

4

Survey Design and Implementation

Creating questions that elicit accurate responses from the persons in the sample requires an appreciation of both the art and the science of questionnaire design. A key assumption that underlies this effort is that people generally want to tell the truth. Survey research is predicated on the idea that lying is not a source of systematic bias. Researchers must craft response choices that make it easy for respondents to be truthful, honest, and candid.

Artistically, survey design requires a keen sense of page layout, vertical question flow, and logical question sequence, features that invite rather than deter participation. Clear, cogent questions with interesting and appropriate response choices prompt accurate, consistent responses. Moreover, the questions should interest respondents. The instrument should be easy to follow and to interpret, both for the interviewer and for the respondents. Visual appeal is especially important for mail surveys, but clear instructions and cogent questions are important regardless of the method

of contact. Clarity, simplicity, and attractiveness are the hallmarks of a professional, polished product. The dividend is a high response rate.

The science of asking questions essentially concerns whether an item is a valid measure of what the researcher wants to know about the respondents. Do the questions mean the same thing to all respondents? A great response rate to a survey with poorly worded questions is pregnant with error and only misinforms and misleads. The challenge for the researcher is to write questions that are valid and reliable measures of what he or she wants to know and to avoid things that diminish these qualities. The most serious of these are ambiguity and bias in how questions are worded, ordered, or asked.

Writing questions to measure concepts is much less exact than the statistical methods available for analyzing the results. The researcher should always be mindful that the use of sophisticated sampling and statistical techniques does not, by itself, ensure that the results will be useful or meaningful for the policy process. What matters is how questions and response choices are worded and whether these measure what the researcher intends. In *The Art of Asking Questions*, Payne (1951) recognizes the laborious but critical nature of this task. His advice is deceptively simple but still appropriate, because question wording remains the most serious defect of contemporary surveys: "We need to develop a critical attitude toward our questions . . . [and] subordinate any pride of authorship to this critical attitude" (p. 16). "As question worders, we must be sure that our meaning is comprehended by others" (p. 21).

Previous researchers have devised measures for many political and psychological concepts in social science. It is always wise for researchers designing surveys to review the survey questions developed by others on topics similar to their projects and to contact the authors of relevant surveys for copies of their instruments. Survey questions are often published in journals such as *Public Opinion Quarterly*. Indexes are published for surveys conducted by the *New York Times/CBS*, Gallup, and the National Opinion Research Center's General Social Surveys. The U.S. Census is also an excellent source for attribute questions related to employment, occupation, family, race, ethnicity, income, wealth, education, marital status, household status, and housing types.

The researcher's literature review on a survey topic should begin with articles that have appeared in public administration journals. *Public Administration Review*, *American Review of Public Administration*, and *Social Science Quarterly*, as well as specialized public policy journals, are good sources of ideas. The researcher should not assume, however, that

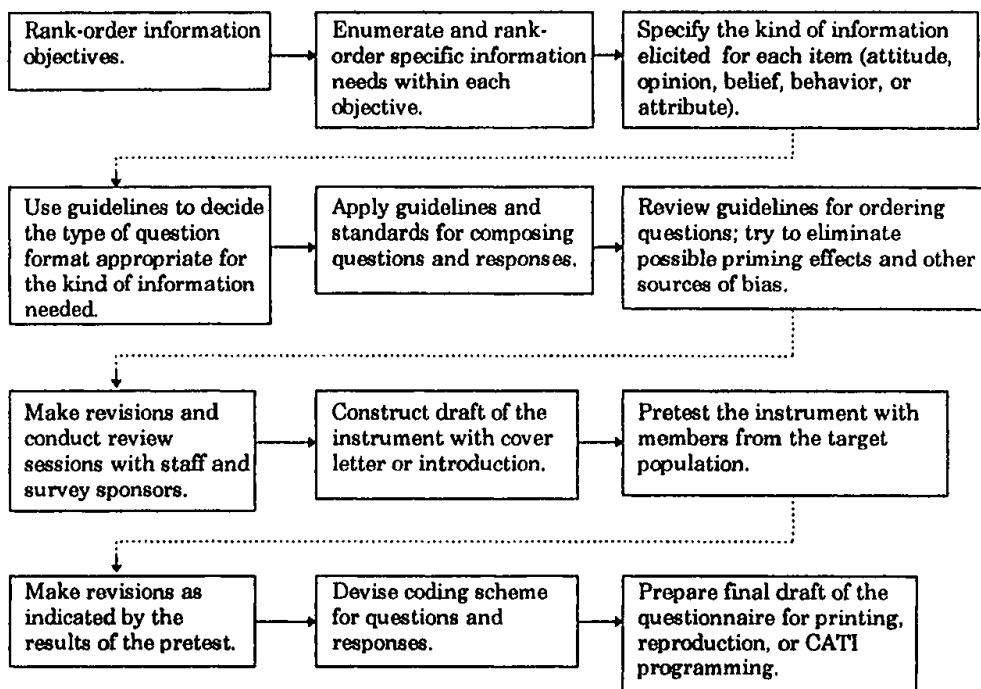


Figure 4.1. An Overview of the Survey Design Process

publication in a scholarly journal means that a question has attained a permanent state of grace with respect to either its validity or its reliability. *All* questions adapted or created for a new survey instrument must be pretested. Language evolves, and the meaning and importance of questions are affected by the context and even their location in the questionnaire (Converse & Presser, 1986). A response that something is “cool,” for example, has a positive connotation for some people. Similarly, at one time “government investment” and “contributions” were not perceived as political codes for higher spending and taxes.

The use of unambiguous words is necessary but not sufficient to avoid bias. The location of a question in the instrument affects the reliability and validity of answers to it. For instance, a well-phrased question about school performance placed *after* a series of questions about a teacher walkout or strike will yield an entirely different response distribution from the one it would elicit if it were placed before this series.

There is much to consider in this stage of the survey research process, as Figure 4.1 illustrates. The ultimate utility of a survey depends on how well the researcher applies the guidelines that constitute the art and science of questionnaire design and wording. If a single maxim could summarize this stage of the process, it is as follows: *Know what you want to ask and*

why you want to ask it; compose clear, unambiguous questions, keep the survey as brief as possible, and have a plan for analyzing the results before the instrument is administered.

The Basics of Question Order

Which structure or format is best for a question? What kind of response choices should be offered? Which type of question is best for the kind of information needed? How much should the question format vary so the respondent does not become bored? Before tackling each of these important issues, the researcher will find it prudent to think first about the general order of items in the instrument. One way to proceed from the proverbial blank sheet of paper to the first rough draft for peer review is to follow these suggestions:

1. Specify and rank-order, from most to least important, the information objectives of the survey.
2. Enumerate the kinds of information needed from respondents that relate to each information objective. Are they opinions, attitudes, beliefs, or attributes? Group related items, such as opinions about the effectiveness of a service or program.
3. Rank the items *within* each topical group in the order of their importance to the study.
4. For *each* item in *each* group, try to answer these questions:
 - Why* ask this? (How is it linked to the central purpose of the survey?)
 - Who* in the target population knows about it, and are they likely to have an opinion about it?
 - How* will the responses to this item be coded?
 - What* kind of statistical analysis will be performed with this variable? (What limits attend the variable's likely level of measurement?)
5. Place the *most* interesting item(s) in the *most* important battery of questions at the beginning of the questionnaire.

The preferred strategy is to begin the questionnaire with the most interesting and most important questions. The objective is to grab the respondent's attention, to pique interest. The initial questions should correspond very closely to the stated purpose of the study. The last section of the instrument normally contains the attribute questions. In between, the

instrument should include nothing that will confuse or embarrass the respondent.

These suggestions produce a logical outline of the order of items that reflects the central objectives of the research. Now the researcher needs to tackle some more discrete questions, such as those posed in the first paragraph of this section. Decisions about question type and format involve a basic choice between open-ended and some type of closed-ended questions. This choice is governed by the level of information about the topic possessed by the respondent *and* the researcher, the *type* of information needed, and the types of analyses the researcher wants to perform with the data.

Question Types

The two basic question types are open-ended and closed-ended. Respondents answer open-ended questions in their own words. For closed-ended questions, the researcher offers limited response choices. The researcher must review the advantages and disadvantages of each in the context of the decision about how best to structure a particular question.

OPEN-ENDED QUESTIONS

Open-ended questions are appropriate under the following circumstances:

- When the researcher has difficulty devising a reasonable number of mutually exclusive categories for a response set
- When it is reasonable to expect that the respondent has some information about or keen interest in the topic
- When it is important for the purposes of the study that respondents feel free to state, in their own words, what they think or how they feel about a particular topic, issue, or proposal
- When decision makers need to probe what people think or feel, even though no one has complete information about a particular problem, issue, or proposal
- When the item solicits nominal data, but including a list of possible answers would be too long for a mail questionnaire or too time-consuming and difficult to communicate orally
- When the researcher needs to measure especially sensitive or socially undesirable behaviors, in which case more people will respond to open-ended questions than to closed-ended questions (Sudman & Bradburn, 1982)

Under these circumstances, the *judicious* and *sparing* use of open-ended questions can stimulate the interest of respondents, generate information that no other question type can ascertain, and offer insights about issues that preconceived response choices may overlook.

Open-ended questions offer the researcher an opportunity to identify specific problems or circumstances known only to specific groups of respondents. Subdivision residents, for example, may know about the location of drainage or flooding problems. A closed-ended question about the seriousness of flooding problems in their neighborhood can be followed by an open-ended question about where these are located.

Open-ended questions are particularly well suited for *exploratory* research questions as long as respondents have some knowledge or information about the topic. The topic likewise should engage the attention and interest of the respondent. Open-ended questions, for example, can be used to ascertain respondents' knowledge about whether a local lake is safe for swimming. The results may suggest revisions in a public safety information campaign. Open-ended questions are futile if the respondents do not have *any* knowledge of or interest in a topic. Increasing the police budget may be a keen concern for the police chief, but most citizens cannot make informed judgments about whether or how much such a budget should be increased.

Open-ended questions are very demanding for respondents. They require more time, effort, and skill to complete than do closed-ended questions. Accordingly, they should be used sparingly. On the instrument for a self-administered survey, the survey designer should allow enough space for a four- or five-line response, but no lines should be inserted in the instrument, as this can constrain comments. Telephone or personal interviews receive more responses to open-ended questions when the target population has a high proportion of illiterates or persons with low educational attainment. These contact methods also permit interviewers to ask follow-up and probe questions to clarify responses.

Open-ended questions are more difficult to code than are closed-ended questions. For mail questionnaires, the researcher can read all of the responses to open-ended questions and then develop categories and codes for various answers. For telephone or personal interview projects, interviewers must record accurately the essence of the respondents' comments, and then categories and codes are devised by one or two people who make judgments about category assignments. This procedure helps to ensure consistency, and later peer review of this effort can help to ascertain the reasonableness and reliability of these decisions.

