CHAPTER 4

The Basics of Crafting Good Questions

ON THE surface, writing a question may seem simple. For example: "How many hours per day do you typically study?" Yet the decision to ask such a question immediately raises a whole host of issues. Will the question produce different answers in different modes? Should one provide answer categories or an open space where respondents can write their answer however they desire? If one chooses to provide answer categories, what should they be, how should they be ordered, and how should they be arranged on the page?

Underneath these questions are more fundamental methodological questions. From what does a respondent draw meaning when reading and interpreting a question? How can all parts of the question work together? How do the words and the visual layout influence how respondents comprehend and ultimately

respond to the question? Does it make a difference what the specific categories are? Are answers influenced by providing other categories? Do some question structures produce higher nonresponse than others? These challenges are illustrated by a recent experiment we conducted (Smyth, Dillman, & Christian, 2007a).

We asked three randomly selected subsets of a sample of students the number of hours per day they studied in three different ways:

1. A low range consisting of five categories from 0.5 hours or less to 2 to 2.5 hours plus a sixth category for those who studied more than 2.5 hours.
2. A high range consisting of an option for those who studied 2.5 hours or less plus five categories for those studying from 2.5 to more than 4.5 hours.
3. An open-ended answer box with no categories at all so that the respondent could choose what to report.

Technically, all three ways of asking the question could accommodate students studying anywhere from 0 to 24 hours, but as Figure 4.1 demonstrates, the scales emphasized different, slightly overlapping portions within that range whereas the answer box did not emphasize any of the range.

As a result of these differences in construction, the percentage of students

reporting in a web survey that they study more than 2.5 hours per day ranged from 30% when the low scale was used to 71% when the high scale was used. This is a difference of 41 percentage points! The answer box version produced a more moderate estimate: 58% of students reported studying more than 2.5 hours. When the same experiment was conducted over the telephone, the results were very similar in that students who received the high scale were much more likely to report studying more than 2.5 hours per day than were those who received the low scale.

A previous mail experiment reported in the second edition of this book found a similar result, with 23% of respondents to the low-scale version reporting that they study 2.5 hours or more per day and 69% of respondents to the high-scale version reporting studying this much (Dillman, 2000a, pp. 32–34). Clearly, the way these questions are constructed influences the ability to accurately measure the number of hours students study. But why? What is producing such differences?

Figure 4.1 Percentage of students reporting studying $2^1/_2$ hours or less and more than $2^1/_2$ hours for three different question formats.

	Low Scale	High Scale	Answer Box
Up to 2½ hours	½ hour or less From ½ to 1 hour From 1 to 1½ hours From 1½ to 2 hours From 2 to 2½ hours	2½ hours or less	
More than 2½ hours	More than 2½ hours	From 2½ to 3 hours From 3 to 3½ hours From 3½ to 4 hours From 4 to 4½ hours More than 4½ hours	

Percentage Reporting Studying Up to and More than 2½ Hours by Format			
Up to 2½ hours	70%	29%	42%
More than 2½ hours	30%	71%	58%

Source: "Context Effects in Web Surveys: New Issues and Evidence" (pp. 427–443), by J. D. Smyth, D. A. Dillman, and L. M. Christian, in *The Oxford Handbook of Internet Psychology,* A. Joinson, K. McKenna, T. Postmes, and U. Reips (Eds.), 2007a, New York: Oxford University Press.

If respondents use only the words in the question stem and the category numbers to interpret the meaning of the question, the answers across these three versions of the question should be the same. Because the answers vary widely, it is apparent that respondents are drawing extra information from the response categories and using that information to help formulate their answers. In other words, respondents are using information well beyond the numbers that are provided to define the parameters of each response option.

For many students, knowing exactly how many hours per day they study is not something they can recall in the same way that they can recall whether they live in a student dormitory or own a car. As a result, they probably have to estimate how many hours per day they study, and doing so probably requires them to average across days during the week and on weekends and probably even across times of year (i.e., a normal day vs. a day during the week of final exams). When respondents have to do this type of mental work to formulate their answers, they often look to the question and its accompanying response options for clues.

When asked how many hours per day they study, respondents might assume that the range emphasized by the scale represents how many hours most students study. As a result, someone who gets the low scale might conclude that *most* students study between 0.5 and 2.5 hours, whereas someone who gets the high scale might conclude that most students study between 2.5 and 4.5 hours. Another assumption that respondents may make is that the middle option(s)

represents the amount that the *average* student studies; thus, those who receive the low scale might assume that the average student studies between 1 and 2 hours, but those who receive the high scale would assume that the average student studies from 3 to 4 hours. Rather than actually counting the hours they study, respondents can instead decide whether they study more, the same, or less than most typical students. In this type of estimating, different assumptions made based on the scale range and midpoint are bound to influence answers.

This example illustrates the challenge of crafting good survey questions that every potential respondent will be willing to answer, will be able to respond to accurately, and will interpret in the way the surveyor intends. Stated quite simply, one must think about many things at once to write a good question, and failure to do so can have significant effects on how the question performs. Factors to consider include what type of question to write (e.g., open- vs. closed-ended, single vs. multiple answer, etc.), how to word the question stem, what response options to offer and how to word them, how to visually present the questions, what type of answer spaces to provide, and whether and where to provide additional sources of information (i.e., instructions).

In this chapter, we address how to answer these questions and others that may arise when developing survey questions. First we describe a holistic approach to crafting survey questions that considers what question format is best and how multiple aspects of the question wording and layout need to work together to reliably provide accurate data about the concept of interest. Then we discuss

general guidelines for question wording and visual presentation that apply to nearly all survey questions. In the chapter that follows, we turn to more specific guidelines for particular types of questions. However, before we start with guidelines, we discuss four issues that need to be considered for each question.

ISSUES TO CONSIDER WHEN CRAFTING SURVEY QUESTIONS

1. What Survey Mode(s) Will Be Used to Ask the Questions?

How one writes a survey question should depend strongly on how that question is going to be delivered to respondents. The key point to keep in mind here is that different survey modes rely on different communication channels. In telephone interviews, respondents give and receive information through spoken words and the hearing system, whereas on the Web and in mail questionnaires, information is transmitted through the visual system. As a result, words take on extra importance in telephone surveys, and memory becomes a significant factor to be considered. In mail and Internet surveys, visual design elements become important. In this chapter we focus mostly on writing questions for mail and Internet surveys, with only brief comments on telephone surveys. We discuss writing questions for mixed-mode surveys in Chapter 8.

2. Is This Question Being Repeated from Another Survey, and/or Will Answers Be Compared to Previously Collected Data?

The answer to this question will influence how much, if any, the question can be changed. If a particular question has been used in another survey and the main objective is to replicate the previous survey or make the new results comparable in some other way, usually no changes or only minimal changes can be made. Examples are government surveys that have asked the same question repeatedly, sometimes for decades, to produce time-series data. For self-administered surveys, this means trying to replicate not only the question wording but also the other aspects of the visual design and layout of the questions. Thus, it is important to ask whether questions will be repeated from other surveys or previous waves of data collection and, if so, whether they can be changed.

3. Will Respondents Be Willing and Motivated to Answer Accurately?

Ensuring that respondents are motivated to respond to each question is a major concern in self-administered surveys because there is no interviewer present to encourage respondents to carefully select and report complete answers. Without

proper motivation, respondents may ignore instructions, read questions carelessly, or provide incomplete answers. Worse yet, they may skip questions altogether or fail to complete and return the questionnaire.

