CHAPTER 2

Writing Questions

The goal of writing a survey question for self-administration is to develop a query that every potential respondent will interpret in the same way, be able to respond to accurately, and be willing to answer. However, in practice, producing good questions is often difficult.

A methodologist friend once described the question-writing task as similar to driving in freeway traffic while drinking a cup of hot coffee and answering an emergency call on his cell phone. Many things are competing for attention, and failure to heed any of them can spell disaster. It is the need to consider many competing things at once that makes it difficult to write questions for self-administered surveys. But unlike freeway driving while tending to coffee and telephone calls, which is not advisable, many things need to be considered when writing questions. Moreover, there is no alternative to understanding and dealing with each of the competing concerns that is addressed in this chapter.

Consider, for example, responses to the question, “How many hours per day do you typically study?” asked in a mail survey of university students (Rockwood, Sangster, and Dillman, 1997). As shown in Figure 2.1, different response categories, a high set and a low set, were used. Five of the six categories in the low set cover the same range of hours (less than 2¼ per day) as that covered by only one of the six categories in the high set (Figure 2.1). Use of the low set resulted in 23% of the respondents reporting that they studied more than 2¼ hours per day, compared with 69% when the high set was used. Clearly, the answers to this question were influenced by more than the words used to describe each response category.

Two other samples of students were interviewed by telephone for this experiment, with half being presented with each set of categories. Among those presented with the lower set of categories, 42% (compared to 23% by mail) reported studying more than 2¼ hours per day. Thus, the choice of surveying by mail versus telephone also influenced people’s answers. These results could have placed the surveyor in the unenviable position of reporting to the study sponsor that as few as 23% or as many as 70% of the students at this university were studying more than 2¼ hours per day, depending upon which version of the questionnaire and survey method she wanted to believe. Unfortunately, the measurement challenge did not end here.

Another, nearly identical question included in these experimental surveys asked about how many hours each day students watched television. For this question the alternative sets of categories had similar effects on responses; 17% reported watching television 2¼ hours or more per day in response to the low set, compared with about 32% when the high set was used. However, in contrast to the question about studying, whether the survey was done by mail or telephone made absolutely no difference in how many people selected each answer category.
Why did these results for the alternative sets of categories and the two questions differ so dramatically? Which answers should be trusted? In addition, how can you ever be sure that whatever questions are asked in surveys obtain answers that provide the best possible estimates of the distribution of the characteristic of interest in the population? These issues and more are the topics of this chapter. In the course of this discussion, I will return to the likely reasons that these questions produced such divergent answers. The purpose of this chapter is to provide principles, as well as specific tools, to help you write questions for self-administered surveys in ways that will produce answers you can trust.

CRITERIA FOR ASSESSING EACH SURVEY QUESTION

Survey questions fail in their purpose for many reasons, ranging from use of the wrong words or an inappropriate structure to simply not being answerable. In addition, a question that will work fine in one survey may not be satisfactory for another. Attempting to help hundreds of people, ranging from first-time surveyors to experienced methodologists, with their surveys leads me to conclude that too often sponsors have not figured out what they want to know from respondents, except in a general sort of way: “I want to know why people have such negative attitudes about . . . .” Other sponsors know fairly precisely the questions they want to ask, and still it is very difficult to progress from a list of draft questions to a set of good survey questions. Some questions are easy to write and ask, while others require draft after draft and much testing. Once a working draft of proposed survey questions and response choices (if any) has been prepared, ask each of the following eight inquiries about each of the proposed survey questions. Answers to these questions will help diagnose problems and guide you towards the structural and wording decisions that are appropriate for your study.

Question 1: Does the question require an answer?

A survey question is more than a general inquiry. It is the surveyor’s tool for gaining responses from subjects in a survey sample that will make it possible to determine the distribution of a characteristic (an attitude, belief, behavior, or attribute of each respondent) in the survey population. In order for an inquiry to constitute a survey question, it must require an answer from each person to whom the question is asked. Neither of the following questions meet that criterion.

If you fixed dinner at home last night, did you eat meat as part of that meal?

- Yes
- No

When you go out to eat, which type of food do you most prefer?

- American
- Italian
- Chinese
- Other

A respondent who ate out the previous night could not answer the first question. A respondent who never eats out could not answer the second one. Use of the introductory words “if” or “when,” invites some respondents not to provide an answer. Inasmuch as not all respondents are required to provide an answer to the above question, we cannot distinguish nonresponse from those to whom the question does not apply. Even if a “does not apply” box were provided, the wording of the first question implies that no response is needed from those who ate out the previous night. In addition, for us to be able to estimate the distribution of a characteristic in the sample and (through it) the population, respondents must be given the opportunity to answer every question they are asked.

Question 2: To what extent do survey recipients already have an accurate, ready-made answer for the question they are being asked to report?

Some survey questions are easier to get accurate responses for than are others. For example, virtually everyone knows their age. People are frequently asked how old they are (birthdays are special events in the lives of most people) and age is a number that people are expected to know. People are also able to report age accurately in response to many ways of asking the question, as shown in Figure 2.2. Assuming a willingness to report age, it does not make a lot of difference how the question gets asked. From the standpoint of expected precision of the answer, the question with three broad categories will, of course, obtain a less precise answer. However, from the perspective of whether the age question will be answered accurately, the choice of question structure makes little difference.

However, consider the other question shown in Figure 2.2, concerning whether tall people are more likely to be elected president. Most people do not have a ready-made answer that they can immediately report. Some might wonder what is meant by “more likely” in the stem of the question. Others might wonder what the difference is between “somewhat” and “strongly” in the answer choices. To answer, they not only have to give definition to the question and answer choices, but they may also try to recall the height of current and past presidents and the relative heights of their opponents. Some might spend a long time and think about a lot of presidents, while others think
Figure 2.2 Examples of questions to which respondents already have a ready-made answer (which produce accurate answers) versus one that asks about an informed opinion that produces inconsistent answers.

**Respondent's age; the exact wording makes relatively little difference in response accuracy:**

33. What is your current age?
   
   _Years_

33. When were you born?
   
   _Day_ _Month_ _Year_

33. What is your date of birth?
   
   _Day_ _Month_ _Year_

33. Which of the following age category describes you:
   
   - [ ] 35 or younger
   - [ ] 36-65
   - [ ] 66 or older

**Respondent's opinion on issue for which wording changes produce substantial inconsistencies in response choice:**

"Tall people are more likely to be elected President of the United States."

Do you strongly agree, somewhat agree, somewhat disagree, or strongly disagree with this statement?

(Or)

Do you very strongly agree, strongly agree, agree, disagree, strongly disagree, or very strongly disagree?

(Or)

On a scale of 1 to 7, where 1 means entirely agree and 7 means entirely disagree, use a number to indicate how strongly you agree or disagree.

only about the current president’s height in relation to the opponent in the last election until a satisfactory answer is reached. Still others may not think about any specific presidents or candidates at all.

In contrast to the near automatic answer for a question about age, when people respond to broad opinion questions they must comprehend the question and answer choices and contemplate what each of them means. They must recall information about the topic, make a judgment about the information they retrieve from memory, and then report a response, a process that has been described in detail by Tourangeau (1992). Each of these stages is influenced by different contextual considerations and processes. People do not have ready-made responses to opinion and belief questions to nearly the same degree that they possess answers to questions about their educational level, the kind of car they drive, or whether they own their own home. The vaguer the question, the vaguer the categories (e.g., using only numbers, as in a 1-7 scale), and the more remote these items are from people’s experiences, the more likely a question is to produce inconsistent responses if we ask the same person to answer this question at different times.

Sometimes it is necessary to spend a lot of time drafting, writing, and testing alternative wordings of questions, and even then we can obtain only an approximate answer. In other instances, writing questions becomes a very easy task, requiring little time and effort.

Thus, a critical step towards writing good questions is to understand the extent to which respondents have an accurate, ready-made answer, and whether creating an answer demands considerable thought that is subject to myriad influences, including the context in which the question is asked. If people do not have a ready-made answer to a question, getting an accurate answer becomes more difficult.

**Question 3: Can people accurately recall and report past behaviors?**

Whereas abstract beliefs and attitudes are hard to measure, it would seem on the surface that people should be able to report past behaviors. However, that is not always the case.