CLOSED-ENDED QUESTIONS

Closed-ended questions are accompanied by lists of possible answers from which respondents choose. Response choice lists that have no “other,” “no opinion,” or “don’t know/not sure” options are called *forced-choice* formats. Good closed-ended questions are difficult to develop, but they have several distinct advantages:

- They are easier for respondents to answer, provided the respondents are familiar with the subject and have thought enough about it to take a position or to form an opinion.
- A good response list offers a common frame of reference to maximize reliability.
- The responses are easily coded, which facilitates comparisons within or across different groups in the target population.
- Attribute and behavior questions are especially well suited to a closed-ended format because the choices can be expressed as ranges.
- Ordered response choices permit the researcher to classify individuals, measure the intensity of attitudes, and perform ordinal-level data analyses.

It can be difficult to craft closed-ended questions for several reasons. The researcher has to decide how much information members of the population have about a topic and whether they are likely to have an opinion about it. The problem of nonattitudes is very real (as discussed in Chapter 2); questions must be directed to and answered by people who are likely to know something about the subject.

The choices that make up the response set must be ones that people commonly understand and will interpret reliably. O’Sullivan and Rassel (1995) report that they once asked respondents to indicate the ethnicity of volunteers, but “only after we received several surveys indicating that ‘all’ volunteers were Native Americans did we realize that our terminology was interpreted differently by our respondents” (p. 192). There is always the risk that someone will misinterpret the meaning of a question or a particular response choice. The objective is to make the question and the options as plainly intelligible as possible. Plain language, common concepts, and clear task explanations facilitate comprehension. Careful pretesting of closed-ended questions increases the likelihood that they will be commonly understood.

A clear response set can salvage an unclear or poorly worded question. A student of mine once submitted an attribute question that read, “Please indicate the number in your household broken down by sex.” I wanted to

answer “zero,” because we are still in good shape, but explicit choices concerning numbers of males and females precluded this response. Response categories can be important cues for understanding questions.

Good closed-ended questions satisfy two measurement principles: *exhaustiveness* and *mutual exclusiveness*. Exhaustive response categories include *all* of the possible responses that might be expected. If they are inadequate, respondents may not be able to indicate their genuine opinions. An incomplete response set suggests that answers are limited to the listed options. Rational people will either skip questions that have incomplete response choices or mark something that *seems* close to their view. Non-exhaustive categories result in an instrument robust in error and low in response, because people quickly conclude that the entire enterprise is a waste of time if none of the choices applies to them.

Mutually exclusive responses offer clear, distinct, and nonoverlapping choices. Instructions to choose the “one best answer” from a list can help to eliminate confusion. Researchers should always avoid overlapping response choices, such as \$10,000-\$20,000 and \$20,000-\$30,000. Similarly, researchers should never list choices that are very close in meaning. The use of *frequently*, *sometimes*, *occasionally*, and *regularly* in the same response set is confusing. Other words to avoid using in the same response set are *usually*, *seldom*, *often*, and *rarely*. The principle is that one should avoid vague terms that are subject to different interpretations by reasonable people. Subtle differences among choices jeopardize the validity of the survey results. The respondent should feel that his or her answer fits one, and only one, possible category.

In the course of a telephone or personal interview, the respondent must understand the choices and then recall the one that applies to him or her. The number of choices in closed-ended questions for these interviews should not exceed five or six. Pretesting questions helps to ensure that the categories are clear, exhaustive, and mutually exclusive. If any interpretation problems are going to arise, the pretest stage is the time to discover them.

Closed-ended questions may have several types of response formats. The simplest and often the most appropriate set consists of *yes*, *no*, and *don't know*. For questions that require more options, investigators can choose partially closed-ended questions, closed-ended questions with ordered response choices, and closed-ended questions with unordered response choices. A *partially closed-ended* question allows the respondent to enter a response not listed by the researcher. The most probable choices are listed, but choices such as *other* and *don't know* are included as well. Researchers should use this structure when they can think of the probable responses but

may be unsure about the exhaustiveness of the options. For instance, an item such as “Please circle the number next to the one city service you think needs the most improvement” may list several services as well as an “other” choice with a blank space next to it (see question 3a in Figure 4.4 for another example).

Partially closed-ended questions should not be employed as escape valves for poorly crafted response sets, and “other” options should not be appended to every closed-ended question. They should be included only when the researcher can identify the most likely responses but needs to measure the diversity that exists beyond those choices. Researchers should be aware that most people will not take the time to write in their choices if they do not appear on the list provided (Schwarz & Hippler, 1991).

Closed-ended questions with ordered choices are especially well suited for measuring questions about attitudes, beliefs, or behaviors. Responses to questions of this type often combine to form a multiple-item index to measure some concept important to the study. Examples are “public safety,” “service satisfaction,” or “civic-mindedness.” The answer choices are a gradation of a dimension of some belief, behavior, or attitude. An “agree-disagree” response set measures opinions about specific statements. Evaluative questions may employ other types of ratings, such as “good,” “fair,” and “poor.” Figure 4.2 presents examples of closed-ended questions with ordered choices.

Closed-ended questions with unordered response choices do not limit responses to gradations of a single concept. The unordered response structure is suitable for acquiring information on opinions, beliefs, behaviors, or attributes. The responses can help the researcher to establish priorities or to rank alternatives. Occasionally, the researcher may need to have the respondent “circle all that apply” in a list. In coding these, each choice is considered to be a distinct variable and is designated 1 if selected and 0 if not chosen. Figure 4.2 also illustrates examples of unordered response sets.

Choosing the Question Type

The key factors the researcher should review in deciding how to structure a particular question are as follows:

- The objectives of the survey and the types of information needed
- The knowledge or information the researcher and respondents have about the topic

Ordered choices:**For a telephone survey:**

1. Are you inclined to agree or disagree with the following statement: "For the local taxes that I pay, the services provided by the city are a good bargain." Would you say that you:

- 1 STRONGLY DISAGREE
- 2 DISAGREE
- 3 NEITHER DISAGREE NOR AGREE
- 4 AGREE
- 5 STRONGLY AGREE

2. How would you rate the job performed by the public service department with respect to the timely removal of brush and leaves in your neighborhood? Would you say that it is:

- 1 POOR
- 2 FAIR
- 3 GOOD
- 4 NOT SURE/DON'T KNOW

Unordered choices:

1. Which of the following best describes the type of structure in which you reside?

- 1 SINGLE FAMILY
- 2 DUPLEX
- 3 APARTMENT IN BUILDING WITH 3 OR 4 UNITS
- 4 APARTMENT IN BUILDING WITH MORE THAN 4 UNITS
- 5 MOBILE HOME

2. Who do you think is most responsible for the success of the riverfront redevelopment project?

- 1 MAYOR
- 2 CITY COUNCIL
- 3 CITY PLANNERS
- 4 CHAMBER OF COMMERCE

3. Some problems in our city are listed below. Please rank these from 1 to 5, with 1 being *most* critical to 5 being *least* critical.

- ___ TRASH ON CITY STREETS
- ___ VIOLENCE IN THE PUBLIC SCHOOLS
- ___ GRAFFITI ON BUILDINGS AND WALKWAYS
- ___ LACK OF SHOPPING FACILITIES
- ___ LACK OF AFFORDABLE HOUSING

Figure 4.2. Closed-Ended Questions With Ordered and Unordered Choices

- How motivated respondents may be to communicate their experiences and thoughts
- The level of measurement of the variable and the type of data analysis planned

Closed-ended questions are less demanding for respondents than are open-ended questions, and they typically constitute 90 to 95% of all questions in an instrument. When crafting these questions, the researcher has to be careful to avoid omitting an important choice or fashioning a response set that does not correspond to how or what respondents think. If the options presented are not exhaustive or mutually exclusive, respondents will be frustrated and response rates will decline. Clear, concise, and unambiguous closed-ended questions with exhaustive and exclusive answer choices are rewarded with higher response rates.

Open-ended questions are easier to write, but they present a more challenging free-recall task for respondents. They require more time and effort both for the respondent to complete and for the researcher to code. The judicious use of open-ended questions, however, may be highly informative, especially when the researcher is uncertain about the exhaustiveness of response choices. Open-ended questions offer respondents the opportunity to express their views or thoughts in their own words. People generally will respond to open-ended questions if they concern something they care about. In some cases, placing an open-ended question near the beginning of the questionnaire or telephone interview can help to establish a “consulting” tone for the remainder of the questionnaire or interview.

Open-ended questions are often the best choice when the investigator is interested in the salience of an issue, the order in which issues are recalled, or the proportions of respondents who recall particular issues without any prompting from a list of options. Closed formats, on the other hand, may remind respondents of alternatives that they otherwise may not have considered and are appropriate when the investigator desires to evaluate a larger set of specific items, to determine the intensity of opinions, or to assess the *relative* importance of choices. Table 4.1 summarizes the major attributes of the different question formats.

Avoiding Bias in Questionnaire Design

Bias exists whenever some feature of the survey instrument or interview process leads to a response that is not a genuine measure of the respondent's true opinion, attitude, belief, or attribute. Bias may occur in the instructions, question wording, question order, response choices, or the format of the instrument. Sometimes a guess about what people think can cause fewer problems than a biased survey whose thin scientific veneer misleads decision makers.

TABLE 4.1 The Merits of Open-Ended and Closed-Ended Questions

<i>Question Type</i>	<i>Applications</i>
Open-ended	Allows respondents to answer in their own words on topics that interest them and on which they have information. Especially useful for exploratory research questions that need to probe people's preferences, priorities, and positions. Appropriate when mutually exclusive and exhaustive response choices are difficult to devise or when such a list greatly increases the complexity of a question.
Partially closed-ended	The most probable or likely choices are presented but the list cannot be exhaustive because there is reason to suspect that opinion diversity exists among a small segment of the population. This question type permits respondents to offer their own answers. Because few persons usually select this option, a large number of responses to the "other" choice may suggest a defective response set.
Closed-ended with ordered choices	Especially useful for determining frequency of participation, intensity of feeling, or degree of involvement or contact. A scale that represents a gradation of a single concept distinguishes this question type. This format is especially useful for a series of attitude and belief questions.
Closed-ended with unordered choices	Used to help establish priorities, decide on alternative policies, or enumerate behaviors as long as the choices are exhaustive and mutually exclusive.

Instrumentation bias refers to defects in the way instructions, questions, or response choices are expressed or organized. Some of the most common sources of instrumentation bias are unclear or arcane vocabulary, poor grammar, excessively specific and demanding questions, double-barreled and loaded questions, unbalanced or overlapping response choices, and reliance on a single question to measure complex concepts. Interviewers may also introduce bias, through voice inflection that suggests preferred responses or by inconsistently phrasing questions. Interviewers must be thoroughly familiar with the survey protocol and sensitive to the importance of their being courteous, tactful, consistent, and helpful.

Sometimes the wording or format of a questionnaire can lead to *acquiescence response set bias*. This is the tendency for people to answer questions in a specific direction (either positive or negative). Respondents may "agree" with the first few questions about the desirability of reducing the size of government, for example, and then respond with or mark the same answer choice on the assumption that this choice applies to the rest of the questions that deal with smaller government or reduced spending. Respon-

dents are quickly bored when they encounter too many questions with the same format, and they may superficially scan for answers they think apply, to end the ordeal quickly.

Singleton, Straits, and Straits (1993) suggest that one way to avoid acquiescence response set bias is to steer clear of the Likert scale format (the agree-disagree index) in favor of explicit response choices. Accordingly, the investigator may phrase a question in this format:

Which of the following statements most closely reflects *your* opinion about the sales tax on food?