In some instances motivational problems stem from poor question design, such as when questions are difficult to read and understand, instructions are hard to find, or the response task is too vague. In other instances, the question topic itself may be the source of motivational problems. This is often the case with questions pertaining to personal financial information. For example, people are more likely to report their income when provided with broad categories from which to choose rather than asked to provide an exact value; however, sometimes a survey, such as the U.S. Decennial Census, requires an exact number, and anything else is unacceptable.

Respondents are also often reluctant to answer questions about certain behaviors that they may find embarrassing or threatening, such as sexual or criminal activity. When asking for sensitive information about people's past or current behavior, changing the wording of the question can encourage reluctant respondents to answer. For example, instead of asking "Have you ever shoplifted anything from a store?" one might ask "Have you ever taken anything from a store without paying for it?" Another strategy is to include the question with others, such as "How often do you go shopping?" and "What types of stores do you shop at?" so that it may appear in context and seem less objectionable. Although steps can be taken to improve the design of sensitive questions, it may

still be difficult to collect accurate information from all respondents.

4. What Type of Information Is the Question Asking For?

It is easier to get accurate answers for some types of survey questions than others. For example, almost everyone knows how old they are, as people are frequently asked to give that information to others. Because people already have an answer in their head when asked about their age, assuming they are willing they can easily provide it and can do so in a number of different ways, as shown in Figure 4.2. Because the information is readily available to the respondent, most surveyors can accurately collect such data as age and other *factual or demographic information* regardless of how they ask for it; however, as we discuss later in this chapter, people can still be encouraged, through question design, to report the answer in a particular format.

Figure 4.2 What type of information is the question asking for?

A question that people can easily answer regardless of how it is asked

What year were you born?

☐ Year born

How old are you?

☐ Age

What is your date of birth?

☐ ☐ ☐
Month Day Year

A question for which people are often more likely to be influenced by context

In your opinion, how effective do you think citizens groups are in helping to solve environmental problems?

- ☐ Very effective
- ☐ Somewhat effective
- ☐ A little effective
- ☐ Not at all effective

In your opinion, how effective or ineffective do you think citizens groups are in helping to solve environmental problems? (Please mark an "X" on the line.)

Very effective ←——————→ Very ineffective

In your opinion, how effective are citizens groups in helping to solve environmental problems?

- ☐ A great deal
- ☐ A fair amount
- ☐ Not very much
- ☐ Almost none at all

Surveys also frequently ask for information that people may have thought little about and will need more time to answer, such as questions about *attitudes and opinions.* The second example in Figure 4.2 provides an example in which respondents are asked how effective they think citizens groups are in helping to solve environmental problems. Most respondents will not have a ready answer available for this type of question and may have to do considerable work to formulate one. Some may consider generally whether they think people can effect change, whereas others will think about examples of environmental citizens groups or various types of environmental problems that people have tried to help solve. For these types of questions, more so than for factual or demographic questions, respondents can be substantially influenced by the context of the question as they work their way through the question–answer process of comprehending the question, recalling relevant information, forming a judgment, and reporting their answer (Tourangeau, 1992). Different elements of the question that can influence the answering process include what type of response is being asked for (e.g., choose one category, mark an X on the scale), the wording of the question and response options (e.g., "a great deal" or "somewhat effective"), and visual layout.

Other questions that are prone to such context effects are those asking about *behaviors and events.* Surveyors often ask about many aspects of people's behavior, such as what they have done, how often (number of times or relative frequency), and when. Frequently survey designers want respondents to provide far more

detail about past behaviors than can be recalled and, as a result, they write questions respondents find difficult, if not impossible, to answer. Doing this causes respondents to draw even more on features of the questions' context rather than their real experiences in formulating their answers, as is the case in Figure 4.1. To avoid this tendency, surveyors should consider three recall problems. First, memory tends to fade over time. Second, individual episodes or occurrences of regular and mundane events are generally not precisely remembered (Rockwood, Sangster, & Dillman, 1997). And third, people usually do not categorize information by precise month or year. Given these limitations, respondents are unlikely to be able to accurately report how many days they drove more than 5 miles during the past 6 months. But they can probably very accurately report how many days they drove their car during the past week or drove more than 200 miles at a time in the past 3 months. Asking questions about behaviors that people can easily recall because they are recent or more memorable can help improve the accuracy of the information people report. In addition to choosing an appropriate reference period for the type of behavior, using definitions and examples can also help improve recall (Schaeffer & Presser, 2003). However, definitions must be easy for respondents to understand, and examples should be selected carefully so as not to influence respondents' answers in unintended ways.

A HOLISTIC APPROACH

Throughout this and the next chapter, we discuss a holistic approach to crafting effective survey questions for mail and Internet surveys. By holistic we mean that multiple aspects of the wording and design of the question need to work together to convey meaning. This approach considers what type of question structure best measures the concept of interest, how questions are composed of multiple parts that work together, and how both the words and the visual presentation of questions are important.

Choose the Appropriate Question Format

There are two broad types of question formats: open-ended and closed-ended questions. *Open-ended* question formats provide a blank space or box where respondents type or write in their response using their own words (or numbers), whereas *closed-ended* question formats or *scalar* questions provide respondents with a list of answer choices from which they must choose to answer the question.

The strength of the open-ended question format is that it allows respondents to freely answer the question as they want without limiting their response. Thus, this format is preferable when the surveyor does not want to influence respondent answers by providing a set of answer choices; when the goal is to collect rich, detailed information from respondents; and when the surveyor is questioning about topics for which little information is known ahead of time.

Additionally, an open-ended format in which respondents provide a numerical response can sometimes be easier for respondents and yield more precise information because respondents report an exact number rather than choose from categories with vague labels or ranges of values.

However, there are also several limitations to open-ended question formats. In self-administered surveys, more respondents skip open-ended question formats than closed-ended formats because the former require more work to answer. Issues of item nonresponse bias arise because some types of respondents may be more likely to skip these questions than other types. If respondents do answer the question, they may provide only a short response. In addition, responses to open-ended questions must be entered and coded before they can be analyzed; however, web surveys make this less time consuming because the responses are already in electronic form. In addition to the time to code responses to open-ended questions, there is often a lot more variation in respondents' answers, so it may be more difficult to analyze and interpret the data, and a variable may not be able to be created based on the responses. In contrast, responses to closed-ended questions can be analyzed immediately (or with minor transformations to the data), and data results can be produced quickly.

Closed-ended question formats should be used when surveyors want respondents to provide an answer after considering or evaluating a set of answer choices. Because researchers provide answer categories in closed-ended questions, the response options they choose have significant impact on how respondents

interpret the questions. Closed-ended question formats can utilize nominal or ordinal scales. In *nominal* scalar questions, respondents are asked to compare a set of categories with no natural order underlying the categories. Because the categories lack an inherent ordered relationship, the difficulty of processing nominal scales increases as the number of categories that need to be compared at one time increases. An adaptation of the nominal closed-ended format allows respondents to select multiple answers (e.g., check-all-that-apply and ranking questions). Examples of nominal variables include grocery stores one has visited, web sites one frequently visits, and brands of personal care products purchased. Surveyors may order these variables alphabetically or group them by type, but any such ordering or grouping is subjective and is usually to help make answering the question easier for respondents. One difficulty discussed below is that such ordering and grouping can sometimes have unintentional effects on answers and actually make responding more difficult.