Asking how many hours a person watched television on the first Sunday of the previous month is an example of a behavior that is hard to remember. Respondents are unlikely to be able to recall something so precise from that long ago. To solve this problem, surveyors often ask how many hours per day a person “usually” watches television, an example of which was provided at the beginning of this chapter. To answer the question, the respondent must recall what she “usually” does and estimate. Recalling the frequency of these routine or mundane behaviors is subject to considerable potential error, as evidenced by the effect of the category choices offered as part of the survey question (Rockwood et al., 1997). Asking people to reconstruct how much time they have spent studying or watching television in the last four hours, on the other hand, is less subject to unintentional error.

Frequently, people who write surveys want respondents to provide far more detail about past behaviors than can be recalled. Thus, determining whether people are able to recall information needed for answering each proposed survey question is important. Keeping recall simple and related to recent events helps to produce high quality survey data.
QUESTION 4: IS THE RESPONDENT WILLING TO REVEAL THE REQUESTED INFORMATION?

The fact that respondents know an answer does not mean they are willing to provide it. Many respondents are reluctant to reveal certain information about themselves, for example, their income. Others may be unwilling to answer questions about previous drug use, or having shoplifted when they were teenagers.

Considerable evidence suggests that people are more likely to give honest answers to self-administered than to interviewer questionnaires (e.g., de Leeuw, 1992; Fowler, Roman, and Di, 1998; Aquilino, 1994). For example, when asked the question, “How often have you driven a car after drinking alcoholic beverages?” only 52% responded “never” to the self-administered questionnaire versus 63% for the comparable telephone survey (Dillman and Tarnai, 1991).

Although self-administered questionnaires are often selected because of respondents’ greater honesty with their answers, there is little doubt that social desirability is somewhat of a problem for this method as well. A recent article suggests that young people responding to computers may provide more truthful answers than with self-administered questionnaires (Turner, Ku, and Rogers, 1998).

QUESTION 5: WILL THE RESPONDENT FEEL MOTIVATED TO ANSWER EACH QUESTION?

Motivation to respond distinguishes self-administered from interview surveys dramatically, and in a way that counts. Designers of interview surveys often write questions independent of any motivational considerations, leaving it to the interviewer to encourage, cajole, or otherwise persuade respondents to carefully select and report complete answers. Unless consideration is given to how people react to questions in self-administered questionnaires, instructions may be ignored and incomplete answers given. Also, the question may be skipped or, even worse, the questionnaire not returned.

For example, questionnaires are sometimes constructed in a way that requires respondents to consult a separate instruction booklet to understand unclear questions. Questions are sometimes presented in matrices with row and column headings that must be connected in order to understand the question. Further, it may be impossible for the respondent to understand the question without consulting a separate instruction booklet to determine what is meant by each row and column heading. Thus, three widely separated items of information have to be connected by the respondent in order to figure out an appropriate answer. Another kind of question with adverse motivational qualities is one that asks people to rank a large number of items from top to bottom, such as a list of 20 “priorities for economic development.”

Motivation can be encouraged in many ways, ranging from incentives and follow-up reminders to respondent-friendly questionnaire design. In some instances, the questions themselves are the source of a motivational problem and no matter how much one does with other aspects of survey design, wording remains a major impediment to accomplishing the survey objectives. Sometimes, modifying questions (e.g., stating them more clearly) or changing a ranking question to a rating question (e.g., asking the importance of each of the 20 economic priorities on a scale) will improve the likelihood of getting an answer.

QUESTION 6: IS THE RESPONDENT’S UNDERSTANDING OF RESPONSE CATEGORIES LIKELY TO BE INFLUENCED BY MORE THAN WORDS?

Responses to the questions about studying and watching television presented at the beginning of this chapter were clearly influenced by more than words alone. Some respondents appeared to see a set of categories that were viewed as running from low to high. Choice of the highest category may have seemed appropriate to respondents who saw themselves as in the “top” group of TV watchers or studiers. Respondents faced with a situation in which they don’t have an obvious answer may respond to these questions partly in terms of where they see themselves in relationship to other students, for example, “I study more than most, so I should choose one of the top categories.” It is important to recognize that category ranges and visual layout (as discussed in the next chapter) also provide important clues used to select the appropriate answer. Words are only part of the question stimulus.

Attitudinal and belief questions typically rely on vague quantifiers, such as strongly favor to strongly oppose, high priority to low priority, agree to disagree, or even vaguer ones that rely on numbers such as -3 to +3, 1-7, or 1-10. Such numerical scales require respondents to give a certain amount of definition to any category they choose to use. The vaguer the question and answer categories, the greater the potential for measurement error.

QUESTION 7: IS SURVEY INFORMATION BEING COLLECTED BY MORE THAN ONE MODE?

Increasingly, more than one survey mode is used to collect information for a single survey. This means that data collected by each mode needs to be comparable with that collected by another. An illustration is the fact that 42% of the telephone respondents versus only 23% of the mail respondents chose more than 2½ hours in response to the hours of study question posed at the beginning of this chapter (Figure 2.1).
It appears from the available research that several different kinds of mode differences may occur between self-administered and interviewer surveys (Dillman, Sangster, Tarnai, and Rockwood, 1996), and that the introduction of electronic survey technologies such as Interactive Voice Response, as discussed in Chapter 11, introduces even more challenges (Srinivasan and Hanway, 1999). However, several remedial actions, ranging from restructuring questions to changing question wording, can be done to minimize mode effects. For this reason, Chapter 6 is devoted entirely to the special challenges of mixed-mode surveys.

Question 8. Is changing a question acceptable to the survey sponsor?

Questions with recognized defects cannot always be changed. Sometimes a particular question has been used in another survey, and the main objective is to replicate the previous survey or make the new data comparable in some other way. Examples are government surveys that have asked the same question repeatedly in order to produce time-series data, sometimes for decades. In other instances the surveyor may be willing to accept a lower response rate or higher item nonresponse in an effort to get the more precise answers only some respondents are able to provide. This is a tendency of some economic surveys, when offering broad income categories is considered an unacceptable alternative to requesting exact income. Often sponsors want information about which respondents have virtually no knowledge and for which formulating a meaningful answer is difficult. Political considerations may also dictate the selection of question and answer categories. Thus, it is important to ask sponsors whether questions that appear troublesome are subject to change and if so, how much.

Writing questions is a difficult challenge precisely because so many factors simultaneously influence whether a proposed question obtains accurate answers. The eight questions listed here constitute the mental checklist I attempt to apply when, inevitably, interviewers asking for help thrust a draft in front of me and ask, “What do you think?” The search for tools and principles for writing questions is a means of overcoming each of these potential problems. This is the topic to which I now turn.

Which Question Structure Is Most Appropriate?

Fundamentally, there are only three different ways a survey question can be structured. One way is to pose a query as an open-ended question, an item for which no answer choices are provided. The others provide answer choices which, in turn, be structured as closed-ended in either of two ways—as ordered or unordered response categories. Shifting from one structure to another is the most fundamental tool available for responding to the kinds of problems suggested above.

Open-Ended Questions

An inability to get adequate answers to open-ended questions is often identified as a chief disadvantage of self-administered surveys. After all, there is no interviewer to respond to an unclear answer with a probe, such as, “Could you provide a little more detail so that I’m sure I understand what you mean?” However, the issue of whether questions that don’t offer answer categories are acceptable for mail surveys is somewhat more complicated. Consider, for example, this open-ended question:

A. Why did you purchase a new automobile?

Reason for purchase

Answers to Question A are likely to be inadequate. This question will not elicit self-administered answers that are as complete as those offered in interview surveys when supported by interviewer probes. People don’t necessarily have a ready-made answer to such a question in the way they do for such things as age, and the answers are prone to considerable unintentional error. Although the above question format could be improved somewhat by providing more space, perhaps 2-3 lines to visually suggest that a longer answer is desired, answers are not likely to be as complete as one would like. The fundamental problem with questions of this nature is that the answer depends upon the extent to which respondents are willing to think hard about the question and write a complete answer. In such cases a surveyor would be well advised to list possible reasons for buying a car and turn the question from open-ended to closed-ended.