- 1 The sales tax on food should be abolished.
- 2 The sales tax on food should not be abolished.
- 3 Not sure/don't know.

Theoretically, a researcher can detect acquiescence response bias if the instrument has two different questions on the same topic. For example, at one point, the following question may be asked: "Some state legislators propose to abolish the state sales tax on food. Would you favor or oppose removal of the state sales tax on food?" Then, later in the survey, respondents are asked whether they agree or disagree with this statement: "The state sales tax on food should be abolished." If substantively different answers are given, the instrument may be biased. A distinct drawback to this strategy is that it takes up valuable time or space and annoys people who may feel their time is being wasted by survey sponsors who ask redundant questions. A better strategy is for the survey designer to write *interesting* questions in the first place and to vary the response choices for items in a series. Use of both questions and statements to measure variables is less monotonous.

Similarly, *straight-line response set bias* may occur when a long series of questions or statements with identical answer choices appears on a page. Use of the same "agree-disagree" scale for a long list of items is a recipe for disaster. A respondent may mark the first few items accurately, find the process boring, and speed through the rest of the instrument by marking the same answer for subsequent statements. In mail surveys this response is sometimes recognizable when a respondent marks a large circle around or a line through the same response choice in the series. Varying the arrangement, structure, and format of questions, and selecting different types of questions, eliminates straight-line response set bias.

Social desirability bias occurs when respondents are unwilling to admit or to report accurately various behaviors or opinions because these are not

considered to be socially acceptable. The challenge for the researcher is to ask questions about socially sensitive issues in ways that elicit honest answers. Some people are reluctant to report their actual behaviors or opinions because they vary from socially acceptable or politically correct norms. The fear of being thought of as extreme, unpatriotic, indolent, or just plain stupid motivates people to avoid or to respond differently to various “mom and apple pie” issues, such as voting, recycling, environmental protection, child restraint seat use, and consumption of substances or foods known to cause health problems. Blunt questions on sensitive topics or issues embarrass people, and rational people want to avoid this emotion.

Social desirability bias can occur for any practice, behavior, or attitude that “conforms to the dominant belief patterns among groups to which the respondent feels some identification or allegiance” (Dillman, 1978, p. 62). Dillman (1978) suggests that this problem is most acute for personal interviews, affects telephone interviews occasionally, and is a less critical concern for impersonal mail surveys. Regardless of the method of contact, the researcher should always try to frame questions so that they will elicit undistorted, honest responses. The following are some examples of unnecessarily blunt questions that invite social desirability bias:

- Did you vote in the last mayoral election?
- About how many hours of television per day do you allow your elementary school child to view?
- Did you attend the last scheduled PTA meeting on the new special education programs offered in the Wildwood Elementary School?
- How much money did you make last year?
- Have you ever taken office supplies home with you?

Questions on sensitive topics should be phrased so that they are easy for respondents to answer truthfully. Weisberg, Krosnick, and Bowen (1989) suggest that questions about voting behavior, for example, can be phrased to assure people that nonvoting is reasonable, as in the following question:

In talking to people about the last city election, we found that some people were not able to vote because they weren’t registered, were sick, or just didn’t have the time. How about you—did you vote in the last city election? (p. 70)

Likewise, accurate information about parental attendance at a socially desirable event might be obtained by this question:

These days parents have many responsibilities and demands on their time and cannot always attend every school meeting. Did you attend the May meeting of the PTA on the new special education programs?

Employees can be asked, "How often, if at all, do you bring work home from the office?" "Do you normally use computer disks [printing paper, etc.] from the office for this homework?"

Response choices that offer *ranges* for sensitive attribute questions (income or age, for instance) are less intrusive and threatening. More people will answer frankly when asked about their annual household income if their choices are offered as ranges of dollar amounts. Likewise, some people are reluctant to report their exact ages, but they will check an age range (for example, 40-45) or report the year in which they were born.

Composing Questions and Response Choices

The central issue is how to formulate unbiased open- or closed-ended questions that will get good-quality responses. Scholars have wrestled with this issue ever since systematic public opinion polling began in 1935 (see, e.g., Belson, 1986; Biemer, Groves, Lyberg, Mathiowetz, & Sudman, 1991; Blankenship, 1940; Bradburn & Sudman, 1991; Payne, 1951; Schuman & Presser, 1981; Sudman & Bradburn, 1974). Although these studies have advanced our understanding of the survey process, two conclusions are evident: First, there is no formula for writing perfect questions and designing a flawless questionnaire; second, the survey method is an incredibly complex activity about which there is still much that we do not know. Our understanding of the social and cognitive processes that affect responses is still in its infancy.

Writing questions places dual demands on the researcher: First, questions must be unambiguous and must adequately measure what the researcher wants to measure; second, the questions must be understood to mean what the researcher intends them to mean by complete strangers who agree to engage in a voluntary social exchange. This exchange between researcher and respondent occurs within the larger context of "social norms governing relations between strangers, general canons of politeness, and ways of treating strangers" (Bradburn & Sudman, 1991, p. 31). Too often, researchers construct questions that make perfect sense to them, but that are misunderstood or misinterpreted by some respondents. To compound this error, researchers often pay too little attention to ways in which they

can increase voluntary participation and make the interview a more pleasant experience for the respondent.

The challenge for the researcher, a daunting one indeed, is to write clear, unambiguous questions that are not so long and complex as to create additional problems of their own. This tension between specificity and simplicity is inherent to the task. A number of scholars have offered guidelines that can help survey designers to craft questions that strike a reasonable balance between these two values (see, e.g., Converse & Presser, 1986; Dillman, 1978; Oppenheim, 1992; Rubenstein, 1995; Weisberg et al., 1989). These prescriptions, founded on common sense and refined by experience, can help investigators avoid egregious flaws in their wording of questions and responses. Later in this section I will review and illustrate these guidelines, but first I want to discuss an approach to the question-answer exchange relationship that can help to encourage participation and reduce the "costs" of respondents' participation.

POSITIVE INDUCEMENTS

Every survey is unique in several respects, and what really happens in the social exchange process varies among surveys. An approach to question wording and instrument design that I have found to work well consistently, regardless of the type or purpose of the survey, consists of two basic practices: a *reality check* and a *consideration check*.

First, the researcher should try to imagine being "in the respondent's shoes" and hearing (or reading) each question for the first time. Every question and draft of the instrument should be viewed from this perspective, with an eye to answering the question, How can this be misunderstood by or bore the respondent? If the researcher can conceive of *any* way someone might misinterpret, misconstrue, or take offense at a question, then the chances are fairly good that some people in the target population also will have this reaction. Removing defects in the language and structure of questions that might produce inconsistent responses or deter participation is a fundamental reality check.

The second practice is for the researcher to remember that the entire survey enterprise depends on inducing people to volunteer their time to perform what amounts to doing him or her a favor. The golden rule of survey design is to show the respondent consistent consideration and respect in every aspect of the instrument, from the cover letter or introductory spiel to the expression of appreciation at closure. Obtaining voluntary cooperation requires the practice of survey etiquette. The language used should be

polite, respectful, and understood by people in the target population. The instrument itself should be attractive, clear, and easy to follow. The cover letter or introduction should explain clearly what the investigator or interviewer is doing and why, as well as why the respondent's help is needed. A polite request for the individual's help, an *accurate* estimate of how long the process will take, and an acknowledgment of the person's permission to proceed are good ways to build rapport. The language of the instrument should be free of jargon, all words should be spelled correctly, and the vocabulary should be appropriate for the mean level of educational attainment of the target population. Transition statements should introduce batteries of questions and perhaps briefly describe why it is important for the respondent to take the time to answer them.

Packing questions densely onto a page creates the perception in the respondent that all the researcher cares about is saving a little money on paper; this communicates little regard for the respondent's task burden. Adequate white space between the vertical ordering of questions and response choices both improves their appearance and facilitates a sense of progress as items are completed. Questions that are neat, orderly, and piloted by directions are easier to follow. Consistent vertical orientation of items makes the product visually smoother. The use of a consistent method for marking responses on similar items—whether checking, circling, or underlining—helps to minimize later data entry errors.

In the interview process, respect and consideration for the respondent often make the difference between a hang-up or door slam and a successfully completed interview. Courteous, polite callers who are thoroughly familiar with the instrument and who have good listening skills can guide respondents through questions quickly, with a minimum of inconvenience or frustration. Interjecting a "please" or "would you mind" where appropriate is a simple act of courtesy that makes people feel they are being treated with civility and respect.

This approach to the social exchange that attends survey implementation requires a conscious effort to attend to all of the details that cumulatively shape a respondent's decision to participate. Building rapport founded on respect and consideration for the individual results in a qualitatively different kind of survey process that should help to increase the quantity of completed surveys. This approach is now distinctive in an era rife with obnoxious and relentless telemarketers and mail questionnaire prize scams.

The researcher may also want to offer a material inducement to potential respondents, such as a copy of the executive summary of the survey results. The idea is to offer something more than just a psychic reward for

Example of an unrevised original

In my pursuit of scholarly research: (Circle the number most appropriate, where 1 = strongly disagree and 10 = strongly agree.) Please skip question if it does not apply to you.

44. I have developed a plan for my research activities and success
Strongly Disagree 1 2 3 4 5 6 7 8 9 10 *Strongly Agree*
45. Most successful researchers have a plan for personal objectives and success, which they monitor and update as conditions change.
Strongly Disagree 1 2 3 4 5 6 7 8 9 10 *Strongly Agree*
46. Because of the serendipitous nature of research, scholarly achievements and success are mainly beyond the control of the researcher.
Strongly Disagree 1 2 3 4 5 6 7 8 9 10 *Strongly Agree*
47. I choose research topics based on likely availability of funds.
Strongly Disagree 1 2 3 4 5 6 7 8 9 10 *Strongly Agree*
48. Whether or not specific research enhances one's career largely depends on its timing and trendiness.
Strongly Disagree 1 2 3 4 5 6 7 8 9 10 *Strongly Agree*
49. Before sending an article to a journal, I carefully research what types of articles and methodologies have been published in that journal previously.
Strongly Disagree 1 2 3 4 5 6 7 8 9 10 *Strongly Agree*
50. I choose research topics primarily based on my professional interests.
Strongly Disagree 1 2 3 4 5 6 7 8 9 10 *Strongly Agree*
51. Before applying for a grant, I often lay the groundwork by checking with the funding sources and/or those sponsoring the RFP.
Strongly Disagree 1 2 3 4 5 6 7 8 9 10 *Strongly Agree*
52. I think quantity and number of publications is an important indicator of research success.
Strongly Disagree 1 2 3 4 5 6 7 8 9 10 *Strongly Agree*

This same format continues for two more pages, through question 83.

Critique

Several questions are arguably double-barreled, use redundant or unnecessary words, and pertain to topics that are apparently unrelated to the subject of the pursuit of scholarly research. Questions 45 and 48, for example, ask for information that no one person possesses. Use of the same response set is fatiguing and invites response set bias. It is also inappropriate for some of the questions. No labels or definitions are offered for values 2 through 9 in the scale. A respondent must guess whether a 5 is a neutral response or a 'neither disagree nor agree' choice.