In contrast, *ordinal* scale questions provide an ordered set of answer categories (but the intervals between categories is unknown), and respondents must decide where they fit along the continuum. Because there is an inherent order to the categories, respondents are particularly influenced by how the categories are distributed and by the overall layout of the response scale. A common ordinal scale asks about levels of satisfaction (e.g., completely satisfied, very satisfied, somewhat satisfied, not at all satisfied), where each category represents a higher degree or level of satisfaction. Someone who is "completely satisfied" is more satisfied than someone who is "very satisfied," but it is not necessarily known

how much more satisfied. Another common type of ordered scale asks about frequency of behaviors or events (e.g., all of the time, most of the time, some of the time, none of the time), where each category represents a greater or lesser frequency. One concern with ordered closed-ended questions is that researchers often use vague quantifiers to describe the answer categories. For example, some people may consider walking every day to be "all of the time," whereas other people may consider walking three times a week to be "all of the time."

One of the fundamental writing tools that exists for creating survey questions is to shift questions from one format to another. Having a working knowledge of different question formats can help surveyors craft effective survey questions because often questions that do not work in one question format can be converted to another format to more effectively measure the concept. For example, one of us was once asked to help a university committee that was preparing a questionnaire to evaluate a dean's performance. All of the questions proposed by the committee were nominal closed-ended questions with unordered response categories, similar to the question structure commonly used in student examinations. The first question, shown in Figure 4.3, asked about both leadership and innovation in a way that would have made it difficult to interpret the results. The proposed solution to the university committee was to break the question apart and to ask two ordered closed-ended questions that focused on how often or to what extent the dean had demonstrated leadership and innovation. Then ask people a nominal closed-ended question where the dean's abilities to lead and innovate are directly compared. Separating leadership and

innovation into separate questions and asking the direct comparison question allowed the committee to test its stated objectives of finding out how the faculty evaluated the dean separately on leadership and innovation and on which attribute she performed better.

Figure 4.3 Choosing the appropriate question format.

Nominal closed-ended

Which of these five statements best describes the 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

Ordinal closed-ended for each concept

To what extent has the dean demonstrated strong leadership qualities?

- ☐ All of the time
- ☐ Most of the time
- ☐ Some of the time
- ☐ None of the time

To what extent has the dean demonstrated an ability to innovate?

- ☐ All of the time
- ☐ Most of the time
- ☐ Some of the time
- ☐ None of the time

Nominal closed-ended—revised to achieve direct comparison of concepts

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

Open-ended for each concept

How would you describe the dean's leadership abilities?

How would you describe the dean's ability to innovate?

Another option for revising the question would have been to replace it with two open-ended questions that asked separately about the dean's ability to lead and innovate and perhaps still follow up with the direct comparison of leadership and innovation in the nominal closed-ended question. Which strategy should be used depends on the ultimate purpose for asking the question. The open-ended questions would have produced more descriptive data on how faculty evaluated the dean's abilities independently; however, the ordered closed-ended questions would have allowed the committee to measure the dean's abilities using a common scale so that results could be easily summarized and compared.

A partially closed question format is a hybrid of the open- and closed-ended formats that includes a set of response categories and an "other" response, thus allowing respondents who do not fit into the provided response categories to specify a different category that they do fit. This format is often used when it would be too burdensome to ask respondents about the entire set of items. For example, in the question in Figure 4.4, if a respondent's favorite sport were basketball, she would select that sport and move to the next question; however, if her favorite sport were rowing, which is not one of the options provided, she would select "other" and write or type "rowing" in the adjacent answer space. The value of this question format is that it reduces the number of items respondents have to consider at once and still collects data for the key items of interest. However, respondents are more likely to select the options provided than to write or type their own other responses. Hence, categories should be included

for all of the key items of interest, and care should be used in drawing conclusions about volunteered categories versus those that are explicitly provided (i.e., it is not accurate to draw conclusions that compare listed to unlisted options, such as "respondents were 10 times as likely to say basketball is their favorite sport than rowing").

Figure 4.4 Example of a partially closed question format.

Which of the following is your favorite college women's sport?

- ☐ Basketball
- ☐ Gymnastics
- ☐ Soccer
- ☐ Softball
- ☐ Swimming
- ☐ Tennis
- ☐ Volleyball
- ☐ Other: Please specify

Figure 4.5 Example of new web response mechanisms.

Automatic calculation tools

License type: **Physician**

Question 4 of 19

During a typical week, approximately how many hours do you spend in the following professional Physician activities? *(Do not include on-call time)*

8	Direct patient care (including patient education)
4	Administration of clinical practice
	Teaching (Physician education)
	Research
	Other professional Physician activities
12	TOTAL (add above items - this should represent your typical weekly hours of work)

Next >> << Back

Drop-down menus

Question 1 of 25

What month and year did you begin your studies at Washington State University?

–Month– –Year–

Visual analog scales

Next, we'd like to change topics and ask you a few questions about grocery stores.

For the aspects of a grocery store listed below, how important would each be in determining at which grocery store you would shop?

On the lines below, click on the position that best reflects how important you consider each to be.

Lower prices of products than competitors

Not at all important — Absolutely essential

Source: Visual analog scale image courtesy of "A Comparison of Visual Analog and Graphic Ratings Scales," by R. K. Thomas and M. P. Couper, March 2007, paper presented at the General Online Research Conference, Leipzig, Germany.

In web surveying, new response mechanisms such as automatic calculation tools, drop-down menus, and visual analog scales have been developed (see Figure 4.5). Each of these formats uses one of the question formats we discussed above, but the way that respondents report their answers differs in each case. *Automatic calculation tools* are usually composed of a series of numerical response open-ended questions for which the computer helps calculate a running total to make responding easier and reduce the number of errors made by respondents.

A *drop-down menu* provides respondents with a list of options where they select one or sometimes multiple responses, similar to a nominal closed-ended question or a multiple-answer question. The main difference is that respondents cannot view the options until they click on the menu, and then they often have to scroll to find the answer they want to select. Heerwegh and Loosveldt (2002a) and Couper, Tourangeau, Conrad, and Crawford (2004) found that response times were not significantly different between a radio button format and drop-down box. However, Couper et al. (2004) also found that respondents were more likely

to select the visible response options when provided with a drop-down box with half of the response options showing than when presented with a drop-down box where they had to scroll to see the response options. Thus, it is important to not preset drop-down menus with only some response options visible. Instead, one should label the boxes with what type of information is being requested and only make specific items visible once the box is clicked, such as in the example in Figure 4.5.

Visual analog scales usually use an ordinal closed-ended question format in which the respondent can interactively slide a marker to the position on the scale that best describes his or her answer. The parallel in a paper survey would be placing a mark such as an X on a continuum. In two recent studies of visual analog scales on the Web, Couper, Tourangeau, Conrad, and Singer (2006) found no differences in the response distributions for visual analog scales than for scales presented horizontally with radio buttons, and Thomas and Couper (2007) found similar validity and self-reported accuracy ratings between visual analog scales and scales presented as a list of options vertically with radio buttons. Both studies found that the visual analog scales took longer to complete than the scales with radio buttons.