When we have proposed that kind of solution, two counter arguments are often extended. One is that surveyors do not know what the possible reasons might be, so they cannot list them. Second, the question is viewed as “exploratory,” that is, an attempt to get some idea of what reasons might exist, which in a future survey could be built into closed-ended formats. Thus, despite the shortcomings of this type of open-ended question, it is sometimes appropriate to use in self-administered surveys.

Question B is quite different from Question A.

B. What kind (make and model) of automobile do you currently drive?

Make

Model of car

Most people can provide the make (e.g., Ford) and, to a lesser extent, the model (e.g., Taurus) of the car they drive. To list all possible makes and models would require an enormous amount of space and require that respon-
dents search the list carefully to find their car. This type of open-ended question works quite well in self-administered surveys. I find it helpful when asking such questions to provide blank spaces followed by the units for reporting (here, make and model) rather than simply printing an unlabeled space.

At the beginning of this chapter, I noted the biasing effects that result from offering response categories for mundane characteristics such as hours of study or watching television. Asking these questions in an open-ended form with a blank space followed by units, for example, "_______ Hours/Day," is a better way to obtain accurate answers.

Frequently, when designing self-administered surveys we have been asked to elicit the respondent's occupation, as in Question C below.

C. What is your current occupation?

________________________ Occupation

This question produces unacceptably vague answers. For example, a "building engineer" might be someone with a college degree in engineering or one who performs custodial services. When faced with the need to obtain occupational information in self-administered surveys, one solution is to break the question into multiple parts, as shown here:

What is your current occupation?

________________________ Current occupation

What kind of work do you do?

________________________ Kind of work

What is the name of your employer?

________________________ Name of employer

Other specific queries, such as the amount of education one normally needs to perform this job, might be added to assure sufficient information for deriving an occupational code. In essence, this format builds in the kinds of probes that an interviewer might ask to get adequate information for understanding a respondent's occupation.

In sum, open-ended questions are frequently very useful in self-administered surveys, but their usefulness depends upon the nature of the questions as well as the way in which they are structured. Sometimes that means changing to an open-ended format, while in other instances it means avoiding it.

Closed-Ended Questions
Consider these two questions:

A. To what extent do you favor or oppose expanding the number of franchises operated by this company?

☐ Strongly favor
☐ Somewhat favor
☐ Neither favor or oppose
☐ Somewhat oppose
☐ Strongly oppose

B. Which one of the following do you think should be our company's highest priority?

☐ Opening more franchises
☐ Improving remuneration for current employees
☐ Expanding our product line
☐ Improving skills of existing employees through training

The first of these two questions provides respondents with a response scale and the answering task is to determine where one best fits on that scale. Thus, it consists of carefully ordered categories. The second question presents categories in no particular order, and respondents are asked to pick the one that best describes their opinion.

The mental effort needed for answering each of these questions is quite different. Whereas the first question requires envisioning a scale and figuring out where on that scale one fits, the other requires comparing each discrete category with the others, a task that usually is more difficult. Typically, questions with ordered categories consist of two basic concepts, with one presented in the stem of the question (expansion of franchises) and another in the answer choices (favor versus oppose). Questions with unordered categories may consist of many different concepts which must be evaluated in relation to each of the others. In addition to the question stem concept (top priority) the answer categories reveal four distinctly different concepts (alternative business goals) for consideration. Typically, closed-ended questions with unordered categories are more difficult for respondents to answer because of the amount of information that must be processed and about which decisions must be made. However, each type can be enormously useful in a self-administered questionnaire. They also provide quite different writing challenges.

Closed-Ended Questions with Ordered Response Categories
This type of question is most useful when one has a well-defined concept for which an evaluative response is wanted, unencumbered by thoughts of alter-
native or competing ideas. The list of scalar concepts that might be used to evaluate a concept or idea seems almost endless. Here are a few possibilities:

- Strongly agree to strongly disagree
- Very favorable to very unfavorable
- Excellent to poor
- Extremely satisfied to extremely dissatisfied
- High priority goal to low priority goal
- A complete success to a complete failure
- A scale of 1-7 (or 1-10 or 1-100) where 1 means lowest possible quality and 7 (or 10 or 100) means highest possible quality
- A scale of -3 to +3 where -3 means completely lacks this characteristic and +3 means completely exhibits this characteristic

These answer choices are frequently referred to as “vague quantifiers,” a moniker that tells a lot about their measurement characteristics. Generally, these types of scales request answers the respondent may not have ready-made, and which are therefore subject to considerable measurement error. However, when a surveyor wants to obtain separate respondent evaluations of many different concepts (e.g., 20 possible areas for spending agency funds) and compare preferences across areas, there may be no alternative to this approach.

Using closed-ended, ordered categories of this nature involves making many decisions. One such decision is how many choices should be offered. A second decision is whether to label each choice and, if so, what to call it. The nature of these decisions and factors associated with making a decision can be illustrated by the experiences of two researchers who became dissatisfied with a community satisfaction measure then in common use (Andrews and Withey, 1976). The traditional four-choice scale shown below was quite common, but because most people tended to be “very satisfied,” the clustering of answers into a single cell limited the analyses they could perform. The researchers extended the scale from four to seven answer categories, and relabeled them as shown in alternative B.

A. Traditional Measure of Community Satisfaction
   How satisfied are you with your community?
   - Very satisfied
   - Somewhat satisfied
   - A little satisfied
   - Not at all satisfied

B. Proposed Measure of Community Satisfaction
   How satisfied are you with the community where you live?
   - Delighted
   - Mostly satisfied

The researchers were successful in dividing people’s answers across more categories and accomplished their analytic objective, but they also produced a scale that has not been used much in the ensuing years. I suspect the reason for the lack of use is that the scale labels strike most potential users as a little strange. At first glance the scale appears to convey two distinct concepts, degree of satisfaction plus the delighted-terrible extremes, an inconsistency that one should try to avoid. Moreover, the combination seemed to be unclear to respondents. The intended use of responses may also make a difference with regard to the suitability of these items. If one plans to present results to community officials for discussions on community goals, there is some risk that the terms delighted and terrible might not be taken seriously in that discussion.

In order to appreciate another potential difficulty with this scale, I suggest that you pick up the telephone and ask a friend to allow you to interview him or her. Read the above questions out loud and then assess how easy it is for the scale to be read and comprehended, and whether you think the question is likely to be taken seriously. The movement towards telephone interviewing in the last few decades has encouraged people to use somewhat fewer categories and simplify their presentation to respondents for easier comprehension. Similarly, as discussed in Chapter 6, the need for unimodal construction has led me to keep my scales simpler as opposed to more complex.

Simplicity has also been achieved by stripping scales of labels completely and asking people to place themselves on one to seven or other length scales. Most people can use such scales by telephone as well as in self-administered questionnaires. The disadvantage of doing so stems from removing a sense of meaning which a label gives to each point on a scale. One surveyor who frequently conducted policy surveys said he was uncomfortable going to a city council meeting to report, for example, that 30% of residents were somewhat satisfied with police protection and another 20% were completely satisfied, but did not feel at ease reporting that people were on average 3.1 on a four-point scale. The more abstract the concept and scale presented to respondents, the greater the potential for any given answer to have a less concrete meaning than when a word label is attached.

Closed-Ended Questions with Unordered Response Categories

Some questions that use unordered or nonscalar response categories, such as, “Do you rent or own your home?” are quite simple. However, choosing from among several or many categories is quite complex, as suggested by these two examples:
A. Show which of these six groups you feel should have the most and least influence in deciding whether the proposed community bypass should be built by putting a "1" in the box for most influence, "2" for second most influence, and so on until you have ranked all six choices:

- Local chamber of commerce
- State highway department
- The city council
- The mayor
- The voters
- Local businesses

B. If the highway bypass is to be built on one of these routes, which would you most prefer?

- A north route that starts west of the city at Exit 21 (Johnson Road) off Highway 30, crosses Division at North 59th Street, and reconnects to Highway 30 three miles east of the city at River Road.
- A modified north route that starts further west of the city at Exit 19, crosses Division at 70th Street, and reconnects to Highway 30 three miles east of the city at River Road.
- A south route that begins west of the city at Exit 19, crosses Division at South 24th Street, and reconnects to Highway 30 east of the city at River Road.