The reliance on a 10-point Likert-type scale decreases the reliability of the instrument. How likely is it that you would choose the same number for an item two days hence? Listing the response choices horizontally is visually difficult to follow and increases the perceived task burden. Use of vertical question flow and white space reduces the perceived task burden.

Figure 4.3. Revising a Mail Questionnaire

participation. Tangible inducements may also include coupons or discounts for various services or programs. These are limited only by the imagination and the budget of the survey sponsors.

It is always easier to critique someone else's attempt to compose questions on complex subjects than it is to create questions that conform to all of the

Possible revisions**A transition statement:**

The following questions will help us to understand how you approach your research and what you think it means to be a "successful" researcher. (Please circle the numbers to the left of your answers.)

1. Have you prepared a formal plan to help you accomplish your research objectives?
 - 1 NO
 - 2 YES

2. Before you submit a manuscript to a journal, is it your practice to determine whether that journal has published articles on topics similar to yours?
 - 1 NO
 - 2 YES

3. Do you review the methodological approaches employed by previous authors before you submit a manuscript to a particular journal?
 - 1 NO
 - 2 YES

4. Which statement *best* describes your perspective on what accounts for research publication success? (Circle the number next to your choice.)
 - 1 Whether a journal accepts my manuscript is primarily a matter of serendipity.
 - 2 Which reviewers the editor decides to send my manuscript to primarily determines whether it will be published.
 - 3 If I submit a manuscript on a currently popular topic in my discipline, it is much more likely to be published.
 - 4 Whether my manuscript is accepted for publication depends mostly on whether I am able to incorporate sophisticated statistical methods that are currently in vogue.

5. If you *had* to choose *one* of the following, which do you think is a *better* indicator of a scholar's "research success" over a 10-year period?
 - 1 Publication of 5 articles in the premier quality journal in your field.
 - 2 Publication of 10 articles in several journals of good quality in your field.
 - 3 Publication of three books of original research by an academic press.

Figure 4.3. Continued

guidelines offered so far in this chapter. This is why peer review and pretesting are so important. Survey designers need practice to be able to recognize all of the possible defects that a draft instrument may contain. One of the best ways to learn how to incorporate these guidelines in one's own work is to critique the many surveys that one receives in the mail. For example, Figure 4.3 reproduces as closely as possible a small portion of a questionnaire sent to coordinators of graduate programs in public administration, along with a critique of some of the instrument's major problems and some suggestions for improvements.

Composition Standards

Simplicity, brevity, clarity, coherence, consistency, and symmetry are the main considerations in writing good questions and response sets. Some word processing programs (Microsoft Word®, for instance) can calculate a Flesch Reading Ease formula, which computes readability based on the number of syllables per word and the average number of words per sentence, and assigns scores ranging from 0 to 100. Standard writing scores average between 60 and 70. The higher the score, the larger the number of people who can readily understand the material. Such a tool can be useful as the researcher begins editorial revisions on the survey instrument.

SIMPLICITY VERSUS SPECIFICITY

Relatively short questions (with fewer than 20 words) are especially important for telephone interviews, because they minimize the task burden of the respondent. Lengthy, complex sentences and ideas should be broken up into shorter ones. Clear, concise sentences also have specific time references. If a technical term is essential, it should be defined so that no ambiguity exists about its meaning. Also, the researcher should consider asking two or three questions, instead of one, to measure the distinct elements of a complex process or concept.

For example, a manager may want to know what employees think about a new performance appraisal system. She may begin with the question, "Do you think your supervisor's judgment is affected by recency errors since the new BARS was devised for your position?" Complexity rather than length is the main concern here. Even a knowledgeable employee may not be sure that BARS is an acronym for *behaviorally anchored rating scale* or that *recency error* refers to judgments based only on recent behaviors rather than overall performance since the last evaluation. Also, will all employees know precisely *when* the new appraisal system was devised? Three questions that employ clear concepts and a common reference period might improve the quality of responses. Figure 4.4 illustrates these improvements.

The need for specificity often requires that a researcher include more complex sentences or more questions. Several drafts *and* pretesting are necessary to achieve the best balance for a particular information objective. The fruits of this toil are questions that are less likely to be misinterpreted. Examples of some first-draft efforts and improved revisions are presented in Figure 4.4.

Unnecessarily complex and difficult:

1. Do you think your supervisor's judgment is affected by recency errors since the new BARS was devised for your position?

Improved:

- 1a. Since July 1997, your supervisor has used a new rating scale to evaluate your job performance. What do you think about the fairness of this new rating scale compared to the old appraisal system? Would you say that it is

- 1 WORSE
- 2 ABOUT THE SAME
- 3 BETTER

- 1b. In thinking about your last job evaluation, do you believe your supervisor *accurately* appraised your performance?

- 1 NO
- 2 YES
- 3 NOT SURE/DON'T KNOW

- 1c. Please describe why you think your supervisor did **not** accurately appraise your job performance in the last evaluation.

Unnecessarily complex and difficult:

2. Considering your company's strategic plan for growth and increased market share, how would you expect your industrial sector's generation of hazardous wastes to change during the next five years? [Open-ended]

Improved:

2. By the year 2000, do you expect the hazardous waste generated at your plant to decrease, stay the same, or increase?

- 1 DECREASE
- 2 STAY THE SAME
- 3 INCREASE

Figure 4.4. Improving Question Clarity**CLARITY**

A researcher should not expect too much from a single question. Questions with compound subjects risk being *double-barreled*; that is, they really ask two or more questions but limit respondents to single answers. Asking double-barreled questions is among the most common errors in survey design. The unrevised Question 3 in Figure 4.4 provides an example.

Unnecessarily complex and difficult:

3. Does your household recycle PET and HDPE plastics at the community drop-off collection center?

- 1 NO
- 2 YES
- 3 NOT SURE/DON'T KNOW

Improved:

3. Do you recycle any materials at the community drop-off collection center?

- 1 NO
- 2 YES
- 3 NOT SURE/DON'T KNOW



3a. If YES, which of the following materials do you usually take to the center?
(Please circle the numbers for all that apply.)

- 1 Plastic milk bottles or jugs
- 2 Plastic soda bottles
- 3 Newspaper
- 4 Aluminum cans
- 5 Clear glass containers
- 6 Other (please specify) _____

Figure 4.4. Continued

Figure 4.5 offers additional illustrations; for instance, Question 1 in Figure 4.5 incorporates several distinct issues. It also is biased because it implies that higher teacher salaries and less spending on athletics affect school quality. As a rule, investigators should avoid *false assumptions* about linkages between concepts. Separate questions are appropriate for measuring support for more spending for teacher salaries and support for less spending on athletics and after-school programs. School quality also needs to be defined. "Better school quality" is an output from an educational system and does not necessarily equate with an input such as higher teacher salaries. Perhaps improvement in reading scores on standardized tests is a better measure. Unambiguous measures of distinct elements improve question clarity.

In Question 2 in Figure 4.5, more than one response choice can apply to the respondent or to family members. In addition, some people might think of "family" as broader than just those living in the immediate household, so validity problems arise with the use of this term. Again, there is a dubious assumption that the respondent has accurate information about the experiences of other family members. One solution would be to ask two questions and to substitute *household* for *family*.

Researchers should be careful to scrutinize any question that contains the word *and* to determine whether or not it is a double-barreled question.

Double-barreled and leading:

1. Do you favor increasing the quality of our public school system by raising the property tax to give teachers higher salaries and to offer students more after-school learning programs?

- 1 FAVOR
- 2 OPPOSE
- 3 DON'T KNOW/ NOT SURE

Revised:

- 1a. Would you support a local property tax increase to give teachers in the district's public schools a 10% salary raise?

- 1 NO
- 2 YES
- 3 DON'T KNOW/ NOT SURE

- 1b. Would you support a local property tax increase to expand after-school learning programs for children in public school grades 1 through 6?

- 1 NO
- 2 YES
- 3 DON'T KNOW/ NOT SURE

- 1c. Do you think how much money teachers make has any impact on their students' performance on standardized reading tests?

- 1 NO
- 2 YES
- 3 DON'T KNOW/ NOT SURE

A double-barreled question:

2. How many times were you or a member of your family a victim of a crime during 1995?

- 1 NOT AT ALL
- 2 ONCE
- 3 TWICE
- 4 THREE OR MORE TIMES
- 5 DON'T KNOW/NOT SURE

Revised:

- 2a. Were you a victim of a crime at any time during 1995 in the city of Mesa?

- 1 NO
- 2 YES

- 2b. If YES, did the Mesa police apprehend anyone suspected of committing the crime against you?

- 1 NO
- 2 YES
- 3 DON'T KNOW/NOT SURE

Figure 4.5. Fixing Double-Barreled Questions

Concepts should be clear and distinct. Separate questions are appropriate if their importance to the survey is justified. Question clarity suffers when the investigator assumes that respondents know as much about a topic as he or she does. Whenever any doubt exists about respondents' knowledge

A double-barreled question:

3. Was the police officer in your most recent contact helpful and courteous?

- 1 NO
- 2 YES
- 3 DON'T KNOW/NOT SURE

Revised.

3a. Was the police officer in your most recent contact helpful to you?

- 1 NO
- 2 YES
- 3 DON'T KNOW/NOT SURE

3b. Was the police officer in your most recent contact courteous to you?

- 1 NO
- 2 YES
- 3 DON'T KNOW/NOT SURE

Figure 4.5. Continued

about a topic, the researcher should insert a screen question to determine respondents' familiarity with it. This practice makes it clear that an answer is not required if respondents have no information about it. It also helps to distinguish genuine attitudes from nonattitudes. An alternative to the use of a screen question is to preface a question with a neutral explanation.

Double negatives mar question clarity. Survey designers should avoid phrasing questions in negative terms, such as "*Don't* you agree that the property tax should *not* be increased this year to cover the projected deficit?" Such questions are easily misinterpreted, and respondents may have different perceptions of what yes and no responses mean. The word *agree* in this example also suggests that people *should* agree. A good practice is to *balance* the language in attitudinal questions and to use *nondirectional* language whenever possible. Phrases such as *agree or disagree*, *favor or oppose*, and *satisfied or dissatisfied* balance a question. Nondirectional language directs attention to clear, mutually exclusive choices, for example, "Which option do you think the city should employ to avoid the projected deficit at the end of the year?"

Consistency in question interpretation is the critical objective. *Unclear or unstated criteria* open a question to misinterpretation. Questions with multiple interpretations are not valid measures. For example, "How important is it that the state maintain the existing income tax rate?" is a question that might mean different things to different people. Many citizens can be counted on to think in terms of their own monetary circumstances and answer with the choice "very unimportant," because they would like to see the tax rate reduced. Industrialists might think the current rate is good for

business and so may answer with “very important.” Still others may believe it is important to increase the tax rate to fund education at a higher level and thus may select “very unimportant.” An improved question with more explicit criteria would be as follows: “Do *you* think the state should increase, keep the same, or decrease the existing 6% tax on personal income?” An open-ended follow-up question may ask *why* the respondent holds this opinion. To maximize question clarity, the researcher must have a clear focus on what information interests the survey sponsors.

