design of the instrument must (1) ensure effective two-way communication between the respondents and the researcher; (2) assist the respondents in recalling and clarifying their experiences, attitudes, and thoughts; and (3) keep the respondents interested and motivated. The principles and procedures discussed in this chapter are intended to facilitate these tasks.

In gathering "raw materials," the survey designer chooses among open and closed questions, direct and indirect questions, and a wide variety of response formats. The researcher may also select visual aids and will almost certainly draw upon questions developed in previous research. Open questions are most appropriate in face-to-face interviews when the researcher is unsure of response categories or is dealing with complex issues that require an in-depth understanding. They often are used effectively in preliminary interviewing to determine appropriate closed questions and response alternatives. However, the use of open questions in the final survey instrument is somewhat limited because responses are difficult for researchers

to code and require more time and effort on the part of respondents. Closed questions are best used when both researcher and respondent have ample information on the topic. While easier for researchers to analyze and for respondents to complete, closed questions are more difficult to develop, may bias results by structuring responses, and may dampen respondents' interest in the survey. Most questions in surveys are linked apparently or directly to what the researcher wants to know. However, when examining characteristics that respondents are unable or unwilling to report directly, researchers may use indirect questions that disguise the researcher's intent.

After considering raw materials, the survey designer makes a "sketch" of the instrument—an outline of topics to be included. Focus group discussions may be used to learn how people think and talk about a survey topic. The researcher decides at this point on opening questions, on the placement of sensitive and routine questions, and on the order of topics. Opening questions are critical for motivating respondents to complete the survey; therefore, they should be interesting, relatively easy to answer, and consistent with respondents' expectations. Sensitive questions are best placed somewhere between the middle and end of the instrument; routine background items are usually placed at the end. But all questions should be ordered in a way that seems natural and logical to the respondent and that does not produce unintended order effects. Transitions may improve the flow from topic to topic.

As one begins to fill in the sketch—to write the items—one should keep the objectives of the study in mind. Properly worded items are clear and precise, are written in a vocabulary appropriate for the respondents, contain a single idea, are free of emotionally loaded words and phrases, and are tactfully worded. To control the respondent's frame of reference in investigating complex topics, one can arrange questions in a funnel sequence, moving from general to progressively more specific questions, or in an inverted-funnel sequence, moving from the most specific to the most general questions. Complex issues also can be examined through reason analysis. This involves the development of an accounting scheme, based on preliminary interviewing, that identifies the key dimensions of reasons for an action and integrates them into a model of the action.

Besides the error and bias introduced by improper question wording, the survey designer must be alert to problems of forgetting and memory distortion and to various response bias tendencies. Response accuracy may be improved by shortening the recall period, by giving respondents more time to respond, by aided-recall procedures (e.g., providing a helpful context or list), or by encouraging the use of records. Among the ways to minimize the social desirability bias are careful placement and wording of questions as well as assurances of anonymity. Response set problems are best handled by varying the type of questions and response format. It is also important for the format of the instrument to be neat and appealing. To save time and prevent adverse reactions, the format might include filter questions that direct respondents to appropriate contingency questions.

The pretest-and-revision process begins early with first drafts of questions and continues until the survey is successfully field-pretested on a group of respondents similar to those in the target population. Initially, cognitive interviewing techniques ("thinkaloud" interviews, probing questions, and paraphrasing follow-ups) are used to diagnose question wording, ordering, and formatting problems. Field-pretesting techniques, including behavior coding, respondent debriefings, interviewer debrief-

ings, response analysis, and split-ballot tests, are then used to evaluate the survey instrument under realistic field conditions.

## Key Terms

open-ended (free-response) question	social desirability bias
closed-ended (fixed-choice) question	acquiescence response set
direct question	position response set
indirect question	contingency question
response format	filter question
Likert response scale	cognitive interviewing
focus group	field pretesting
double-barreled question	behavioral coding
leading question	respondent debriefings
funnel sequence	interviewer debriefings
inverted-funnel sequence	split-ballot tests
reason analysis	response analysis
response bias tendency	

## Exercises

1. Suppose you are constructing a questionnaire for the purpose of conducting a survey of sex-role attitudes. What would be the best placement (beginning, middle, end) of the following questions?
  - a. How many sisters do you have?
  - b. Does your mother work outside the home?
  - c. Mothers should put their children before themselves.
    - ( ) Strongly agree
    - ( ) Agree
    - ( ) Disagree
    - ( ) Strongly disagree
  - d. Would you say that women nowadays are more likely to work outside the home than they were when you were growing up?
2. The text identifies five common wording problems in constructing survey questions: (1) lack of clarity or precision, (2) inappropriate vocabulary, (3) double-barreled question, (4) loaded word or leading question, and (5) insensitive wording. Identify the wording problem(s) in each of the following questions and then rewrite the questions to make them more satisfactory.
  - a. How many siblings do you have? ( ) 0-2 ( ) 3-7 ( ) 8 or more.
  - b. Do you think the man should initiate and pay for the first date?
  - c. In divorce or separation cases, the man has just as much right as the woman to have custody of the children.
    - ( ) Strongly agree
    - ( ) Agree
    - ( ) Disagree
    - ( ) Strongly disagree

d. Because women are less aggressive than men, a woman's place is in the home.