Responding to either of these questions requires considerable effort. The first question requires comparing six groups, then five, then four, and so forth to complete the ranking. The second question requires absorbing considerable detail, identifying differences between the choices, and then selecting the most preferred route. In both cases, providing answers requires considerably more effort to comprehend and decide how to answer than is usually the case for closed-ended questions with ordered categories. Yet these are precisely the types of question structures that can sometimes provide the most useful information to survey sponsors.

One of the challenges of creating questions of this nature is to keep the demand on respondents from getting out of hand. Each answer choice adds another concept which must be compared with other choices. Trying to provide detailed information on alternatives, as in the second example above, increases the likelihood of some small detail on a proposal being overlooked (e.g., the exact location of South 24th Street). On paper, a question that lists 15 items to be ranked from top to bottom may not look much more difficult than one which has only six items to be ranked, but the respondent demand is obviously far greater. If all 15 options need to be presented, then the question might be simplified by asking for a ranking of only the top three.
people to choose from among alternatives. In other cases, I have been given a confusing mixture of good ideas in formats that intermingled ordered and unordered categories. For example, I was once asked to help a university committee that was preparing a questionnaire to evaluate a dean's performance. All of the questions they proposed were structured as closed-ended questions with unordered response categories, similar to the one shown in Figure 2.3.

The first question sought information about both leadership and innovation skills but seemed destined to produce results that were difficult to interpret. The proposed solution was to break the question apart using both the ordered and unordered question structures, as shown in the first revision in Figure 2.3. Doing so allowed the committee to accomplish its stated objectives of finding out how faculty viewed the dean separately with regard to leadership and innovation skills, and also on which attribute she scored best.

An examination of questions posed about election candidates also illustrates how a knowledge of different kinds of question structures can help get information of greatest use to survey sponsors (Figure 2.4). When one thinks

Figure 2.3  Restructuring a question to provide interpretable answers.

An uninterpretable combination of closed-ended ordered and unordered categories:

6. Which of these five statements best describes this dean:
   - Innovative but lacking leadership qualities
   - About the same on innovation and leadership qualities
   - Stronger on leadership than innovation
   - A born leader
   - A real innovator

Revision—Ordered categories for each concept:

6. To what extent has the dean demonstrated strong leadership qualities?
   - All of the time
   - Most of the time
   - Some of the time
   - Seldom
   - Never

7. To what extent has the dean demonstrated an ability to innovate?
   - All of the time
   - Most of the time
   - Some of the time
   - Seldom
   - Never

Revision—Unordered categories that achieve head-to-head comparison of concepts:

6. Which one of the following do you feel best describes the dean?
   - A strong leader
   - A strong innovator
   - Both a strong leader and innovator
   - Neither a strong leader nor innovator

3. Who do you most like to see elected President of the United States in the next election?

Name of person I'd most like to see elected President of the United States

Closed-ended with scalar categories:

3. For each of the following candidates, please tell how qualified you feel he or she is for becoming president:

Emma Cain
   - Very well qualified
   - Fairly well qualified
   - Somewhat qualified
   - Somewhat unqualified
   - Not well qualified
   - Not at all qualified
   - And so forth for remaining candidates

Closed-ended with unordered categories:

3. If the election for President were being held today, for which one of these candidates are you most likely to vote?

   - Martha Holmes
   - David Badger
   - Harold Garwell
   - John Wilby

Partially closed-ended question:

3. If the election for President of the United States were being held today, for whom are you most likely to vote?

   - Laurel F. Logar, Democrat
   - Charles Young, Republican
   - Other (Write in name of other choice)
of election surveys, it is usually expected that head-to-head questions, as represented by the third example listed here, will be used. However, when candidates have not yet declared, and the question of much interest is who has public visibility and exists in people's minds as a potential candidate, then a completely open-ended question may be the best type to ask. If one is thinking about "dimensions" of the candidate, perhaps as a campaign manager devising beneficial advertisements, knowing how one's candidate is perceived and how that differs from perceptions of other candidates would be enormously useful. The scalar question for asking how well qualified each candidate is to be president will provide a sense of "distance" between candidates rather than a comparison of who would be best and worst. The head-to-head comparison, or closed-ended unordered structure, would be of obvious use during an election campaign, providing feedback that is typically used to design campaign strategies. The partially open-ended question is of use when write-in candidates are involved in the race. Thus, each type of question has much usefulness for determining voting preferences.

Changing question structures in this way enables us to formulate appropriate questions for achieving survey objectives. Each type of question structure performs a role that no other type can achieve as well, making an understanding of question structure a fundamental tool for drafting acceptable survey questions.

**PRINCIPLES FOR WRITING SURVEY QUESTIONS:**

**THE COMBINING OF WORDS AND STRUCTURE**

Words are the building blocks for all question structures, but deciding which words to use and in what order is far from simple. The wrong choice of words can create any number of problems, from excessive vagueness to too much precision, from being misunderstood to not being understood at all, and from being too objectionable to being uninteresting and irrelevant. No one would deny that it is important to find the right choice of words. However, making the right choice in any given situation and knowing when you have achieved it are issues on which agreement is far less likely.

Perhaps no one has summarized the dilemma of writing questions as elegantly as Stanley Payne (1951) nearly 50 years ago in *The Art of Asking Questions*. In one chapter, he presents 41 versions of a single question before finding one that he considers acceptable for a survey. Even this question is cautiously labeled "passable," and the reader is admonished that pretesting might shoot this 41st version full of holes. The concluding chapter of the book summarizes Payne's rules for wording questions, with the subtitle of the chapter, "A Concise Check List of 100 Considerations," describing the surveyor's dilemma.

The rules, admonitions, and principles for how to word questions, enumerated in various books and articles, present a mind-boggling array of generally good but often conflicting and confusing directions about how to do it.

There is no shortage of simple admonitions on what to do and what not to do. For example:

- Use simple words
- Do not be vague
- Keep it short
- Be specific
- Do not talk down to respondents
- Avoid bias
- Avoid objectionable questions
- Do not be too specific
- Avoid hypothetical questions

The problem is that these "how to do it" rules often get in one another's way. "Use simple words" is usually good advice, but frequently interferes with advice to "keep it short." It is interesting, for example, to note that Payne's 41st version of the question mentioned previously expanded from only eight words to 28 as difficult words were changed to simple ones. Using simple words also increases the risk of talking down to the respondents rather than communicating with them at the same level. The well-founded advice not to be vague often produces questions that are too specific. The advice to keep questions from being too direct and therefore objectional sometimes results in not heeding the advice to avoid hypothetical questions. In addition, although biased wording is certainly to be avoided, it is precisely such questions that may be required for building many kinds of attitude scales.

The reason that seemingly good advice, taken literally, may turn out to be bad advice is that questions are not written in the abstract. Writing questions for a particular questionnaire means constructing them for a particular population, a particular purpose, and placement next to another particular question. Words that are too difficult for use with some populations may be perfectly acceptable for others. A question that is fairly vague may satisfy the exploratory objectives of one survey, but not the analytic ones of another. A question that makes little sense by itself may be quite clear when asked after the ones that precede it in the questionnaire. A list of admonitions, no matter how well intended, cannot be considered absolute. With these cautions in mind, the remainder of this chapter is devoted to principles I have found useful in making the wording and structural changes necessary for turning an initial draft of respondent queries into acceptable survey questions.

**Principle 21: Choose simple over specialized words.**

We begin by trying to find synonyms that are likely to be understood by more people, and substituting the first word for the second.
When a word exceeds six or seven letters, chances are that a shorter and more easily understood word can be substituted, although it should not automatically be assumed that all shorter words are acceptable. For example, it would not be advisable to substitute “deter” for “discourage.”