Unfortunately, there is no dictionary of “commonly understood survey terms.” Payne (1951) offers a “Rogue’s Gallery of Problem Words” (pp. 158-176) that is worth consulting, but the best strategy for a researcher to employ after making his or her best effort to draft clear questions is to pretest them on a group from the target population.

AVOIDING BIASED TERMS AND LOADED QUESTIONS

Biased questions lead people to respond in ways that do not reflect their true opinions or attitudes about an issue. Survey designers should excise any terms, phrases, and clichés that appeal to emotions or that evoke visceral responses. Examples of terms laden with emotional or ideological baggage include *cops*, *bureaucrats*, *government planning*, *justice*, *welfare*, *dump*, *liberal*, and *conservative*.

Leading questions suggest that some answers are more acceptable than others, for example, “Do you favor spending more money to make our streets safer by expanding the local jailhouse to hold a larger number of drunk drivers being convicted in our city?” This is obviously a biased, double-barreled, and leading question. Rhetoric, pretentiousness, and demagoguery are not consistent with objective question composition. The researcher should strive to be as straightforward as possible about what he or she wants to know. An improved version of the preceding question might read: “Do you favor or oppose a 2 cent increase in the property tax rate to build additional cells in the local jail?”

Questions should be nondirectional; that is, nothing in the language should suggest, even subconsciously, the desirability of one response over another. A subtle kind of misleading question is illustrated by the following: “Would you consider voting for someone other than Bill Clinton for president if the 1996 election were held today?” Well, who would not at least *consider* voting for someone else? Opinions on this question are likely to differ markedly from a more straightforward query: “If the 1996 election

were held today, how likely is it that you would vote for Bill Clinton for president?" The response choices could be "not at all likely," "somewhat likely," and "very likely."

TIME REFERENCES

An explicit *time frame* minimizes the recall burden for respondents and facilitates interpretation of responses. A mail survey typically requires 6 to 8 weeks to complete. Use of an imprecise time period in a question, such as "the past 2 months," means that those who answer the first mailing and those who respond to the third will have different time frames in mind. Specific, recent time references are easier for people to recollect accurately.

The memory burden on respondents is less demanding if response choices include ranges. Instead of asking an open-ended question such as "How many times did you use the city trolley system during 1997?" the researcher can insert "About" at the beginning of the question and then offer frequency ranges, such as "not at all," "fewer than 5 times," "from 6 to 10 times," and "more than 10 times." Ranges are not as precise as interval-level data, but they reduce the respondent's recall burden and increase the reliability of responses.

Questions should not demand extremely precise recollection or calculation by the respondent. A question such as "What is your mean net family income per week?" is difficult for most people to answer, because few people think in those terms. Instead, the question should ask for total household income for a particular year. The researcher can always calculate an average monthly income later in the data analysis stage.

SYMMETRY

Balanced response choices improve symmetry and question objectivity. Response sets should have equal numbers of negative and positive choices and, where appropriate, a middle category for those who have no opinion. Unbalanced rating scales seem to appeal to fast-food chain marketing departments, whose customer evaluation cards typically offer only "excellent," "very good," "good," and "poor" choices, so that three of the four options are obviously positive. Unbiased response choices are always symmetrical and balanced. Figure 4.6 shows some questions with unsymmetrical and symmetrical response choices and suggests other possible balanced response sets.

Examples of unbalanced response choices:

1. How would you rate the food at Nebraska Bob's Raw Oyster Bar?

- 1 EXCELLENT
- 2 VERY GOOD
- 3 GOOD
- 4 FAIR

2. Overall, how satisfied are you with the quality of education in the county school system?

- 1 VERY SATISFIED
- 2 SATISFIED
- 3 SOMEWHAT SATISFIED
- 4 DISSATISFIED

Examples of balanced response choices:

1. How would you rate the overall quality of food that you ordered this evening at Che Rivera?
(Please check one.)

- 1 EXCELLENT
- 2 GOOD
- 3 AVERAGE
- 4 POOR
- 5 VERY POOR

2. Overall, how satisfied are you with the quality of education in the county school system?
(Please check one.)

- 1 VERY SATISFIED
- 2 SATISFIED
- 3 NEITHER SATISFIED NOR DISSATISFIED
- 4 DISSATISFIED
- 5 VERY DISSATISFIED

Examples of other balanced response sets:

<ul style="list-style-type: none"> excellent good neither good nor poor poor very poor 	<ul style="list-style-type: none"> increase stay the same decrease 	<ul style="list-style-type: none"> too much about right too little 	<ul style="list-style-type: none"> strongly agree agree disagree strongly disagree 	<ul style="list-style-type: none"> good fair poor
<ul style="list-style-type: none"> desirable somewhat desirable somewhat undesirable undesirable 				

Figure 4.6. Balancing Response Choices

Question Order for Mail Surveys

Question order affects responses in a variety of ways, only some of which are known to scholars. In between the gripping questions that lead off the questionnaire and the more sensitive items that bring up the rear,

Copyright © 1996, SAGE Publications, Inc. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

much can happen to shape responses in ways the researcher should *try* to anticipate. In particular, the investigator should consider how preceding questions might direct or influence answers to later questions. The objective is to order questions within topical sections so as not to bias later responses.

Question order should not create *priming effects* for later questions in a series. For example, the researcher should avoid asking a series of questions about local crime and perceived threats to public safety before asking the respondent to rank in order of importance a list of five problems, one of which is "the local crime rate." The respondent in such a situation will be sensitized to the crime issue and may rank it higher than he or she might have otherwise. A strategy to help rectify this problem is to use a *funnel sequence* for questions. When topics are already important to respondents, the researcher should start with *general* questions and then funnel to more specific ones. General rankings or ratings of various problems, goals, or programs before specific items on these topics minimize respondents' sensitization. This avoids the artificial elevation of a particular item in the mental response hierarchy of a respondent. For example, several questions about recycling followed by a question about what citizens can do to help their community will result in recycling being mentioned more frequently or ranked more highly than if this general ranking question preceded specific questions about this practice. When topics are not particularly stimulating, the researcher should start with specific questions to provide respondents with a frame of reference and then ask more general questions about the topic. This is often called an *inverted funnel sequence*.

There are always potential problems in obtaining consistent rankings and ratings of items in a series. Research in the fields of advertising and sociology suggests that some respondents assign more importance to items listed first in any series. Because the positions of items may affect their likelihood of being selected or endorsed, interviewers conducting personal or telephone interviews can be instructed to modify the order in which items are read. It is usually prohibitively expensive to print different versions of self-administered surveys, but items in such instruments can be presented in alphabetical order, or statements can be included that indicate the items are listed in random order.

The most sensitive questions in the instrument should be placed *at the ends* of their respective sections. Attribute and background questions usually make up the last section of the questionnaire, with the most sensitive questions, such as those concerning income, race, and political party identification, located at the end of this section. The rationale is to obtain responses

to the most critical questions before the respondent gets turned off by queries on more sensitive subjects.

Constructing Indexes and Scales

A single question is insufficient for measuring complex concepts such as socioeconomic status, participation, freedom, quality of life, safety, service satisfaction, government responsiveness, business climate, fiscal stress, employee morale, knowledge, faithfulness, piety, and ideological orientation. Several questions are necessary to capture the multiple dimensions involved in such concepts. Multiple indicators of a concept provide more valid, reliable, and precise measures, and results are not affected as much by the wording of a single question. For example, if a researcher asks people what their socioeconomic status is, he or she is likely to be met with blank stares. A better way to ascertain this information is to create a composite measure by asking questions that respondents can answer about the three variables researchers typically use to define this concept: income, education, and occupation.

Many composite measures already exist, and some are useful for the kinds of surveys conducted by public administrators. As I have noted, researchers should always review the published research on the topics in which they are interested, as well as the indexes constructed by national polling organizations. However, advances in our understanding of social and political phenomena often stem from better, more innovative measures of complex concepts. If a researcher can create an index that is more appropriate than those used in the past for the information needs of a particular project, he or she should not feel shackled by those already published.

An *index* is a composite measure of scores on individual items that relate to a specific concept. Indexes abound in daily administrative life. The FBI's Uniform Crime Index, the Department of Labor's Consumer Price Index, the Dow-Jones Index of stocks, the "misery index" (popular near election time), and even a student's grade point average are examples of composite measures of variables that attempt to measure complex concepts.

Constructing an index is a straightforward task that involves the following steps:

- Define the concept clearly.
- Choose the indicators for the concept.

- Combine the items to create an index.
- Weight items in the index (optional).

To define a concept means to specify its key elements or components. Social scientists derive these from theories, practice, and previous research. A concept is measured by indicators or index items that serve as the operational definition of the concept. Indicators that are logically related to a concept have *face validity*, or a commonsense linkage with the concept. The index is constructed by assigning a range of possible scores to each item. The sum of the scores on the individual items is the composite index score for a case. Weights can be assigned to indicators that are more important or influential for an index of a concept.

For example, there are several indicators of a “downtown’s image.” Questions can measure an individual’s positive or negative orientation to each indicator. Adverbs modify the intensity of opinions. Respondents may be “very satisfied,” “satisfied,” “neither satisfied nor dissatisfied,” “dissatisfied,” or “very dissatisfied” with items. Similarly, they can be asked whether they agree or disagree with statements about each item. Possible indicators of a downtown’s image are the cleanliness of sidewalks and streets, accessibility, traffic congestion, the appearance of storefronts, the aesthetics of signs and buildings, the convenience of parking, and a sense of personal safety when walking downtown. Each indicator should add something to a person’s composite image of a downtown. Figure 4.7 gives examples of a few of the questions that measure these indicators.

The items of an index are chosen carefully to represent the most important dimensions of a concept. Balanced questions or statements control response set bias. Changing the “direction” of items forces respondents to consider their answers more carefully. In Figure 4.7, Questions 1 and 3 are positive statements, whereas Questions 2 and 4 are negative. A higher value is assigned if a respondent agrees with 1 and 3, and a lower value is assigned if the respondent agrees with 2 and 4. This coding scheme assigns higher scores to more favorable attitudes about the downtown’s image. Cases with a less favorable image will have lower scores. Alternatively, the values for a statement can be recoded in the data analysis stage rather than appear in reverse order as shown here for purposes of illustration.

How many response choices are appropriate for an item in an index? It is desirable to capture the full range of opinions or orientations on a subject, but there is also a need to have a sufficient number of cases in each response to make analysis of the results meaningful. A large number of response choices make the questionnaire more complex and time-consuming. The

Instructions. Please circle the number next to your answer to each question.

1. I am usually able to find a parking space in a convenient location when I travel downtown.

- 1 STRONGLY DISAGREE
- 2 DISAGREE
- 3 NEITHER AGREE NOR DISAGREE
- 4 AGREE
- 5 STRONGLY AGREE

2. The streets and sidewalks downtown are usually dirty and littered.

- 5 STRONGLY DISAGREE
- 4 DISAGREE
- 3 NEITHER AGREE NOR DISAGREE
- 2 AGREE
- 1 STRONGLY AGREE

3. I feel safe when walking in the downtown area.

- 1 STRONGLY DISAGREE
- 2 DISAGREE
- 3 NEITHER AGREE NOR DISAGREE
- 4 AGREE
- 5 STRONGLY AGREE

4. Most of the buildings that border Main Street appear to be run-down and poorly maintained.

- 5 STRONGLY DISAGREE
- 4 DISAGREE
- 3 NEITHER AGREE NOR DISAGREE
- 2 AGREE
- 1 STRONGLY AGREE

Figure 4.7. Items in an Index for Perceptions of a Downtown's Image

compromise can be as straightforward as a yes or no response, or as complex as a five-part agree-disagree scale. Usually, fewer response options are preferred.