Although these new response mechanisms provide alternative ways for respondents to report their answer or help them provide an answer, many of them are not new question formats, so we discuss them with the specific question structure to which they apply.

The Anatomy of a Survey Question

Survey questions are made up of multiple parts that must work together in concert to produce high-quality data about the topic of interest. If one part of the question fails or provides a conflicting message with another part, it can undermine the accuracy of responses. Crafting good survey questions requires understanding how each component of the question conveys meaning independently to respondents as well as how all of the parts work together to convey meaning.

The most important part of any survey question is the *question stem*, or the words that form the actual query itself. In Figure 4.6, the question stem is "How many years have you lived in Washington?" The question stem provides the most explicit and direct information about what the question is asking (e.g., how long have you lived in Washington) and how respondents should provide their answer (e.g., in years). It may also include *additional instructions*, definitions, or examples that will help respondents comprehend the meaning of the question or of key concepts. Additional instructions might include verbal instructions (e.g., "select only one" or "please round to the nearest whole year"), or they may consist of numbers, graphics, or symbols that further inform respondents how to answer (e.g., $, #, YYYY). Each question has *answer spaces or choices* that provide additional information to respondents about what responses are possible and how

to record their answer. Answer choices limit the available possibilities from which respondents can choose (e.g., less than 2 years, 2–5 years, etc.) and often provide cues about the type and number of answers to provide (e.g., size of answer space, reminder of units requested). Throughout this chapter and the next, we use the specific terms italicized above for each of the particular parts, and we use the term *question* to mean the entire anatomy of the survey question, including all of the parts.

Figure 4.6 Examples of the components of open- and closed-ended question formats.

Open-ended question

Question stem with additional verbal and numeric instructions

How many years have you lived in Washington?
Please report only whole numbers. (For example, if you have lived in Washington 20 months, please round to 2 years.)

Answer space with additional verbal and symbolic instruction

☐ # of years

Closed-ended question

Question stem with additional verbal instruction

How many years have you lived in Washington?
Please choose the category that best describes the total number of years you have lived in Washington during your lifetime.

Answer choices

- ○ Less than 2 years
- ○ 2 to 5 years
- ○ More than 5 years, but less than 10 years
- ○ 10 or more years

Words used in the question stem, additional instructions, and response options

are the primary sources of meaning that respondents draw upon when comprehending the meaning of survey questions. In addition, numbers, symbols, and graphics can also influence how respondents answer questions. Finally, the visual layout and presentation of questions can also have significant effects on how respondents interpret and answer questions in mail and web surveys. Thus, in open-ended questions, respondents draw meaning from the question stem, any instructions that accompany the stem or answer spaces, and the visual display of the answer spaces (e.g., size, location, etc.). In closed-ended questions, both the wording and the visual display of the answer choices, in addition to the question stem and any instructions provided, can influence how people respond. Therefore, crafting survey questions involves both choosing words to form the questions and deciding how to visually present the questions, including each of the component parts, to respondents.

In the remainder of this chapter, we present guidelines for crafting survey questions that apply to nearly all types of survey questions. First we discuss guidelines for choosing words and writing good survey questions. Then we discuss further how the visual layout of survey questions can influence responses to questions in mail and web surveys. Lastly, we present general guidelines for the visual presentation of survey questions that apply to nearly all types of survey questions.

GUIDELINES FOR CHOOSING WORDS AND FORMING

Guideline 4.1: Make sure the question applies to the respondent

Imagine you are responding to a survey that asks "What type of Internet connection do you have in your home?" but you do not have an Internet connection. How do you respond? What if you ate dinner at a restaurant last night but are asked "If you made dinner at home last night, about how many minutes did it take to prepare the meal?" (see Figure 4.7). Sometimes people who write questions for mail surveys try to reduce the number of questions (i.e., save space) and avoid skip instructions by asking questions such as these.

Figure 4.7 Make every question require an answer.

A question that does not require an answer from every respondent

If you made dinner at home last night, about how many minutes did it take to prepare the meal?

☐ minutes

A revision that uses a filter question

Did you make dinner at home last night?

☐ Yes
☐ No

If yes, how many minutes did it take to prepare the meal?

☐ minutes

These two questions have a common problem (i.e., they do not require answers of every respondent), but the source of the problem is slightly different

for each. The first question contains an embedded assumption of having an Internet connection that may not be true for all respondents. The second avoids making assumptions by including an "if" statement but does not apply to those who do not fit the "if" criteria. Even if a "does not apply" box were provided, the use of the word *if* implies that no response is needed from those who ate out the previous night.

This type of question can be particularly damaging in web surveys when respondents are required to enter an answer for every question before they are allowed to advance to the next question. In this situation, the respondent has to choose between two bad options: knowingly entering false information or quitting the survey altogether. However, questions that do not apply to every respondent are still problematic in questionnaires that do not require answers. Aside from potentially confusing respondents, the methodological problem with these questions is that it is impossible to distinguish between those who did not respond because they were unmotivated (i.e., nonrespondents) and more motivated people who, nonetheless, did not respond because the question did not apply to them. Moreover, in order to be able to estimate the distribution of a characteristic in the sample population, the surveyor must give respondents the opportunity to answer every question they are asked. Therefore, a good rule to apply is that in order for an inquiry to constitute a survey question, it must require an answer from each person of whom it is asked. The two questions above should only be asked of respondents who answer "yes" to filter questions such as "Do you have an Internet connection in your home?" or "Did you make

dinner at home last night?" (see Figure 4.7). Techniques to ensure that respondents follow the required skip instructions properly are discussed in Chapter 6.

Guideline 4.2: Make sure the question is technically accurate

Asking a question that is not technically accurate can confuse respondents and make answering difficult. For example, many avid horse people might be confused by this seemingly simple question: "How many feet tall is your horse?" This is because horses are often measured in a different unit—hands (one hand equals 4 inches). A more appropriate way to ask this question would be "How many hands tall is your horse?" Ensuring that questions are technically accurate becomes more challenging when one is asking questions about topics that apply to very specialized populations such as equestrians or business executives. Failing to do so, however, can compromise the quality of responses as well as diminish the perceived credibility and authenticity of the surveyor, possibly resulting in reduced motivation or even break-offs on the part of respondents.

Guideline 4.3: Ask one question at a time

On first take, the advice to ask one question at a time seems like a no-brainer, yet it is striking how often what appears to be one question actually contains

two components about which respondents may feel differently. Consider, for example, the question in Figure 4.8. This question actually contains two questions: "Do you subscribe to any periodicals, magazines, newsletters, etc.... ?" and "Do you regularly read any periodicals, magazines, newsletters, etc.... ?" As this question is written, it poses problems for respondents who subscribe to but do not regularly read these items, or for those who regularly read them but do not subscribe. Such respondents will not know for which component to provide their response. The question also poses a problem for the survey or anyone using the resulting data, as they will not know which component the respondents were referring to when they marked "yes" or "no."

One possible solution to this problem, as demonstrated in the first revision, is to untangle the original question by asking the two questions separately. If even more precision is desired, each of the two questions can be written in a forced-choice style so that respondents are asked to indicate whether they subscribe to and whether they read each type of literature, as shown in the second revision in Figure 4.8.