Next, we focus on words that can be simplified only by using combinations of shorter words:

people who live here ... occupants of this household
your answers ... your responses to this questionnaire
what you do after school ... post-school extracurricular activities
job concerns ... work-related employment issues
area of the county ... subnational region

In addition, we search for specialized words or abbreviations that are commonplace for the survey sponsor, but require some translation for survey respondents. Such terms are especially prevalent in government surveys. For example:

Oil inventory questionnaire ... Form 822
Bureau of Labor Statistics ... BLS

Sometimes it is unnecessary to find substitutes for what appear to be difficult words. Virtually all occupational groups share a particular vocabulary that is not understood by outsiders. The use of such a vocabulary facilitates efficient communication, and the use of simpler words would only confuse matters. In a survey of city planners it seems quite reasonable to talk about “annexation” instead of “an addition.” Similarly, in a survey of physicians it seems reasonable to talk about “pharmaceutical companies” instead of “companies that sell medicines.” To do otherwise may even suggest a lack of knowledge and understanding of the topic of the survey.

However, the fact remains that people who write questionnaires are far more likely to overestimate than underestimate the knowledge and vocabulary of respondents. Thus, when in doubt, it is prudent to use the simpler of the available alternatives. A thesaurus is an indispensable tool for finding synonyms that might be used. In addition, there is no substitute for asking someone with less education than the survey designer to go through a questionnaire and identify words that are confusing. Pretesting with actual respondents is also very important, because it helps the surveyor identify the commonly shared vocabulary of the study population.

**Principle 2.2: Choose as few words as possible to pose the question.**

At first glance this goal may seem to contradict the frequent necessity of using several words as a substitute for a single, more complex word. That is the reason I place this goal second, rather than first. Initially, we must be sure that any words we choose are understood by virtually all respondents. Having done that, we attempt to keep questions short. Consider the following example tested for a U.S. Census questionnaire:

How many people were living or staying at this residence on Saturday, March 3rd, 2000? To make sure each person in the United States is counted only once, it is very important to:

Include everyone who lives here whether related to you or not, and anyone staying temporarily who has no permanent place to live;

But not include anyone away at college, away in the Armed Forces, in a nursing home, hospice, mental hospital, correctional facility, or other institution.

A cognitive test of this question resulted in one respondent appearing somewhat embarrassed. He then said, “I don’t have any idea how many people live in the United States.” As a result of this and other interviews, the well-intentioned second sentence that explained the importance of the inclusion/exclusion definition was removed (Dillman and Allen, 1995).

The problem with long questions stems from the fact that when people read questions they attempt to be efficient, thus giving uneven attention to each word. Important words get missed and unimportant ones sometimes receive undue emphasis, especially when the questionnaire language is a second language for some of the respondents.

Figure 2.5 shows a question in which the answer categories are given in the stem of the question and listed separately below. Such redundancy across many questions is a particularly strong indicator to respondents that it is okay to skip words, and may result in the rest of the sentence also being unevenly read. Redundancy can be eliminated easily by not including the answer choices as part of the question stem. However, the problematic version would be appropriate in an interview questionnaire when only the question stem would be read to each respondent.
Principle 2.3: Use complete sentences to ask questions.

It is tempting to meet the goal of minimizing words by using incomplete sentences for surveys. It is true that few people will misunderstand, “Your name,” or even “Age.” However, the series of questions in Figure 2.6 once caused many respondents to provide erroneous answers to the second and third questions. Nearly 20% of the respondents listed the number of years they had lived in the city or town and the county. In addition, several other respondents listed “U.S.” for country, a word which is only one letter different from country. Writing each question as a complete sentence would have helped solve both problems. In addition, county is changed to Idaho County in order to minimize the possibility of listing the United States as the respondent's county of residence.

Government-sponsored business surveys are some of the most frequent violators of this principle, a topic discussed in more detail in Chapter 10.

Principle 2.4: Avoid vague quantifiers when more precise estimates can be obtained.

The ease of using the same categories for many different kinds of questions has led to a tendency to use vague quantifiers when more precise answers can easily be obtained. Consider, for example, the question about attendance at religious services in Figure 2.7. Although it is likely that an answer to this question can easily be provided, enormous variation may exist in what respondents mean by their answers. In some religions, “regularly” implies once a week, whereas in others it may imply several times a month.

Also, the standard for regular attendance within religions may vary among respondents. For example, among adherents of a Christian religion, one person may think of regularly as implying at least once or twice a week, whereas another member of the same religion may think that annual attendance at holiday services, such as Christmas and Easter, implies regular attendance. Changing answer categories to numerical amounts, as shown in the revised form, eliminates the possibility of widely varied interpretations for regular or occasional attendance.
Problem:

26. How often did you attend religious services during the past year?
   □ Never
   □ Rarely
   □ Occasionally
   □ Regularly

A revision:

26. How often did you attend religious services during the past year?
   □ Not at all
   □ A few times
   □ About once a month
   □ Two to three times a month
   □ About once a week
   □ More than once a week

PRINCIPLE 2.5: AVOID SPECIFICITY THAT EXCEEDS THE RESPONDENT'S POTENTIAL FOR HAVING AN ACCURATE, READY-MADE ANSWER.

Nearly as troublesome as too much vagueness is the request for too much specificity. Figure 2.8 asks respondents to list the number of books they have read for leisure during the past year. Most people will have a reasonable idea of how many books they have read, but cannot list a precise number for such a long period of time. As a result, some people may simply make a reasonable guess. It is also likely that the open-ended question will elicit a high-item nonresponse because the respondent is not able to offer a precise answer. One solution to this problem is to provide answer categories, as shown in the first revision.

However, because people sometimes respond to the position of the categories as well as the category labels, great care should be taken to determine whether categories are needed and which ones are appropriate. As discussed at the beginning of this chapter, widely varied estimates of the number of hours students reported studying were obtained by varying the size of categories. Thus, people who think they read more than most people might pick the top categories regardless of the labels. Selecting categories on the basis of a pretest and other known characteristics of a population may help the surveyor make the middle categories correspond to the average numbers of books read by the survey population. This example illustrates one of the significant problems associated with simply taking question and answer choices from a survey of one population and using them for another.

Problem:

12. About how many books have you read for leisure during the past year?

□ None
□ 1-2
□ 3-5
□ 6-10
□ 11 or more

A revision:

12. About how many books have you read for leisure during the past year?

□ Less than 10
□ 11–25
□ 26–50
□ 51–75
□ 76 or more

PRINCIPLE 2.6: USE EQUAL NUMBERS OF POSITIVE AND NEGATIVE CATEGORIES FOR SCALAR QUESTIONS.

The fact that respondents draw information from the number of categories as well as from labels means that the midpoint for number of categories can easily be interpreted as the neutral point. Figure 2.9 shows an example drawn from a customer satisfaction survey in which three positive (satisfied) categories were used but only one negative (dissatisfied) category was used. Thus, the visual midpoint on the scale became "somewhat satisfied."

When a recommendation was made to change this format, it was met with objections from the survey sponsor. She pointed out that far more people who responded were satisfied than dissatisfied. Further, she felt that the most important goal of the survey was to obtain significant gradations of satisfaction, a situation similar to that which prompted development of the delighted—terrible scale discussed earlier in this chapter. The solution in this case was simply to add additional categories in a second revision, which then
questions.

Problem:

25. How satisfied were you with the service you received when you bought your air conditioner?

☐ Completely satisfied
☐ Mostly satisfied
☐ Somewhat satisfied
☐ Neither satisfied nor dissatisfied
☐ Somewhat dissatisfied
☐ Dissatisfied

A revision:

25. How satisfied were you with the service you received when you bought your air conditioner?

☐ Completely satisfied
☐ Somewhat satisfied
☐ Neither satisfied nor dissatisfied
☐ Somewhat dissatisfied
☐ Completely dissatisfied

Another revision to maintain gradations of satisfaction:

25. How satisfied were you with the service you received when you bought your air conditioner?

☐ Completely satisfied
☐ Mostly satisfied
☐ Somewhat satisfied
☐ Neither satisfied nor dissatisfied
☐ Somewhat dissatisfied
☐ Mostly dissatisfied
☐ Completely dissatisfied

allowed the three levels of satisfaction to be distinguished from one another while the scale remained balanced.

PRINCIPLE 2.7: Distinguish undecided from neutral by placement at the end of the scale.

Sometimes attitudinal questions are posed without giving the option of a neutral opinion or no opinion at all. In other cases, researchers prefer to allow for no opinion in order to distinguish true opinion holders from those who are being “forced” to choose on a topic to which they have given little or no thought.