The researcher may decide that some indicators are more important than others and deserve more weight in the computation of the index. Normally, items will have equal weight unless there is some historical, theoretical, empirical, or other reason to believe that one or more of the items are significantly more important in shaping a person's standing or position on some index. An example is an index of a person's environmental activism. A monetary contribution to an interest group, for instance, might be weighted more heavily than just support for stricter air pollution control. In the survey's final report, the researcher should describe the reasons for the weighting of various items in the index.

For a simple additive index, each response to an item is assigned a value. The sum or the mean of the item values constitutes the index score. For

example, a hypothetical measure of environmental activism might consist of four questions about membership in an environmental protection organization, the regular practice of solid waste recycling, monetary contributions to an environmental preservation cause, and participation in an organized demonstration about an environmental issue. The responses to each question (or variable) in the index may be coded as no = 0 and yes = 1, depending on whether or not the person checked any of the listed behaviors. Each case has a potential index score of 0, 1, 2, 3, or 4, depending on the number of items in which the respondent engaged. If the researcher considers participation in an organized demonstration to be a more important indicator of activism than the other activities listed, he or she could assign a yes response a weight of 2. The *score weight* products are then added to compute the index score. In this case, a score may range from 0 to 5.

What if a respondent does not answer all of the questions that form an index? How are these missing data handled? Nonresponses are coded 9 or 99, or any number other than the one for a legitimate answer. The frequency distributions indicate the extent of a missing data problem for each question. There are several strategies that researchers use to deal with missing data. If only a few cases have missing data, these cases may be excluded from the analysis. If one of the indicators in a scale has an extraordinarily large number of nonresponses, it can be excluded and a new operational definition of the concept presented. Another, less satisfactory, method is to compute the average score for a case on the questions that have been answered and then assign this mean to the unanswered question(s). Still another strategy is to assign the middle score or the mean score of an item to a case that has not answered the question(s). With large samples, the first strategy is preferred. For a survey with a small *n* or response, the exclusion of all cases with missing data may bias the findings. Regrettably, there is no highly satisfactory solution to the problem of missing data, but the methods listed above are in rough order of preference.

To validate an index, it is necessary to determine whether each of the items in it contributes to the explanatory power of the composite measure. *Item analysis* is the method used to ascertain whether each of the items is correlated with the composite measure or index score. In item analysis, the index score is the independent variable and the item is the dependent variable. A measure of association (described in Chapter 6) indicates whether an item is strongly associated with the index. If it is not correlated with the index, that item should be dropped from the composite measure. If the items all correlate strongly with the index, then it is a good measure of the concept.

A commonly used index is the *Likert scale*. This index, incorrectly termed a scale, presents respondents with statements about which they are asked to agree or disagree. Usually the choices consist of a five-part agreement scale, as illustrated in Figure 4.7. The items in an index are ordinal measures. In other words, they indicate “more or less” of something but not *precisely* how much between cases. The composite scores rank cases on a particular concept.

Technically, one should use only the statistics appropriate for the level of measurement of the data. In practice, some scholars use interval-level statistics for composite scores from Likert scales to compute means or correlations as part of the data analysis stage. Empirical purists reject this practice, but the debate about it continues. In practice, this means that some analysts assume that a score from an ordinal index is an interval-level measure. This practice is ill-advised, but if the researcher employs it, he or she should be very cautious about imputing any great policy significance to fine distinctions among respondents’ “mean” scores.

Indexes can improve the accuracy and validity of the measurement of complex concepts, but researchers should avoid the overuse of Likert scales. Whenever possible, it is better to employ simple yes or no response choices, because these provide accurate and reliable measures of what people think about the various dimensions that constitute concepts.

Designing the Mail Questionnaire

The appearance of a mail questionnaire is especially important. The pages of the questionnaire should have plenty of white space to make it easier to read and to allow the respondent to follow the vertical flow of the questions. The goal is to make the questionnaire inviting, pleasing to the eye, and easy to complete. Concise instructions should accompany different types of questions, and transition statements should lead off each cognitively related group of questions. Several suggestions to help investigators prepare an appealing instrument are listed in Table 4.2. Figure 4.8 shows some hypothetical questions that conform to these criteria.

THE COVER LETTER

A good cover letter encourages participation, and a poor one does little to persuade respondents not to toss the questionnaire in the trash. Altruistic appeals for participation work, but the inducements discussed earlier, such as offering to send the respondent an executive summary, enhance the

TABLE 4.2 Guidelines for Constructing Mail Questionnaires

-
- Never cram too many questions onto a page so that it looks dense and cluttered. This increases the respondent's perceived task burden.
 - Whenever possible, do not break a question between pages.
 - The instrument length generally should not exceed 12 page faces.
 - Lead off the instrument with an interesting, easy-to-answer question that is clearly related to the purpose of the survey as explained in the cover letter. An open-ended question asked early often helps to establish a consulting tone.
 - Vary the types of questions to avoid response set bias. Long series of "yes or no" and "agree or disagree" questions fatigue respondents and lower the rate of completed surveys.
 - Incorporate the coding scheme into the question response sets to facilitate data processing.
 - Provide sufficient space for the respondent to write answers to open-ended questions, but do not provide lines, as this may constrain any comments.
 - Use a legible type in an easily readable font size. Capitalization of response choices often helps to focus the interviewer's and respondent's attention on them.
 - Allow sufficient space between questions and response choices to minimize coding and data entry errors.
 - Establish a vertical flow for questions and response choices. This helps to prevent inadvertent omissions and facilitates the perception of making rapid progress through the questionnaire. This technique automatically creates white space on the instrument.
 - Group questions by topic and use transition statements between topics to enhance the flow and continuity of the questionnaire.
 - Place the most sensitive questions at the ends of their respective sections. Attribute and background questions should usually constitute the final section.
 - Always thank the respondent for taking the time to complete the questionnaire.
 - Always pretest the questionnaire on a small sample of the target population.
 - Always place identification numbers on the survey instruments to track responses. This practice is essential for the direction of follow-up mailings.
-

credibility of the social exchange. The cover letter should be printed on the official stationery of the organization responsible, to lend legitimacy to the survey. It should be signed (not signature stamped) by the official in charge of the project. People are more likely to respond to surveys conducted by organizations or persons they know about. Dillman (1978) offers several excellent suggestions for composing a cover letter; Table 4.3 summarizes these ideas, and Figures 4.9 and 4.10 provide examples of cover letters for first and second mailings.

THE MAIL QUESTIONNAIRE BOOKLET

Questionnaires prepared in booklet form are attractive and reduce the number of sheets needed. There are several options for preparing questionnaire booklets, but some general recommendations include using paper

1. First, we would like to know what you think is the single *most* important problem facing the city. Please tell us in the space below what you think that most important problem is:

2. Considering the services that you receive, do you think that your city property taxes are:

- 1 TOO LOW
- 2 ABOUT RIGHT
- 3 TOO HIGH
- 4 DON'T KNOW or NOT SURE

3. Please take a few moments to rate the quality of each of the following city services. Please *circle* the number that corresponds to your rating of that service's overall quality.

Service	Quality Rating					
	Very Poor	Poor	Average	Good	Very Good	Not Sure/ Don't Know
Crime Control	1	2	3	4	5	6
Fire Protection	1	2	3	4	5	6
Pothole Patching	1	2	3	4	5	6
Garbage Collection	1	2	3	4	5	6

4. Some people say that on-street parking on Main Street downtown should be eliminated to improve traffic flow and reduce congestion. Other people say that on-street parking on Main Street should remain for the convenience of workers and shoppers downtown. How do you feel about this issue? (Please select the one choice that best fits your opinion.)

- 1 On-street parking on Main Street downtown should be eliminated.
- 2 On-street parking on Main Street downtown should not be eliminated.
- 3 Don't know/Not sure

Figure 4.8. Mail Questionnaire Page Format

with at least some recycled content and printing the pages with a laser printer. High-quality photocopying on both sides of all pages reduces costs. Off-white, light blue, or light green paper enhances type legibility.

The designer can use a mock-up to determine how the page faces should be numbered so that they are in the correct order when the printed full-size pages are folded and stapled to form a booklet. For example, there are 12 page faces to be numbered if three sheets of paper are needed for all of the

TABLE 4.3 Guidelines for Composing a Cover Letter

-
- The letter should be brief and concise; a length of three or four paragraphs is ideal.
 - The letter should be printed on official stationery, should be dated, and should be signed (not signature stamped) by the official responsible for the project.
 - Explain the purpose of the study in the first paragraph. Communicate the importance and utility of the study in the second paragraph. Assure the reader of the confidentiality of responses in the third paragraph. Thank the prospective respondent for taking a few minutes to help with the project in the final paragraph. Inclusion of a "deadline" for completion and return of the questionnaire is optional.
 - Avoid using worn-out phrases that may turn off respondents (e.g., "Enclosed is a questionnaire"; "This is a survey"; "I am conducting research"; "You are important to our study").
 - Do not inject bias either for or against particular officials, organizations, or issues.
 - If the budget permits, offer to send a copy of the survey results. Ask respondents to mark a space provided on the questionnaire to indicate their interest in receiving an executive summary.
-

questions. The front cover should include the title of the survey and the official seal of the organization responsible for the project (or some other attractive design). The remaining 11 pages will contain the survey questions. When the printed pages are laid flat, before they are folded and two staples are put in place in the spine, the bottom sheet will have the cover page and page 11 on the back side and pages 1 and 10 on the front. The middle sheet will have pages 2 and 9 on the back and pages 3 and 8 on the front. The top sheet will have pages 4 and 7 on the back and pages 5 and 6 on the front. When the pages are folded, the booklet's pages will be in order. Questionnaire booklets can be made using either letter size or legal size paper.

Implementing the Mailing

Publicity about a survey in appropriate media outlets may result in an improved response rate. Press releases can explain the purpose of the survey. Internal office memos may introduce a forthcoming questionnaire for employees. If a sample is selected, not everyone in the target population will receive the questionnaire, but those who do may recall a reason to distinguish the questionnaire from junk mail if they have heard something about the survey already.

Before the envelopes are stuffed, the researcher should check to make sure that each booklet has an identification number and that this number is



The Town of Farragut
David Farragut
Mayor

May 7, 1997

Dear Ms. Jones:

The town staff is developing a budget proposal for next year and we need your help! What you think about how the council should allocate citizens' hard-earned tax dollars is *very* important to us if we are to provide the kind and quality of services that *you* want. Please take about 10 minutes from your busy schedule to tell us what you think about funding priorities and how we might be able to serve you better.

Your household is one of a small number randomly selected to participate in this project. All of your responses will remain completely confidential. No names or addresses will *ever* be connected with them. The identification number in the upper right-hand corner only helps us to confirm that you have responded, so that you won't be bothered with follow-up mailings.