- ( ) Strongly agree
- ( ) Agree
- ( ) Disagree
- ( ) Strongly disagree

e. Do you hold traditional sex-role attitudes?

f. Does your mother work?

g. Is the leadership in your family matriarchal, patriarchal, or egalitarian?

3. Suppose you want to know why students choose a specific major (e.g., sociology). Applying reason analysis, construct a series of questions to find out the reasons for students' choice of major.

4. Imagine that you want to conduct a campus survey on voluntarism. Specifically, you are interested in who volunteers, why they choose to volunteer, where they volunteer, how often they volunteer, and what kind of volunteer work they do. To reduce costs, assume you choose to do either a mail questionnaire survey or Web survey and you need to construct a survey instrument.

- a. Outline the question topics to be covered.
- b. Write at least one question or find an existing question for each topic.
- c. Give an example of a good opening question.
- d. Identify possible sensitive questions. Where will you place them in the questionnaire?
- e. Give examples of a contingency question and a filter question that you will include.

5. Once you have developed questions for your campus survey (exercise 4), pretest your questions on a small sample of potential respondents using "thinkaloud" and probing questions.

## Notes

1. For experimental investigations of the effects of including "don't know" options in opinion questions, see Howard Schuman and Stanley Presser (1981).

2. Even when the researcher has had relevant personal experiences, this should not take the place of pretesting.

3. Indirect questioning should not be confused with proxy reporting, which involves asking respondents information about others in their households or social environments (spouses, relatives, co-workers, neighbors, etc.). Proxy reporting reduces interviewing costs because information about several persons (e.g., all household members) may be collected during a single interview. In the November 1984 Current Population Survey (CPS), for example, about 40 percent of the reported presidential election participation (voting turnout) figures were based on proxy reports (U.S. Bureau of the Census, 1986:table 18). The rationale underlying indirect questioning, that respondents will perceive others to be like themselves, is a potential source of error in proxy reporting.

4. Forcing a response is likely to dampen respondent motivation and to violate typical institutional review board requirements that responses are voluntary. Instead, a Web survey could branch to a follow-up screen query to ascertain that the respondent intended to leave the previous question blank.

5. This response format is used in, but not limited to, the semantic-differential scaling technique. For details see Charles Osgood, George Suci, and Percy Tannenbaum (1957).

6. This rule applies to interviews only. As indicated earlier, open-ended questions are usually inappropriate in mail surveys and, if placed at the beginning of a self-administered questionnaire, may discourage respondents from completing the survey.

7. A test of question wording shows how the word "communist" can alter the tone of the question. Schuman and Presser (1981:284-86) compared the responses to two questions regarding support for American military intervention: "If a situation like Vietnam were to develop in another part of the world, do you think the United States should or should not send troops?" and "If a situation like Vietnam were to develop in another part of the world, do you think the United States should or should not send troops to stop a communist takeover?" When these questions were asked in separate but comparable national sample surveys in 1974, 1976, and 1978, an average of 15 percent more people favored sending troops when the communist-takeover phrase was added.

8. For certain underreported behavior, this type of question may be necessary (Sudman and Bradburn, 1982). Although the marijuana question appears to assume that the respondent smokes, a "never" answer is still possible.

9. Research supports this practice. A famous study by Donald Rugg (1941), replicated by Schuman and Presser (1981:276-78), compared responses to two freedom-of-speech questions, one using the word "forbid" and the other the antonym "allow": "Do you think the United States should allow public speeches against democracy?" and "Do you think the United States should forbid public speeches against democracy?" When these questions were asked in three separate but comparable national sample surveys in the 1970s, over 22 percent fewer people, on average, were willing to "forbid" than to "not allow" such speeches.

10. This section is adapted, in part, from the discussion in Zeisel (1968).

11. If the UCLA administration had hired us to study the impact of financial aid programs on the quality of undergraduate admissions, for example, our accounting scheme would definitely incorporate financial considerations; and the sample under study also should include UCLA applicants who are accepted but choose to go elsewhere to school. If, instead, our client were the UCLA student government, which wanted to improve programs serving the needs of first-year students, our accounting scheme would focus on these needs and would likely omit information about other schools considered.

12. Sometimes an accounting scheme guides the entire research design, not just the wording of a few questions. Nine years after the first major announcement in 1953 of a possible link between cigarette smoking and lung cancer, Bruce Straits (1967) conducted a survey to study why the health reports were apparently having only slight effects on cigarette smokers. His study was designed around the accounting scheme that this anxiety-arousing information was having negligible impact on smokers because they (1) were unaware of the reports, (2) disbelieved them, (3) were not motivated to give up smoking, or (4) were unable to translate their motivation into action.

13. In some panel surveys interviewers reduce telescoping errors by having the respondents compare reference-period recalled events with those given during the previous wave to avoid duplications (Groves et al., 2004:218).

14. Like many recent methodological advances in survey research, the EHC method is informed by theoretical knowledge in cognitive psychology, specifically on autobiographical memory. Such knowledge indicates that memory is structured in a hierarchy, with salient life events at the top and other events indexed within this hierarchy. Retrieval cues embedded within EHC interviewing take advantage of this to optimize recall (Belli, Shay, and Stafford, 2001).