If an “undecided” category is offered to respondents, it makes a great deal of difference where that category is placed. An experiment by Willits and Jianota (1996) compared the first and second alternatives shown in Figure 2.10 for presenting an undecided choice to respondents. When undecided was placed in the middle, respondents were consistently more likely (across 13 items) to
than doubled, from 5 to 13%. When the answer was placed in the last position (first revision), respondents were more likely to select one of the directional opinion categories. Thus, it appeared that when placed in this end position, "undecided" responses were being separated from neutral opinions. This revision appears to pick up some of the desirable features of both of the other scales, providing less of an invitation to avoid a directional response while still providing an opportunity for people who have no opinion to say so.

Another alternative is provided by the second revision, offered in Figure 2.10. It allows respondents to report being neutral on the issue, but also allows for having no opinion at all. The middle and final categories are carefully worded: "Neither agree nor disagree" and "No opinion" are used to make the meaning of choices as clear as possible.

**PRINCIPLE 2.8: AVOID BIAS FROM UNEQUAL COMPARISONS.**

Closed-ended questions with unordered categories may become unbalanced, albeit for somewhat different reasons than in the case of scalar questions as discussed under Principle 2.6. In these questions, the topic of the question appears in the answer choice. Consider the wording of the question in Figure 2.11, designed to find out whom respondents think is most responsible for outbreaks of violence in schools. The term "ineffective" places a value connotation on the first category that is not present in the other choices. Although it is unclear whether unbalancing questions in this way leads to more or less frequent selection of such categories (Schuman and Presser, 1981), the credibility of responses to such questions is inevitably open to question.

The difficulty of revising such questions is that true balance may be extremely difficult to achieve. The first revision uses less emotionally charged words, "The way children are raised by parents," but results in a category with many more words than the school and television choices. The last two categories could be made more specific by mentioning school discipline policies and violent television programs, but it is unclear without extensive pretesting whether that would improve or detract from the balance. The challenge of achieving balance on such closed-ended questions often leads to reducing choices to simple nouns (parents, schools, television), a solution that also increases the vagueness of the categories. One might, for example, wonder what aspect of television is being referenced: its use in schools, how much television students watch, or the content of the programming. Still another revision that might be considered is to completely restructure the question, converting to a closed-ended ordered question with a detailed concept presented in the stem of the question, as shown in the final revision offered in Figure 2.11.

**Problem:**

19. Which one of the following do you feel is most responsible for recent outbreaks of violence in America's schools?

- [ ] Irresponsible parents
- [ ] School policies
- [ ] Television programs

**A revision:**

19. Which one of the following do you feel is most responsible for recent outbreaks of violence in America's schools?

- [ ] The way children are raised by parents
- [ ] School policies
- [ ] Television programs

**Another revision (simplest form):**

19. Which one of the following do you feel is most responsible for recent outbreaks of violence in America's schools?

- [ ] Parents
- [ ] Schools
- [ ] Television

**Still another revision (retaining more complex descriptions):**

19. To what extent do you feel that the way children are raised by parents is responsible for recent outbreaks of violence in America's schools?

- [ ] Completely responsible
- [ ] Mostly responsible
- [ ] Somewhat responsible
- [ ] Not at all responsible

And so forth for the remaining concepts.

**PRINCIPLE 2.9: STATE BOTH SIDES OF ATTITUDE SCALES IN THE QUESTION STEMS.**

It is tempting to reduce the number of words in questions by mentioning only one side of an attitude scale when posing a question. For example, the question in Figure 2.12 asks the extent to which people agree with a statement, leaving out any mention of disagreement. Structured in this way the question may encourage people to think of a scale ranging from "not at all" to "strongly
Problem:

14. To what extent do you agree with this statement: "It's easier for people to find work in this community than it was about one year ago."
   □ Strongly agree
   □ Somewhat agree
   □ Somewhat disagree
   □ Strongly disagree

A revision:

14. To what extent do you agree or disagree with this statement: "It's easier for people to find work in this community than it was about one year ago."
   □ Strongly agree
   □ Somewhat agree
   □ Somewhat disagree
   □ Strongly disagree

Another revision, to avoid acquiesence:

14. Do you feel it is easier, the same, or more difficult for people to find work in this community than it was about one year ago?
   □ Easier
   □ The same
   □ More difficult

agree." Substituting "agree or disagree" conveys to the respondent that the scale has a greater range. It also lets the respondent know that disagreement is an acceptable answer. Mentioning both sides of a scale is even more important when the terminology of a scale changes from one end to another, as it would for a scale that goes from "strongly favor" to "strongly oppose."

However, one of the shortcomings of agree/disagree is that there is a cultural tendency in many societies for people to agree rather than disagree, or to acquiesce in their responses. Consequently, the second revision, which maintains the use of closed-ended ordered categories but changes to an "easier, the same, more difficult" format, would seem to be a better solution.

Principle 2.10: Eliminate check-all-that-apply question formats to reduce primacy effects.

In an effort to reduce respondent burden, queries containing unordered response categories are sometimes structured as "check-all-that-apply" items,
Our main concern with mutual exclusivity is when it is used for response categories of survey questions, where there is considerable likelihood that its presence will go unnoticed. Figure 2.14 shows a question in which respondents have been asked to choose one answer to the question of how they learned about a disaster, but choices combine sources as well as location. The revision simply breaks the question into two parts, one about source and the other about location.

Mutual exclusivity provides yet another reason for avoiding check-all-that-apply question formats. If people can check more than one category in the example used in Figure 2.15, it might be reasoned that mutual exclusivity is not a problem. However, Israel and Taylor (1990) have reported experimental results that provide further evidence that check-all-that-apply questions create difficulties for surveys. The check-all-that-apply question about forages fed to cattle in winter months (Figure 2.15) was asked in a survey of Florida beef producers. The answer categories for this question were presented in different orders with quite different results.

Switching the "native range" to fourth on the list decreased the proportion

**Figure 2.14** Develop response categories which are mutually exclusive.

**Problem:**

7. From which one of these sources did you first learn about the tornado in Derby?
   - Radio
   - Television
   - Someone at work
   - While at home
   - While traveling to work

**Revision:**

7. From which one of these sources did you first hear about the tornado in Derby?
   - Radio
   - Television
   - Another person

8. Where were you when you first heard about it?
   - At work
   - At home
   - Traveling to work
   - Somewhere else
Figure 2.15 Results from check-all-that-apply question that stems from the lack of mutual exclusivity (Israel and Taylor, 1990).

Question: Which of the following forages are used during the winter months to feed your cattle? (Check all that apply.)

<table>
<thead>
<tr>
<th>First Order</th>
<th>Category</th>
<th>Percent Selected</th>
<th>Second Order</th>
<th>Percent Selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Native range</td>
<td>70%</td>
<td>(4)</td>
<td>30%</td>
</tr>
<tr>
<td>2</td>
<td>Deferred grazing (save pasture for fall and winter)</td>
<td>37%</td>
<td>(2)</td>
<td>48%</td>
</tr>
<tr>
<td>3</td>
<td>Hay</td>
<td>84%</td>
<td>(5)</td>
<td>79%</td>
</tr>
<tr>
<td>4</td>
<td>Silage</td>
<td>1%</td>
<td>(1)</td>
<td>2%</td>
</tr>
<tr>
<td>5</td>
<td>Winter pasture</td>
<td>29%</td>
<td>(3)</td>
<td>36%</td>
</tr>
<tr>
<td>6</td>
<td>Other</td>
<td>14%</td>
<td>(6)</td>
<td>15%</td>
</tr>
</tbody>
</table>

deciding what answer to choose. Although it is sometimes suggested that an advantage of self-administered questionnaires is that respondents can take their time in completing them and therefore give more thoughtful answers, we know of no evidence to support that as being the typical response behavior (Dillman and Tarnai, 1991).

In addition, Jove and Mingay (1989) have reported for interview surveys that recall accuracy on visits to health providers can be improved from 41% to 53% by asking about details of the visit. For example, people can be asked who the health care provider was and how they got to the office. Respondents who are asked to recall such visits over a specified period of time, such as the last six months, can also be asked to reconstruct a calendar of important events in their life (e.g., vacations, weddings, etc.) over this period of time before asking about visits to their doctor.