Guideline 4.4: Use simple and familiar words

One way to establish legitimacy and credibility with respondents is to present them with a formalized and professional questionnaire. Generally this is good practice. One way it can backfire, however, is if efforts to formalize the questionnaire lead to the use of complex words or phrases and technical

terminology that not all respondents will understand. Many complex words and phrases can be easily replaced by more generally understood terms, as seen in Figure 4.9. When drafting questions it may be advisable to consult a grammar book or writing manual, as they commonly provide more extensive lists of replacement terms for complex words and wordy phrases. A good rule of thumb is that when a word exceeds six or seven letters, a shorter and more easily understood word can probably be substituted. However, it should not automatically be assumed that all shorter words are acceptable. For example, it would not be advisable to substitute "deter" for "discourage."

Another common tendency, especially in government surveys, is to inadvertently use abbreviations or specialized phrases that are commonplace for the survey sponsor but require some translation for respondents. An example is shown in Figure 4.10. Although the survey sponsors may know what Form SS-4 is, many respondents may not. Others may not immediately know what IRS stands for, further confusing them and making it even harder to figure out what form SS-4 is. A clearer statement of the question is provided in the revision, in which the form is referred to by its full name and "IRS" is replaced with "Internal Revenue Service."

Figure 4.8 Ask one question at a time.

A double-barreled question

Do you personally subscribe to, or regularly read, any periodicals, magazines, newsletters, etc. that are specifically related to your occupation?

☐ Yes
☐ No

A revision to ask each question separately

Do you personally subscribe to any periodicals, magazines, newsletters, etc. that are specifically related to your occupation?

☐ Yes
☐ No

Do you regularly read any periodicals, magazines, newsletters, etc. that are specifically related to your occupation?

☐ Yes
☐ No

Another possible revision to collect more specific information

Please indicate whether or not you personally subscribe to each of the following sources of information specifically related to your occupation.

	I do subscribe	**I do not subscribe**
Periodicals	☐	☐
Magazines	☐	☐
Newsletters	☐	☐

Please indicate whether or not you regularly read each of the following sources of information specifically related to your occupation.

	I regularly read	**I do not regularly read**
Periodicals	☐	☐
Magazines	☐	☐
Newsletters	☐	☐

In most instances it is desirable to replace complex and specialized words, but there are instances when this is not necessary. Virtually all occupational groups share a particular vocabulary that is not understood by outsiders but that facilitates efficient communication within the group. Replacing this specialized vocabulary with simpler words would only confuse matters for these groups. In a survey of city planners, for example, 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.

Figure 4.9 Words and phrases that can be simplified.

Replacing complex with simple words

Exhausted.....Tired

Candid.....Honest

Top priority .....Most important

Leisure.....Free time

Employment.....Work

Courageous .....Brave

Rectify.....Correct

Replacing complex with simple phrases

Occupants of this household .....People who live here

Your responses.....Your answers

Post-school extracurricular activities What you do after school

Subnational region .....Area of the country

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. Ultimately, though, the best way to determine if the vocabulary is appropriate is to pretest questions with members of the population of interest to identify potential difficulties, a topic we return to in Chapter 6.

Figure 4.10 Use of simple and familiar versus complex or specialized words.

Unnecessary use of specialized words and abbreviations

Have you filed Form SS-4 with the IRS?

☐ Yes
☐ No

A simplified revision

Have you filed an application for an employer identification number (Form SS-4) with the Internal Revenue Service?

☐ Yes
☐ No

Guideline 4.5: Use specific and concrete words to specify the concepts clearly

Suppose you are interested in different elements of family cohesiveness and you pose the question in Figure 4.11 to a sample of mothers with school-age children. One strength of this question is that it clearly specifies a reasonable time referent and the units in which one should report an answer. The problem, however, is that it contains several vague concepts. What does "eat" mean? Should we count snacks? What if we all gathered for a smoothie a couple hours before dinner; does that count, even though it is a thick drink? And what does "together as a family" mean? Does the lunch we got from the drive-thru and ate in the car on the way to grandma's house count? What about the pizza we ate in the living room while watching a movie on Saturday night? Do we count the end-of-season potluck for the baseball team at which we were all present but interspersed with other families?

One respondent might take a very liberal interpretation of this question and include all gatherings where her whole family was present and food or drinks were consumed, whereas another may take a very conservative view and count only full meals for which the family was gathered at home around the kitchen table. The problem is that the two interpretations would most likely result in quite divergent answers that are due to different interpretations of the question, not differences in family cohesiveness. This question could be improved by specifying that you are interested in meals consumed at home, as in the first revision. It could be made even more specific by specifying that you are interested in sit-down meals shared as a family, as in the second revision.

Figure 4.11 Use specific and concrete words to specify the concepts clearly.

Question with vague concepts

How many times did you eat together as a family last week?

☐ # of times

A revised question with more specific and concrete concepts

How many meals did you eat together as a family at home last week?

☐ # of meals

A more specific revision

How many meals did you sit down to eat at home as a family last week?

☐ # of meals

This example illustrates a common problem for many writers of survey questions. Once the units are specified, many concepts such as age, height, and

weight are very straightforward. Others, however, are not as straightforward as they seem. It seems like everyone would know what it means to eat together as a family, but as the example shows, once one begins to factor in the complexity of family life in this day and age, the concept of eating together as a family is opened up to much interpretation. Thus, it is important to make sure the concepts in survey questions are clearly defined and communicated in order to minimize the amount of interpreting and defining that respondents need to do.

Guideline 4.6: Use as few words as possible to pose the question

Part of keeping questions simple is keeping them short and to the point. The longer the question, the more information the respondent has to take in and process, and the higher the likelihood for misunderstanding or misreading. When presented with the question in Figure 4.12, for example, one respondent in a cognitive interview answered, "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 reason for the directions about who to include and who to exclude was removed (Dillman & Allen, 1995).

The goal of keeping questions short sometimes contradicts the previously stated goals of using familiar and simple words and using specific and concrete words to specify concepts clearly. Substituting several simpler words for a more complex word or carefully specifying concepts can lengthen questions. In these instances, we recommend subordinating the goal of keeping the question short to

the goals of using simple and familiar words and using specific and concrete words. Once one is sure that any words chosen are understood by virtually all respondents and the concepts are clearly specified, one can attempt to keep the question short.

Figure 4.12 Use as few words as possible to pose the question.

Long question with potentially confusing information

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 shorter revision with potentially confusing information removed

How many people were livingor staying at this residence on Saturday, March 3rd, 2000? Please:

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.

Figure 4.13 Eliminating wordy and redundant expressions.

Due to the fact that	Because
At this point in time	Now
A small number of	A few
A considerable number of	Many
Small in size	Small
Has the ability	Can
Ascertain the location of	Locate
Concerning the matter of	About
If conditions are such that	If
In the majority of instances	Usually
Make a decision	Decide
Take into consideration	Consider

There are several ways to do this. One way is to replace wordy and redundant expressions such as those shown in Figure 4.13 with simpler wording. More comprehensive lists of commonly used wordy expressions and their replacements can be found in most grammar and writing manuals.

Another technique is to avoid including answer categories in the question stem, as in the following example: "Are you very likely, somewhat likely, somewhat unlikely, or very unlikely to visit Glacier National Park again?" If the respondent will be able to see the response options after reading the question stem, it is unnecessarily redundant to include these options in the stem as well. Such redundancy across many questions is a particularly strong indicator to respondents that it is okay to skip words, and it may result in the rest of the sentence also being unevenly read. Therefore, this question should be shortened to "How likely or unlikely are you to visit Glacier National Park again?" However, it is important that this technique only be used in single-mode self-administered survey designs in which respondents will be reading the question for themselves. In surveys where an interviewer administers the survey or when self-administered surveys are combined with interviews in mixed-mode studies, the answer categories may need to be included with the question stem for better respondent comprehension and consistency across modes (see also Chapter 8).