Thank you for helping us with this project! Your ideas and opinions matter a great deal to us. If you would like to receive a copy of the results, just check the box on the back of the booklet and I'll see that a copy is sent to you.

Sincerely,

Joe Smith

Manager
Town of Farragut

Office of the Manager • 11320 Kingston Pike • Farragut, Tennessee 37922

Figure 4.9. Example of a Cover Letter for a First Mailing

recorded next to the name and address of the household or individual on the master list of the sampling frame. Returned, completed questionnaires should be logged on this list. The mailing package should consist of the cover letter, the questionnaire booklet, and a postage-paid self-addressed return envelope that is large enough to accommodate the booklet. Mailings should take place early in the week and all on the same day. As mentioned in Chapter 2, the use of business reply-type return envelopes can save some



The Town of Farragut
David Farragut
Mayor

May 22, 1997

Dear Ms. Jones:

About two weeks ago I wrote to you to ask for your help in establishing funding priorities for next year's town budget. Many citizens in our sample have responded so far, but we have not yet received *your* response, which is *vitally* important for the accuracy and completeness of our study.

Please take just a few minutes to answer these questions that will help us serve you better. I am enclosing another copy of the questions for your convenience. Please return them in the postage-paid envelope by May 29. I'd be happy to send you an executive summary of the results if you check the box on the last page.

Thanks so much for your help and timely action on this request. If your original response is already in the mail, you may recycle this second copy.

Sincerely,

Joe Smith

Manager
Town of Farragut

Office of the Manager • 11320 Kingston Pike • Farragut, Tennessee 37922

Figure 4.10. Example of a Cover Letter for a Second Mailing

mailing expenses, because the post office will charge the account only for the questionnaires actually returned. This also eliminates any temptation some people may have to peel off the stamp for their own use and throw away the questionnaire.

About 3 weeks after the original mailing, a second mailing should be sent to nonrespondents. In this mailing, a revised cover letter, with the new date, should courteously remind the recipients that their responses (which are vital to the study's accuracy) have not yet been received. Depending upon the response rate, a third mailing should be sent about 2 to 3 weeks after the second mailing. The cover letter in the third mailing should plead for the recipients' cooperation, underscoring the importance of their participation. If the budget permits, copies of an executive summary of the survey results may be offered to those who complete the questionnaire. If it is possible to acquire phone numbers for the procrastinators, the researcher may find that making personal requests for their help in completing the mail survey is productive. Time, effort, and attention to detail should pay off in a good return rate.

Designing the Telephone Survey

A well-designed mail questionnaire will not necessarily facilitate a good telephone interview. For telephone survey instruments many of the same guidelines apply with respect to clarity, simplicity, consistency, question order, and pretesting, but telephone surveys require attention to the needs of *three* audiences: the respondent, the interviewer, and the data entry personnel.

In a telephone interview, all communication with respondents is verbal and therefore, clarity is paramount. The respondent does not read a cover letter, examine the questions, and then decide whether or not to respond. Consequently, the vertical flow of questions and the visual attractiveness of the instrument are not important. However, the nature of the social interaction between the caller and the respondent is critical. Everything the caller does should help to establish a good rapport and a clear understanding of the questions.

The interviewer must explain clearly the purpose of the survey and indicate whether the survey is confidential or anonymous in nature. *Anonymous* responses have no identifying name or identification number; the identity of the subject is unknown even to the researcher. An assurance of *confidentiality*, in contrast, means that "the researcher agrees to limit access by others to data that can be linked to the participant" (Kimmel, 1988, p. 86). A confidential rather than anonymous interview enables the researcher to contact the same people in the future or to confirm that the interview actually took place, as a control on cheating by interviewers (Weisberg et al.,

1989). For these reasons, some researchers prefer to guarantee confidentiality rather than anonymity of responses. The case number permits the researcher to identify the respondent, but the respondent's name, affiliation, or any other identifying information is *never* linked publicly with his or her responses. The researcher must strictly limit access to the secret file that lists names and case numbers for purposes of follow-up contacts.

The dependence on oral communication in telephone interviews requires that they include less complex questions with fewer response categories than can be used in mail surveys. They should also make use of screen or branching questions to reduce the memory burden for respondents. Liberal use of transition statements and clear instructions are essential to communicate what is expected of respondents. Interviewers should read the questions at a moderate pace to give respondents time to think. It has been found that interviewers who have the lowest refusal rates are those who vary their voice pitch when reading questions (Rubenstein, 1995). Respondents find this mode of speaking more persuasive and credible than delivery of questions in a monotone or in a constant high-pitched voice. The interviewer's voice should be a little above normal volume and nonhesitant. *Key-word* summaries of response choices build in redundancy that helps respondents to understand the questions.

Telephone interviewers require questionnaires that are presented in a clear, easy-to-follow format for all questions. Pretesting of the questions is useful for detecting possible tongue twisters that can impede a listener's comprehension and test the caller's patience. The instructions in the instrument should supplement the training regime and help the caller to make smooth transitions between items. Regular placement of questions and response choices on the page helps to reduce interviewer error in marking responses.

The physical design of a telephone interview should be centered on the caller's needs. For this reason, questions should not be continued on separate pages, nor should too many questions be placed on a single page. The sheets must be easy to mark, and adequate space should be provided for interviewers to record open-ended responses.

The ability of the caller to establish credibility and rapport with potential respondents will determine the success of these social interactions. Callers must sound friendly, polite, and sincere. Patience, professionalism, empathy, and assertiveness are all desirable traits in a telephone interviewer.

Callers should read all questions precisely as they are written, without variations in wording or phrasing. Callers must also avoid any intonation that may implicitly suggest preferred responses. A telephone interview is not an occasion for "educating" respondents about desired behaviors or

opinions preferred by the caller. The interviewer may at times, however, need to clarify questions that have been misunderstood, for example, "From your response, I think you must have misunderstood what we meant by a capital crime. We did not mean to refer to the corruption in the state house, but whether legislators should expand the types of crimes subject to the death penalty."

Telephone interviewing requires patience and endurance. Fatigue is a factor that should be considered. A clear, consistent coding scheme (discussed in the next chapter) will help to alleviate the strain somewhat. Generally, callers who complete four 10-minute interviews per hour have good productivity. Those who consistently contact fewer than half of the numbers on a call sheet may need additional training.

THE INTRODUCTORY SPIEL

The key to success for most telephone interviews is the introduction, which should be designed to appeal to a particular target population. The primary objectives of the introductory spiel are to persuade respondents to participate, to legitimate the survey, to build trust that the responses will remain confidential, to explain why certain information is needed, and to suggest the benefits of participation.

No single introduction is perfect for every survey, but most introductions should include the following elements:

- A greeting and the caller's complete name and affiliation
- Brief information about the purpose of the call
- A description of the selection procedure and the strict confidentiality of responses
- Information about the approximate length of the interview
- A request for permission to begin, or a request to set up an appointment to conduct an interview with the desired respondent in the household

Good callers are sensitive to potential refusals and will try in a polite way to persuade recalcitrant contacts. In an effort to salvage an interview, a caller may say something such as, "I realize that I may have called at an inconvenient time; may I call back at another time that's better for you?" or "Your opinions are really important for our study, and we want to make this as easy for you as possible. Would you prefer that I call back at a more convenient time?" If the response to such a suggestion is affirmative, the caller can then establish the time when he or she should call back. Once agreement is obtained, the caller should explain the voluntary nature of the

An introductory spiel.

Hello, my name is Grant Neely, and I'm calling for the University of Tennessee Social Science Research Institute. We're interviewing citizens of Farragut to determine what they think about crime in the city and what actions city officials should take to prevent or reduce various criminal activities. Your phone number, [555-5555], was selected at random from a list of all city residents. Is this number correct? [IF CORRECT, CONTINUE. IF NOT, POLITELY TERMINATE.]

The questions I need to ask will only take about 7 minutes to complete, and your responses will be completely confidential, okay?

Examples of alternative screen and selection criteria:

A screen question: First, do you reside within the corporate limits of Farragut?

[IF NO, POLITELY TERMINATE.]

Example of respondent selection criteria: Are you 18 years or older?

[IF NO, ASK TO SPEAK WITH A PERSON IN THE HOUSEHOLD WHO IS 18 OR OLDER.]

Example of the last birthday selection method: To make sure our results accurately reflect the opinions of all city residents, I need to speak with the person in your household over 18 who had the most recent birthday. Are you that person?

[IF YES, BEGIN QUESTIONS. IF NO, ASK TO SPEAK TO THAT PERSON OR SCHEDULE A TIME TO CALL BACK WHEN THAT INDIVIDUAL WILL BE AVAILABLE AND REPEAT SPIEL THEN.]

Figure 4.11. Basic Format for an Introductory Spiel and Alternative Selection Criteria

interview with a statement such as, "Should we come to a question that you choose not to answer, just let me know and we will move on to the next one." Figure 4.11 shows a hypothetical introduction to a telephone interview and alternative selection criteria.

QUESTION WORDING AND ORDER FOR TELEPHONE SURVEYS

It is essential that the researcher take special care in crafting questions for telephone surveys because a misunderstanding of just one or two words will affect the validity of responses. The first few questions are especially

Example of question that includes a key-word summary:

1. First, I need to ask whether you think certain crimes are a problem or not in your neighborhood. How about VANDALISM? Do you think it is "not a problem," "a minor problem," or "a major problem" in your neighborhood?

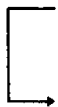
How about AUTO THEFTS? Do you think they are not a problem, a minor problem, or a major problem in your neighborhood? How about HOME BURGLARIES?

Matrix question format:

Crime	Not a Problem	Minor Problem	Major Problem	DK/NS
VANDALISM	1	2	3	4
AUTO THEFTS	1	2	3	4
HOME BURGLARIES	1	2	3	4

2. The city council is considering a proposal to establish a municipal police force. Would you support or oppose the idea of establishing a city police force for Farragut?

- 1 NO [Go to 3.]
- 2 YES [Go to 2a.]
- 3 NOT SURE/ DON'T KNOW [Go to 3.]



2a. To pay for the annual cost of establishing a city police force, would you oppose or support the establishment of a city property tax with a rate of \$1.50 per \$100 of assessed value?

- 1 STRONGLY OPPOSE
- 2 OPPOSE
- 3 NEITHER OPPOSE NOR FAVOR
- 4 FAVOR
- 5 STRONGLY FAVOR

Figure 4.11. Continued

important, because these are the "hooks" that help to convince the respondent that the survey is important and pertains to an interesting subject. If these questions flow easily from the introduction, a sense of order is projected that can be maintained throughout the interview.

Both Dillman (1978) and Frey (1989) recommend using a simple closed-ended question to lead off, followed by an open-ended item. Whatever

question type is chosen, there should be no ambiguity about its connection to the purpose of study. Questions in a telephone interview should abide by the normal rules of conversation, which means the researcher should strive for a balance between simplicity and specificity. Rapport is disrupted if questions are too complex. A question should be explained to a respondent if the need or request arises. This strategy suggests a more active role for the interviewer than traditionally has been the case, and it underscores the need for the survey manager to brief callers thoroughly on the "purpose of questions and what is being measured" (Bradburn & Sudman, 1991, p. 36).