Figure 2.15 shows a quick recall format and cognitively-designed alternative for reporting seat belt use in a survey conducted shortly after a law was passed in Washington state requiring seat belt use, when it was reasoned that people might forget to wear seat belts when starting on a trip in their vehicle, but buckle-up along the way (Dillman and Tarnai, 1991). Use of this cognitively-designed format produced significantly different responses than did the quick recall format; moving people away from a generalized response about the proportion of time the respondent typically wore a seat belt was therefore thought to improve the accuracy of responses.

Although I believe that a cognitive recall sequence of questions has much potential for improving the accuracy of people's responses to survey questions, its use for all questions would result in increased length of questionnaires, not to mention the greater tediousness of responding. Thus, I tend to reserve its use for the most important survey questions—those for which the most accurate estimates of behavior must be made.

**Principle 2.13: Provide appropriate time references.**

Respondents are frequently asked to report whether or how often they have engaged in a particular behavior during a recent period of months or years. There are several distinct problems associated with the use of time references in surveys. First, memory tends to fade and people usually do not categorize information by precise month or year periods. Even if they try very hard to remember, estimating behaviors over a period of the last three years (as suggested by the first example in Figure 2.17) may be impossible. Some people may end up not being able to make a meaningful distinction. One solution, which may or may not be acceptable for some studies, is to shorten the time period and use a cognitive recall set of questions prior to asking for an answer.

Another type of problem is when something is too regular and mundane in life, such as studying, watching television, or eating away from home, as
Figure 2.16 Use cognitive design techniques to improve respondent recall.

Problems:

10. Next, we would like to ask about the most recent time you drove or rode anywhere in an automobile or other vehicle such as a pickup or van. During this most recent ride, would you say that your seatbelt was fastened...
   - All the time; that is, every minute the car was moving
   - Almost all the time
   - Most of the time
   - About half the time
   - Less than half the time
   - Not at all during the time the vehicle was moving

(Question repeated for second and third most recent rides)

Revision:

10. Next, we would like to ask you to please think about the last three times you drove or rode in a car or vehicle such as a pickup or van.
First, when was the last time you drove or rode in a motor vehicle?
   - Today
   - Yesterday
   - Sometime before that

11. Could you tell us generally where this most recent trip began and where it ended?

12. About how long was this trip?
   - Less than a mile
   - One to five miles
   - Longer

13. During this trip were you the...
   - Driver
   - A front seat passenger
   - A back seat passenger
   - Other

14. During this trip, would you say your seatbelt was fastened...
   - All the time; that is, every minute the car was moving
   - Almost all the time
   - Most of the time
   - About half the time
   - Less than half the time
   - Not at all during the time the vehicle was moving

(Repeat entire sequence for second and third most recent trips)

Figure 2.17 Provide appropriate time referents.

Problem:

4. How many times in the last three years have you gone to see a doctor or other health care practitioner?
   Number of visits in last three years

A revision with shorter time referent:

4. How many times in the last six months have you gone to see a doctor or other health care practitioner?
   Number of visits in last six months

Problem:

11. How many times in last six months have you eaten away from home?
   Number of times in last six months

A revision that uses an estimation strategy:

11. On average, how many times per week do you eat a meal away from home?
   Average number of times per week

Problem:

12. How many times have you played golf so far this year?
   Number of times played this year

Revision to take into account different return doses for questionnaires:

12. During the last calendar year, 1998, about how many times did you play golf?
   Number of times played in 1998

may differ greatly by day of week, so that different responses may be obtained depending upon whether one completes a questionnaire on a Monday, following a weekend, or a Thursday, following three weekdays.

The latter problem has some similarity to the final problem question presented in Figure 2.17, in which respondents are asked how many times they have played golf this year. The problem with this question in a mail survey is that responses are likely to be higher later in the year. The revision for this question asks the number of times played in the previous year. We chose a recall for the previous year rather than a general estimation strategy because
people who play golf often keep track of the number of times they have played, or play with sufficient regularity (every Thursday afternoon from June through August) that they can provide fairly accurate estimates of times played in the previous year. The phrase, "your best estimate is fine" is sometimes added in order to limit item nonresponse. Another possibility is to prelist categories, as shown earlier in Figure 2.8.

Principle 2.14: Be sure each question is technically accurate.

Few features of a questionnaire reflect as negatively on a surveyor’s credibility as a technical error. A visible or error in certain respondents or a survey sponsor. Such errors range from the misspelling of a political candidate’s name to providing an incorrect name for an organization asked about in a survey question. Such errors are very easy to make and few can claim never to have made them.

Some errors are more subtle, such as the one in Figure 2.18, where police are identified as being responsible for catching and fining traffic violators. They only do the former; the judicial system, and not the police, is responsible for determining and administering fines. Sometimes as attitude and opinion questions are drafted and redrafted, technical issues such as this one creep into surveys where they have only a minor role, but nonetheless bring into question the surveyor’s competence.

I mention this concern as a separate issue because most surveyors, at some time in their careers, design surveys on topics about which they know little.

Figure 2.18 Be sure that each question is technically accurate.

Problem:

31. Another activity of the police department is the catching and fining of traffic violators. Should this activity receive greater, about the same, or less emphasis than at present?

☐ Greater
☐ The same
☐ Less

A revision:

31. Another activity of the police department is the catching of traffic violators. Should this activity receive greater, about the same, or less emphasis than at present?

☐ Greater
☐ The same
☐ Less

Checking questionnaires for technical accuracy becomes a very important aspect of pretesting, which I discuss in the next chapter.

Principle 2.15: Choose question wording that allows essential comparisons to be made with previously collected data.

Comparison of survey results with previously collected data often represents a major survey objective. Sometimes the comparison is being done with a previous survey of the same population in order to measure change. Another likely comparison is with U.S. Census data for the same geographical area.

Figure 2.19 illustrates the dilemma one may face in deciding whether to replicate the wording of previous surveys. This question asks whether people rent or own their homes, and is much like the structure we have seen used in hundreds of surveys. The revision is far wordier, specifies type of home to some degree (house, apartment, or mobile home), and is shown here exactly as used in the 2000 U.S. Census.

This example also illustrates one of the difficulties with many questions asked in national surveys. The wording is determined through a long process involving input from many stakeholders. As a result, the wording may be somewhat more cumbersome than many surveyors feel is needed for their particular survey. Similar issues exist with respect to the racial and Spanish/Hispanic/Latino questions to be asked in the 2000 Census, which will

Figure 2.19 Choose question wording that allows essential comparisons to be made with previously collected data.

Problem:

13. Do you own or rent the home in which you live?

☐ Own
☐ Rent

A revision:

13. Is this house, apartment, or mobile home:

☐ Owned by you or someone in this household with a mortgage or loan?
☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
☐ Rented for cash rent?
☐ Occupied without payment of cash rent?
provide comparison information for much of the next decade. The race question provides 14 categories and allows for multiple selections. The ethnicity question provides four Spanish/Hispanic/Latino categories plus an "other" category. For policy surveys in which it is crucial to be able to demonstrate similarity or differences from Census data, use of the same questions may be desirable. For other surveys, a different question structure may be quite acceptable, or even more desirable.

**Principle 2.16: Avoid asking respondents to say yes in order to mean no.**

It seems obvious that questions should not include double negatives, thus requiring a respondent to say yes to mean no, as for this question, "Should the city manager not be directly responsible to the mayor?" Yet, such questions are commonly asked in surveys. One of the reasons such questions are so prevalent in surveys is because voters are often asked in elections to vote for measures where a yes vote would result in something not being done, as illustrated by the tax approval question in Figure 2.20. Surveyors are often reluctant to pose the question differently than it would be expressed on the ballot. However, because people tend to read questions quickly, it is likely that some people will miss the word "not." In addition, the mental connection of favoring a "not" is difficult for most people.

Two different solutions for this problem might be considered. The first revision simply asks whether people favor or oppose requiring 60% approval by voters in order to raise state taxes. To help clarify what favor and oppose means for purposes of the question, the answer categories specify what favor or oppose means. This wording would seem appropriate during discussion of an issue before it has reached the ballot measure stage. A second revision, indicating that a vote will be taken, specifies the measure exactly as it will appear on the ballot and asks whether respondents are for or against approval of the measure. The switch of categories from favor-oppose to for-against is also an attempt to bring the language of the question more in line with the voting situation.