Guideline 4.7: Use complete sentences with simple sentence structures

It is tempting to save space by using incomplete sentences for paper surveys. It is true that few people will misunderstand "Your name" or even "Age." However, the series of questions in Figure 4.14 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 "United States" for *county,* a word that is only one letter different from *country.* Writing each question as a complete sentence would have helped solve both problems. In addition, in the revision "county" is changed to "Idaho county" in order to minimize the possibility of listing the United States as the respondent's county of residence.

Figure 4.14 Use complete sentences.

Common use of incomplete sentences

Number of years lived in Idaho

☐ Years

Your city or town

☐ City or Town

Your County

☐ County

A revision using complete sentences

How many years have you lived in Idaho?

☐ Years

In what city or town do you live?

☐ Name of City or Town

In what Idaho county do you live?

☐ Name of Idaho County

Using complete sentences is even more important in web surveys with page-by-page construction. When there is only one question per screen, respondents can easily lose track of the context in which an inquiry is made. Here, incomplete sentences become isolated in ways that can make their meaning even less clear.

When asking for very specific information it is tempting to add extra clauses onto the sentence to help specify the focus of the question; however, these may confuse respondents or result in their misunderstanding the question. The problem with doing this is that reading sentences with multiple clauses or a complex sentence structure requires more skill than reading sentences with simple structures. Just as with wording, it is advisable to avoid complex sentence structures and replace them instead with simple structures.

Guideline 4.8: Make sure "yes" means yes and "no" means no

It seems obvious that questions should not include double negatives, or, in other words, require a respondent to say "yes" to mean "no" as in the following question: "Should the city manager not be directly responsible to the mayor?" Yet such questions are commonly asked in surveys. One of the reasons they are so prevalent 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 4.15. Surveyors are often reluctant to pose the

question differently than it will 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, the answer categories specify what favor and oppose mean for the purposes of this question. 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 that of the voting situation.

Figure 4.15 Make sure "yes" means yes and "no" means no.

A question containing a double negative

Do you favor or oppose not allowing the state to raise taxes without approval of 60% of the voters?

- ☐ Favor
- ☐ Oppose

A revision with no double negative

Do you favor or oppose requiring 60% approval by voters in order to raise state taxes?

- ☐ Favor
- ☐ Oppose

A revision that preserves important wording

In the September election, you will be asked to vote on this referendum: "No state tax can be raised without approval by 60% of those voting in a statewide election." If the election were held today, would you vote for or against approval?

- ☐ For
- ☐ Against

Guideline 4.9: Be sure the question specifies the response task

Another way of stating this guideline is that the response task should be clear to the respondent after having only read the question stem. There are two pieces to this guideline. First, *the question stem has to clearly state the response task.* This means that the question stem needs to ask for exactly the kind of information and level of detail (e.g., units) the surveyor wishes to collect. Second, *the response format and/or options provided must match the task as it is stated in the question stem.* In other words, do not change the rules in the middle of the game. A yes/no question should only have "yes" and "no" as substantive response options (nonsubstantive options such as "don't know" or "not applicable" are still appropriate). Similarly, if the question asks how many days something occurred, the response options should be numbers appropriate to the time referent (e.g., 0–7 if the referent is a week) or a number box labeled "number of days." Asking for the number of days in the question stem and then providing options such as "Always," "Most of the time," "Sometimes," "Rarely," and "Never" represents a mismatch between the question stem and the response options that forces the respondent to undertake the extra, and sometimes difficult, step of determining which category fits best (i.e., is 5 days "most of the time" or "sometimes"?).

HOW DOES THE VISUAL PRESENTATION OF SURVEY QUESTIONS MAKE A DIFFERENCE?

Prior to the 1990s, most of the survey methodology research focused on how the wording of questions influences how respondents answer them; the visual design and layout of questions was mostly viewed as an art form. Since then, researchers have drawn on research from the science of visual perception (Hoffman, 2004; Palmer, 1999; Ware, 2004) to formulate theoretically based rationales for how the design of questions and questionnaires can be improved to help respondents process survey questions and navigate through self-administered questionnaires (Christian, Dillman, & Smyth, 2007a; Jenkins & Dillman, 1997; Redline & Dillman, 2002; Redline, Dillman, Dajani, & Scaggs, 2003). In addition, a growing body of experimental research has demonstrated how the independent and combined effects of manipulating various aspects of the visual design and layout of survey questions influence how people respond to paper and web surveys (Christian & Dillman, 2004; Dillman & Christian, 2005; Smyth, Dillman, & Christian, 2007b; Smyth, Dillman, Christian, & McBride, in press; Smyth, Dillman, Christian, & Stern, 2006b; Tourangeau, Couper, & Conrad, 2004, 2007).

Visual design and layout influences respondents as they organize information presented in the survey questionnaire and as they focus their attention on responding to individual survey questions. Visual design features can help guide

respondents to self-administered surveys, much like an interviewer would do in a face-to-face or telephone survey, to ensure that each respondent receives the same question stimulus delivered in the same way. Thus, this research is requiring survey methodologists to apply new visual design concepts to the practice of crafting survey questions and constructing questionnaires.

In this chapter, we briefly introduce the visual design concepts, summarized in Figure 4.16, necessary for understanding how survey respondents process and respond to survey questions. We follow this discussion with guidelines that can be applied to the visual presentation of all types of survey questions to help ensure that respondents attend to each component and process the parts of the question in the intended order. In Chapter 5, we present guidelines for the wording and visual presentation of specific types of open- and closed-ended questions. We continue our focus on visual design in Chapter 6, where we describe how respondents visually process and navigate entire questionnaire pages. There we present guidelines for questionnaire construction to help respondents organize the information presented on each questionnaire page and to encourage them to process questions in the intended order.

Survey questionnaires include four types of visual design *elements* that communicate meaning to respondents: words, numbers, symbols, and graphics. *Words* are the most powerful source of meaning that respondents draw upon when answering survey questions; however, *numbers* used in the question stem, in instructions, and in or as labels for answer choices also communicate additional

meaning to respondents. *Symbols* can also be used to add special meaning, often without occupying very much physical space. For example, an arrow may communicate to respondents where to focus their attention next. Finally, *graphics* (e.g., text boxes, squares, html boxes, circles, and radio buttons) are another type of visual design element that can be used in designing survey questions. In addition, graphics can include more complex images and logos that layer various elements. The use of these types of graphics has increased, particularly in web surveys, because of the ease and affordability of including them, but also in mail surveys because the cost of printing them has decreased.

Figure 4.16 Visual design concepts that guide question design.

Visual design elements that communicate information to respondents

Words are the fundamental source of meaning that help respondents understand what is being asked of them.

Numbers are used to convey meaning and sequence or order to respondents.

Symbols are figures that add special meaning based on what they represent to respondents.

Graphics are shapes and visual images that can be simple or complex and convey meaning to respondents.

Visual design properties that modify how elements are presented visually and the meaning respondents assign to them

Size: Changes in the size of elements influence how elements are perceived and whether they stand out visually.