Frey (1989), Dillman (1978), and Payne (1951) suggest additional principles researchers should apply when deciding on question wording and order:

- Group questions by topic and use transitional statements to introduce survey topics.
- Think about how the order of questions might affect later responses.
- Vary the question structure where possible to hold respondents' interest and to reduce fatigue, which usually occurs after about four questions of a similar type.
- Make sure that all response categories are balanced and mutually exclusive.
- Avoid breaking a question between pages and use sufficient spacing between items to enhance readability of the instrument for the callers.
- Use lowercase type for questions and uppercase type for answers and interviewer instructions.
- Write clear instructions for all screen or filter questions.
- Pretest the questionnaire on a small sample of the target population.

Pretesting the Questions and the Instrument

Pretesting questions is one of those hallowed practices of survey research honored too often in the breach (Converse & Presser, 1986). This is unfortunate, because pretesting is a critical quality-control device. It permits the researcher to discover just how well the survey instrument works. This is the opportunity to identify and correct problems with question wording, questionnaire structure, or administration. The time spent planning and pretesting the instrument is directly related to the quality of results.

The researcher should resist the temptation to implement the questionnaire immediately after consensus is reached among the staff and survey

sponsors about what to ask and how to ask it. Such consensus should not be considered to be chiseled in stone. Questions that make perfect sense to everyone in the office may confuse respondents and evoke unexpected answers. In this event, the validity of such items is doubtful. Questions examined individually may appear satisfactory, but they might be misplaced in the questionnaire, disrupt the flow of an interview, or create priming effects and position bias. Moreover, questions that *read well* may *sound* confusing to a respondent during a telephone or personal interview. Confusion can be created by homophones such as *cash* and *cache*, *no* and *know*, and *profit* and *prophet*. Also, interviewers might find some words difficult to pronounce.

What should researchers look for in a pretest? Fowler (1993) suggests that supervisors listen to telephone interviews with 20 to 50 respondents drawn randomly from the target population. *Each* question should be evaluated on three criteria:

- How easily the caller can read the question as worded
- Whether respondents understand the question consistently
- Whether respondents answer the question accurately with the response choices provided

Problems occur when interviewers do not read each question as worded, respondents regularly ask for clarification of questions' meanings, or respondents give inadequate or inappropriate answers (Fowler, 1993). A simple "problem-no problem" rating can be used to evaluate questions. If problems occur for given questions in more than 15% of the interviews, "those questions are either highly likely to produce distorted data or distinctively susceptible to interviewer effects" (Fowler, 1993, p. 102). Interviewer debriefings and the observed difficulties should indicate to the researcher which questions need to be revised, relocated, or deleted.

Pretesting self-administered questionnaires is even more challenging than pretesting telephone questionnaires. Instruments that will be self-administered merit especially close scrutiny during pretesting because they offer no opportunity to clarify questions' meanings or to probe for more complete responses. A two-stage procedure is recommended. First, a focus group session provides an excellent opportunity for the researcher to observe and record the time it takes individuals to complete the questionnaire. Participants can be asked about the clarity of instructions and the meanings of questions. The researcher can then correct any problems in the questionnaire that lead to difficulties in understanding or answering items

prior to the second stage of the pretest. In the second stage, the researcher mails questionnaires to 20 to 50 persons in the target population. This exercise can give the investigator some idea about what type of return rate to expect and which questions are skipped or misinterpreted.

Converse and Presser (1986) offer additional guidelines for evaluating specific questions and the instrument as a whole. Tests for specific questions include variation, meaning, task difficulty, and respondents' interest and attention. Tests for the questionnaire as a whole include examination of question flow, question order, skip patterns, and completion time.

By pretesting the instrument, the researcher can ascertain whether a question measures an acceptable level of *variation* in the target population (Converse & Presser, 1986). There may be too many or too few response choices, depending on the distribution of results. If hardly anyone selects the extreme categories of a response set, then perhaps the number of choices in the scale can be collapsed. If individuals fill in or state preferences that are not offered, the choices can be expanded. If everyone selects the same response, then the question may be retained, dropped, or structured differently.

One of the most important purposes of a pretest is to clear up any confusion about the *meaning* of a question. People expect survey questions to be reasonable. They tend to "transform obscure questions into ones that seem sensible from their standpoint as they strain for meaning" (Converse & Presser, 1986, p. 57). This makes it difficult for researchers to detect misinterpretations, but one way to do so is to compare responses on similar items, as described above in the section on acquiescence response bias. For example, a seemingly innocuous question such as, "How many times during 1996 did you visit the county courthouse?" may create problems. A suspicious person might interpret this to be a subtle inquiry about his or her criminal activity if he or she thinks only about the court cases that are heard at the courthouse. Other reasonable people might think of a "visit" as a guided tour that tourists take. If both types of respondents answer "never" to this question, yet also indicate in response to another question that they paid their vehicle registration fees in person at the courthouse, there is an obvious misinterpretation. Focus groups can detect these types of problems.

Task difficulty refers to questions that tax a respondent's memory or present an unfamiliar context. Memory distortion, or "telescoping," is common when persons try to answer questions concerning events that took place over an extended period. Questions should incorporate measurement units that are familiar. Few people think about their income in "net terms"

or about the economic status of the households in which they were raised in terms of a choice between classes such as “poverty,” “near poverty,” “working class,” “middle class,” or “affluent.” Clear, mutually exclusive categories should be employed with the use of labels that most people understand.

No doubt stimulating questions help to hold a *respondent's interest and attention*, but the fact is that most of the subjects in a typical survey are hardly likely to increase anyone's respiration rate. Short questionnaires with varied question types can help to reduce boredom and monotony. The pretest should help the researcher identify those questions and portions of the instrument where respondent interest or attention flags. Focus group participants can rate how interesting they find the instrument as a whole.

A sensible, logical *question flow* includes brief transitions to introduce batteries of interesting questions on different topics or issues. Following the guide for question order discussed above, the respondents in the pretest can be asked how they view the connection between the leadoff questions and the stated purposes of the questionnaire. The pretest also gives the researcher the chance to correct problems with *skip patterns* dictated by interviewer instructions, which direct interviewers to vault over or to include certain sections, depending on previous responses.

The actual *time* required to complete the instrument can be observed in the pretest. Researchers should never rely on how long it takes them to complete an instrument in simulations. In practice, it always takes longer for members of the target population to complete a questionnaire.

Training Callers and Implementing the Telephone Survey

Publicizing a forthcoming telephone survey in a community helps to reduce the number of hang-ups and refusals. Press releases to local newspapers and radio stations can stimulate stories that describe the purpose of the survey and who is conducting it. The precise days when calls are scheduled to begin should not be publicized, however, in order to avoid the possibility of sabotage by certain groups or individuals in a community.

The time required to complete a telephone survey project depends on the size of the sample, the number of callers and phones available, the length of the questionnaire, the skill and experience of the callers, and the target population. Calls should be placed at times when it is most likely respondents can be reached. Typically, fewer than half of all interviews are

completed successfully on the first dialing. Interviews should be scheduled for the days and times when the highest proportion of calls are completed on the first dialing. Usually, these times are after 5:30 p.m. on weekdays, 10:30 a.m. to 5:00 p.m. on Saturdays, and 1:00 p.m. to 8:00 p.m. on Sundays. Researchers should avoid scheduling interview calls on holidays, football Saturdays, and other special-event days unique to certain jurisdictions. Calls made only during normal business hours tend to reach mostly retirees, homemakers, and college students. The number of interviews that can be completed per hour will depend upon time lost in dealing with busy signals, answering machines, nonresidential numbers, refusals, and making appointments for callbacks. The more screen questions included in the instrument, the fewer the number of completions possible per caller per hour.

An “in-house” telephone survey can employ volunteers (either public employees or others) as long as they are reliable individuals who have a pleasant demeanor and clear diction. Telephone interviewing is demanding work, and attracting and retaining good callers usually requires payment of fair compensation—more than just minimum wage. It is imperative that the researcher train all callers thoroughly. Each should understand the purpose of the interview, the meanings of questions, how to remain non-directive and unbiased in reading questions, how to handle all screen and branching questions, and how to record respondents’ answers on the interview sheets (or how the software does that if a computer-assisted telephone interviewing system is used). Callers also must understand the protocol for making callbacks and for recording the disposition of attempted contacts on the call sheets. In addition, they should know how to respond to the most probable types of respondents’ questions. Typically, these include questions such as the following: Who is paying for the survey? Who is in charge of the survey? How did you get my number or name? How can I be sure that my answers will be confidential? Are you going to ask me for money? Can I get a copy of the results?

A supervisor should monitor interviewer performance and be available to answer any peculiar questions that arise. Permitting callers to conduct interviews from their home telephones is not recommended. The supervisor is responsible for distributing the call sheets to each interviewer for each shift (giving workers fewer call sheets than they can complete during a shift is one way to impart a sense of progress). The supervisor should track the disposition and progress of the calls and inspect the accuracy and readability of the completed questionnaires. Prompt, thorough oversight of telephone interviews helps to reduce errors and avert potential disasters.

The callers should sign their call sheets so that the supervisor can review each person's work and assess individual callers' productivity.

Training telephone interviewers is similar to training interviewers who conduct face-to-face interviews. All questions must be asked in the proper sequence and precisely as written. Interviewers should speak slowly, clearly, and with well-modulated voices (Dillman, 1978; Singleton et al., 1993). A good training strategy is for the researcher to explain the project's purpose, process, and embedded survey instructions and to conduct *mock interview sessions* in which supervisors assume the roles of caller and respondent. Then, each of the callers practices the role of interviewer, with supervisors playing the role of respondent. This is an opportunity for the callers to experience some of the typical curves that will be thrown at them during actual calls, such as refusals, misunderstandings of questions, abusive attitudes, and unclear or incoherent responses to open-ended questions. An investment of a few hours in this type of training will yield large dividends in later consistency and accuracy. Callers will have more confidence and enthusiasm, make fewer mistakes, and appreciate how vital it is to obtain cooperation from those contacted.

Summary

Composing good questions is the key to the achievement of survey success and requires the researcher to balance specificity and simplicity to obtain valid and reliable responses. Clear, succinct, and relatively brief questions are usually preferred. The use of open-ended or closed-ended question formats depends on what information the researcher needs to know and whether exhaustive and mutually exclusive response categories can be devised. The investigator should recognize problems with double-barreled questions, false assumptions about respondents' knowledge, double negatives, loaded phrases or terms, unclear criteria, unbalanced response options, and possible sources of social desirability bias. Questions on sensitive topics require special craftsmanship and should be asked near the end of the survey. Valid opinion measures for complex concepts can be achieved through the construction of indexes. A logically organized and attractively designed instrument is a fundamental courtesy and signifies respect for the respondent's time.

The order and presentation of questions will affect the response rate and the operational validity of the questions. For a mail survey, presentation of questions in an interesting, visually pleasing way should be complemented

by a cover letter that persuades potential respondents to complete the instrument. For telephone and personal interviews, training the survey personnel and supervising their work is essential for the realization of the advantages of these methods of citizen contact. Callers who establish rapport and communicate questions clearly, consistently, and accurately are more productive and make fewer errors on instruments designed especially for their use.