**Principle 2.17: Avoid double-barreled questions.**

A somewhat similar problem occurs with questions that contain two components which require one answer, but about which a respondent may feel differently. At first glance the question in Figure 2.21 appears to be simply another case of asking respondents to say yes in order to mean no. However, this question about whether people want a swimming pool to be built with lap lanes, but without a winter enclosure, is really asking two separate questions with no opportunity to respond to each part. This question was justified by its proponents because that seemed to be the policy issue facing voters—whether to approve a larger pool with lap lanes that would be too expensive to enclose or a smaller one that could be enclosed for winter use.

The first revision is a conventional one that allows people to respond to each of the issues and is effective in eliminating the inability of people who wanted both lap lanes and a winter enclosure to express their true opinion.

The second revision seemed more preferable in this case. By writing the question with a closed-ended unordered structure, and placing more wording into the answer choices, five clear alternatives could be presented to respondents. However, if the objective of this survey were to find out whether people would vote for or against the first proposal, then the third revision offered in Figure 2.21 might be the best question structure.

**Principle 2.18: Soften the impact of potentially objectionable questions.**

One of the most worrisome aspects of conducting self-administered surveys is that people can skip questions they object to answering, or decide not to answer any more questions. The sources of objections differ, as do the solutions.
Figure 2.21 Avoid double-barreled questions.

**Problem:**

16. Should the city build a new swimming pool that includes lanes for swimming laps that is not enclosed for winter use?
   - Yes
   - No

**A revision:**

16. Should the city build a new swimming pool that includes lanes for swimming laps?
   - Yes
   - No

17. Should the city build a new swimming pool that is enclosed for winter use?
   - Yes
   - No

**Another revision:**

16. It has been proposed that the city build a new swimming pool that could include or not include lanes for swimming laps and be enclosed for winter use. Which one of the following do you most prefer?
   - I prefer that no pool be built.
   - I prefer a pool with lanes for swimming laps and winter enclosure.
   - I prefer a pool without lanes for swimming laps or winter enclosure.
   - I prefer a pool with lanes for swimming laps that is enclosed for winter use.
   - I prefer a pool without lanes for swimming laps.

**Still another revision (for specific policy situation):**

16. It has been proposed that the city build a new swimming pool that includes lanes for swimming laps, but will not be enclosed for winter use. If the election were held today, would you vote for or against this proposal?
   - For
   - Against

Perhaps the one question most likely to elicit negative reactions from respondents is the request for income. The open-ended format presented first in Figure 2.22 is the format most likely to be left unanswered. Not only does it require respondents to record something they are likely to consider no one else's business, but for some people the question is difficult to answer. That is the case for people who do not know their exact income on a yearly basis, and this format provides an easy justification for skipping. Switching from the open-ended format to broad categories, as shown in the revision in Figure 2.22, will reduce item nonresponse. For some studies these categories may be too broad, and anything less than an exact number may be considered unacceptable. The U.S. Decennial Census, for example, asks not only for exact total income, but also asks for exact amounts from eight potential sources of income. Nonetheless, the use of categories helps greatly in overcoming the difficulties of obtaining income.

The second question in Figure 2.22 presents a different problem. Here the respondent is being asked whether she has performed an illegal behavior, such as shoplifting. Sometimes questions of this nature can be softened by changing the wording, as also shown in the revision in Figure 2.22. Another
Another revision:

26. The questions which follow are being asked to help us understand things that have happened to people over the years, and how their lives have been affected. We really appreciate your help and that of the thousands of others who have been asked to complete this national survey.

Have you ever hit someone with your fist?

☐ Yes
☐ No

27. Have you ever taken anything from a store without paying for it?

☐ Yes
☐ No

Problem:

30. “Most religions are a parasite on society.” Do you agree or disagree?

☐ Agree
☐ Disagree

A revision:

30. Here are many different opinions on religions we have heard from others. Please tell us whether you agree or disagree with each of them.

“Most religions try to help members as well as nonmembers.”

☐ Agree
☐ Disagree

31. “Most religions are a parasite on society.”

☐ Agree
☐ Disagree

Possibility shown here is to embed the question in a series of items that starts with a reason for answering, then lists other, less objectionable behaviors before asking about the illegal one.

Some questions elicit objections because of the context in which they are presented. The questionnaire is in many respects a conversation between the surveyor and the respondent, and questions that seem particularly nosy, abrupt, harshly asked, or unclear may motivate people to discontinue the conversation. An example shown in the third question in Figure 2.22 is the abrupt statement, “Most religions are a parasite on society. Do you agree or disagree?” Presented in this form, the statement appears to be the opinion of the surveyor. The revision embeds the statement in a brief context and uses a buffer item, as was done for the revision of the shoplifting question. I consider it important to examine the flow of written items, even reading them out loud to contemplate whether an introductory explanation or change in context will make questions seem less objectionable to respondents.

Principle 2.19: Avoid asking respondents to make unnecessary calculations.

When respondents are asked to report percentages, they are required to make an implicit calculation. Some will take the time to come up with the specific numbers and do the appropriate mathematical calculation, but others will simply make an estimate. To improve accuracy it is usually desirable to ask for the numerical information and reduce the burden on respondents by not asking them to do the calculations. An example is the question about percentage of nights spent away from home for business reasons, as shown in Figure 2.23. Separation into questions about total number of nights away from home, followed by one about number away on business, solves the problem by leaving the calculations to the survey analyst.

An exception to this principle is when the respondent is unlikely to know the base numbers, but might be able to make a reasonable estimate directly. For example, a worker in a large corporation might be able to offer an opinion of what proportion of men usually wear neckties to work every day. Having to determine first the number of employees and then the number who...
CONCLUSION

In this chapter a formidable gauntlet of concerns has been presented that must be addressed by the would-be writer of questions for self-administered surveys. Eight criteria have been posed for assessing potential survey questions. Three fundamentally different ways of structuring questions were presented, with advice on when one or another might be most appropriate. Finally, 19 principles for reconciling question wording with the assessment criteria and structures were described and illustrated.

Blind adherence to these criteria and principles is hardly a guarantee of success in writing questions. We have noted that principles sometimes come into conflict, and different survey situations exhibit different needs. Reconciling the issues discussed in this chapter with one's survey situation has sometimes been described as the art of asking survey questions (e.g., Payne, 1951). The information discussed here provides fundamental tools for successfully accomplishing this reconciliation.

However, the outcome of the question writing process must be viewed as incomplete. The complete stimulus presented by any question in a self-administered survey also depends upon the order in which questions are seen and answered, and whether all of the words are read by each respondent. In contrast to interview surveys, there is no guarantee that respondents will read each question in the manner intended by the questionnaire writer. However, careful design and layout of questionnaires, the topic to which I turn next, can greatly increase that likelihood.

CHAPTER 3

Constructing the Questionnaire

Self-administered questionnaires can be constructed in ways that make them easy to understand and answer. However, they are sometimes designed with features that result in questions being misunderstood, items, or even pages, being skipped altogether. Often these problems have little to do with question wording, as discussed in Chapter 2. Instead, they are the result of unfortunate decisions about questionnaire format, question order, and the appearance of individual pages. Examples encountered through many years of providing consultation to study sponsors include:

- In a university faculty-sponsored survey the questions were listed in the order they were developed. Respondents were therefore required to switch back and forth to answer questions on similar topics at different places in the questionnaire.
- A several-part income question was placed on page one of a questionnaire for a study about the purchase of health care services. The sponsor considered it the most important question in the survey, and was concerned that including this question on the last page, as I recommended (to reduce perceived costs), would result in its being skipped.
- For a nationwide survey of the general public, a proposed questionnaire was printed on one sheet of paper and folded first to form six pages, and once again so that it would fit into a half-size envelope. Observation of people who tried to fill out the questionnaire revealed that some of them unknowingly skipped pages and others made refolding mistakes so that the return address on the refolded questionnaire did not show through the window of the return envelope.
- In a well-funded survey of electric utilities a variety of g graphical forms were intermingled, including font changes, shaded and unshaded backgrounds, boxes that isolated certain text, and combinations of bold,