Font: Changes in the shape and form of elements influence the legibility of words and how elements are perceived.

Brightness/Contrast/Color: Changes in shading and color influence how elements are perceived and whether they stand out visually from the background.

Location: How near or far elements are from one another (the spacing and alignment) influences whether they are perceived as related or unrelated.

Gestalt grouping principles that guide how respondents perceive relationships among information

Pragnanz: Elements that are organized into the simplest, most regular, symmetrical objects will be easier to perceive and remember.

Proximity: Placing visual elements closely together will cause them to be perceived as a group.

Similarity: Elements sharing the same visual properties (color, shape, size, orientation, etc.) will be grouped together.

Elemental connectedness: Elements connected by other elements will be grouped together.

Common region: Elements within a single closed region will be grouped together.

Continuity: Visual elements that can be seen as continuing smoothly will be perceived that way.

Closure: Elements that together create a "closed" figure will be perceived that way.

Common fate: Elements that move or imply movement in the same direction will be grouped together.

Each of the four visual design elements (words, numbers, symbols, and graphics) can be presented in different ways: They can be large or small, light or **dark**, close to one another or far apart, static or in motion, gray scale or in color. Thus, in addition to using the four types of visual design elements, surveyors can also manipulate the *properties* of each of these types of elements to increase or decrease the attention and change the meaning respondents assign to them. Visual design properties include size, font, brightness or contrast, color, location or proximity, shape, orientation, and motion. For example, attention can be drawn to particular words, numbers, or symbols by changing their size, contrast, or color in relation to the surrounding text (e.g., **bolding** or *italicizing* an important word or phrase in the question stem). A dollar symbol ($) could flicker (motion) to remind respondents to report their income in dollars. Symbols can also be located in proximity to the answer spaces where respondents will need to use them at the time of response. In addition, differently shaped graphics for response scales, for example a ladder versus a pyramid shape, can influence how respondents interpret and respond to the scales (Schwarz, Grayson, & Knäuper, 1998).

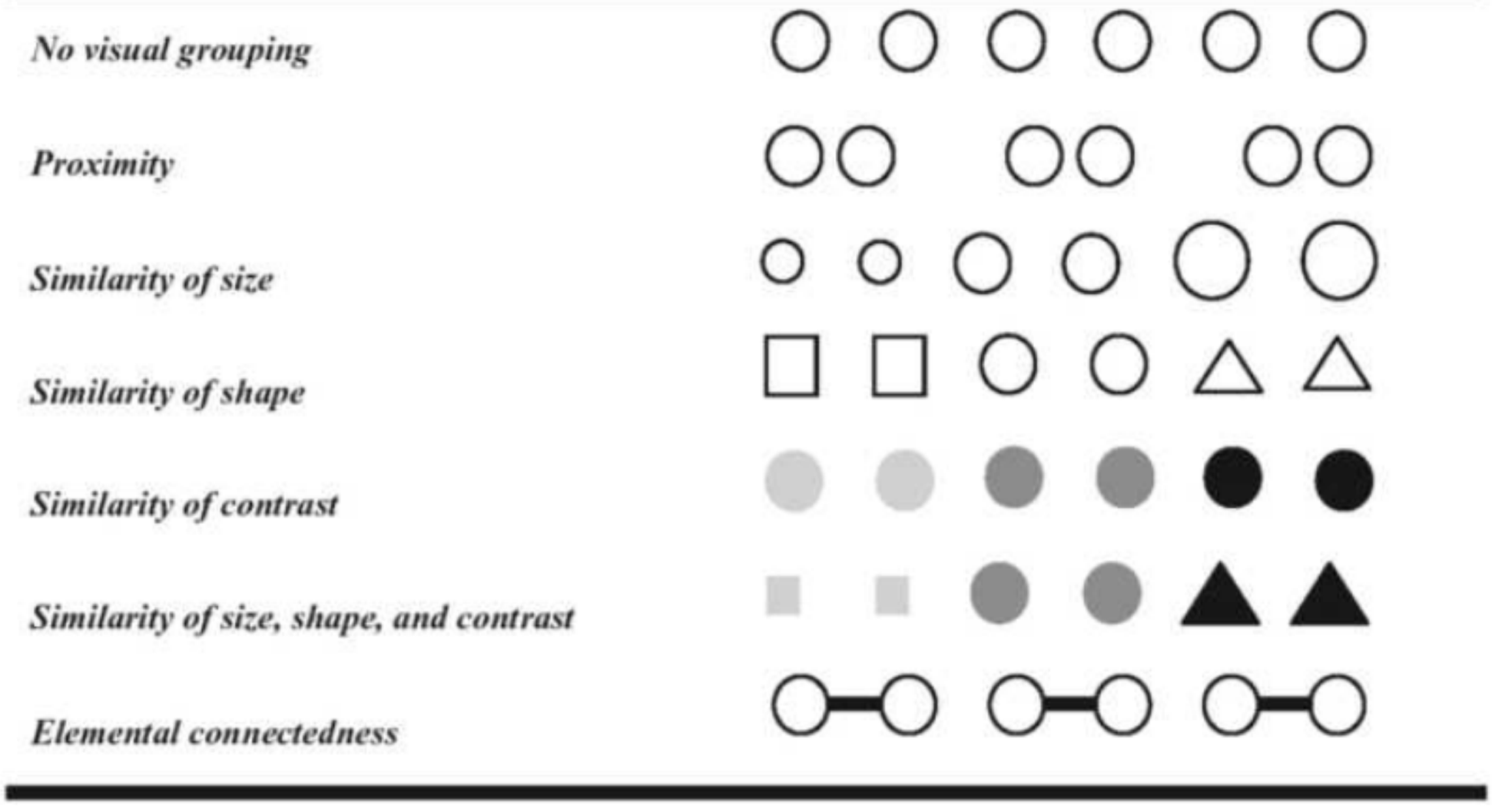
When respondents are presented with a questionnaire, they first take in the entire scene and organize the information presented to them. During this process, respondents begin to distinguish the various visual elements on the page, and the properties of these elements influence whether they are noticed (Ware, 2004). After respondents organize the information on the page, they focus on answering

the questions. Here, the visual processing is attentive and conscious, the visual field narrows to about 2 degrees or 8 to 10 characters in width (the *foveal* view), and attention is focused on only a few elements (for definitions of *attentive processing* and the *foveal view,* see Figure 6.4 in Chapter 6). During focused attention, the properties of the visual design elements or how they are displayed can strongly influence the meaning respondents assign to them.

The Gestalt psychology principles of pattern perception can help surveyors understand how respondents perceive groups among visual elements with shared properties and then assign meaning by viewing the grouped elements as conceptually related (see Figure 4.17). According to the principle of *proximity,* locating elements that should be grouped closer to each other than to other elements encourages respondents to perceive them as related. Surveyors can also use *similarity* of contrast, color, size, or shape to encourage respondents to perceive elements as a group. Respondents also perceive elements that are enclosed in a *common region,* such as a box or an area with a common background color, as a group. Likewise, connecting visual elements by using another element, such as a line, encourages respondents to perceive the *connected* elements as a group. Finally, elements that continue smoothly will be perceived as a group; *continuity* is probably used in surveys mostly in presenting complex graphics where multiple elements are layered in a continuous manner so that they are perceived as a group. These Gestalt principles can help surveyors in deciding how to present visual design elements and which properties to apply to them. In addition, using two or more of these principles to layer properties on

the same element(s) can send a stronger stimulus to make them appear as a group than varying only one property of the elements.

Figure 4.17 Examples of Gestalt grouping principles.



The following example illustrates how visual design elements and properties

used in accordance with the Gestalt grouping principles can improve responses to a simple survey question. In a recently published article based on a series of experiments, we found that visual design changes to the question "What month and year did you begin your studies at WSU?" increased the percentage of respondents reporting their answer in the desired format (i.e., two digits for the month and four digits for the year) from 55% to 96%, an increase of 41 percentage points (Christian et al., 2007b). In a follow-up study, Christian (2007) reported the sequential impact of a series of visual and verbal manipulations to the same question. The results of her study can be seen in Figure 4.18. In the initial version at the top of Figure 4.18, respondents were provided with two equal size boxes for the month and year, and only 44% provided their answer using the desired format. However, when the size of the month box was reduced by half (consistent with the expectation that the month be reported in half the number of digits as the year), more respondents reported the year using four digits, raising the percentage using the desired format to 57%. Thus, the size of the answer boxes communicated additional meaning to respondents, beyond the graphics alone, about how many digits they should use in providing their answer.

In the next manipulation, adding the verbal instruction to "Please provide your answer using two digits for the month and four digits for the year," resulted in a 21 percentage point increase, bringing the percentage using the desired format to 78% and demonstrating that respondents were processing the words in the instruction and applying them when providing their response.

Providing the verbal instruction directly after the question stem helped to ensure that respondents would see and process it just prior to providing their answer. Finally, replacing the word labels "Month" and "Year" with a symbolic instruction MM YYYY beneath their respective boxes increased the percentage using the desired format to 94%. The symbolic instruction was designed so that the letter *M* was used to represent month and *Y* to represent year, with the number of letters indicating the number of digits to use when reporting the month and year. In this version, respondents gained additional meaning from the symbolic instruction about the number of digits to use in their response. Locating the instruction near the answer spaces and within their foveal view also helped to ensure that respondents would notice and apply the instruction when reporting a response.

Figure 4.18 Example of the influence of visual design on how respondents report date information.

Experimental Treatment	Used desired format (2-digit month, 4-digit year)
Question 1 of 30 **What month and year did you begin your studies at WSU?** Month Year	44%
Question 1 of 30 **What month and year did you begin your studies at WSU?** Month Year	57%
Question 1 of 30 **What month and year did you begin your studies at WSU? Please provide your answer using two digits for the month and four digits for the year.** Month Year	78%
Question 1 of 30 **What month and year did you begin your studies at WSU? Please provide your answer using two digits for the month and four digits for the year.** MM YYYY	94%

Source: *How Mixed-Mode Surveys Are Transforming Social Research: The Influence of Survey Mode on Measurement in Web and Telephone Surveys*, by L. M. Christian, 2007, Pullman, WA: Washington State University. Unpublished doctoral dissertation.

In the previous experiment reported by Christian et al. (2007b), providing the symbolic instruction with the answer spaces had the highest impact on use of the desired format (increased use by 35 to 42 percentage points in the two experiments), but there was no additional instruction presented with the question stem. That the instruction was so important in Christian's (2007) experiment (see Figure 4.18) without the MM YYYY symbol to indicate the desired number of digits suggests that verbal instructions are particularly effective in the absence of adequate visual information. Results of telephone experiments confirm this conclusion: The instruction to use the desired format on the telephone, where there is no visual information available, raised the percentage reporting in the desired format from less than 1% to 59% (Christian, 2007). Nevertheless, in self-administered surveys, it is now clear that the use of visual information is key to obtaining desired responses and can contribute to this effort above and beyond question and instruction wording.

GUIDELINES FOR THE VISUAL PRESENTATION OF SURVEY QUESTIONS

Because the visual presentation of survey questions influences how people answer them, choosing words and forming clear questions is not enough; surveyors also need to think about how to put all of the components of the question together.

We now turn to describing general guidelines for the visual presentation of survey questions that apply to designing nearly all types of questions. We continue to use the visual design concepts we just discussed in presenting these guidelines.

The poorly designed question in the top panel of Figure 4.19 would undoubtedly turn many respondents off. Some might not even be able to understand that this is a survey question or what it is asking, making it difficult for most respondents to provide an answer. For the next five guidelines, we use this question as an example. The overall problem with this question is that it is unorganized and cluttered, so there is no clear message sent by the visual design. But such a broad observation does not necessarily give us enough information to start revising the question. When we look closer, however, we can identify a number of more specific problems that we can begin to address:

- It is difficult to tell where the question stem ends and the response options begin.
- The response options run together.
- It is not immediately clear how one should mark an answer.
- Certain options stand out more than others, making them more likely to be selected.

Figure 4.19 Implementing general visual design principles to construct individual questions.

Poor design

Which one of the following best **describes** the reason for your most *recent visit* to the Southgate Mall? ♡ **Shopping** for fun/entertainment ▷ Shopping for a needed item [ ]MALL WALKING/Exercise ☐ Other _____ ○Dining at the mall ☆Hanging out with friends Meeting new people ♧ Conducting Business

Revision with improved design

Which one of the following best describes the reason for your most recent visit to the Southgate Mall?

- ☐ Shopping for fun/entertainment
- ☐ Shopping for a needed item
- ☐ Mall walking/exercise
- ☐ Dining at the mall
- ☐ Hanging out with friends
- ☐ Meeting new people
- ☐ Conducting business
- ☐ Other _____

Poor design

How much do you favor or oppose implementing a merit-based pay system for elementary school teachers? [1] Very much in favor [2] Somewhat in favor [3] Neutral [4] Somewhat oppose [5] Very much oppose

Revision with improved design

How much do you favor or oppose implementing a merit-based pay system for elementary school teachers?

- ☐ 1 Very much in favor
- ☐ 2 Somewhat in favor
- ☐ 3 Neutral
- ☐ 4 Somewhat oppose
- ☐ 5 Very much oppose

- The purpose of the bolding, underlining, and reverse print are unclear.
- One would have to process the "other" option before processing all of the options provided by the surveyor.

The revision is clearly an improvement on the original design. To construct it we used the following five guidelines.

Guideline 4.10: Use Darker and/or Larger Print for the Question and Lighter and/or Smaller Print for Answer Choices and Answer Spaces

The first thing we needed to do was create subgrouping within the question. Good subgrouping helps the respondent quickly recognize and process the parts of the question. We used the design property of *contrast* to create separation between the question stem and the response options. To create differences in contrast, we bolded the question stem but not the response options. This is the standard use of bolding that we have adopted for most of the examples throughout this book. If we wanted to reserve bolding for another purpose in our questionnaire, however, another property we could manipulate is text size (see the bottom example of Figure 4.19). Increasing the size of the text in the question stem but not the response options helps differentiate these two parts of the question.

Guideline 4.11: Use Spacing to Help Create Subgrouping within a Question

The Gestalt psychology principle of *proximity* states that items located close to one another will be perceived as belonging to a group, and items located farther apart will be perceived as not belonging together. We applied this principle to help reinforce the subgrouping within the question. We started by moving the first response option onto its own line of text and adding some extra space between it and the question stem. We then moved each response option onto its own line and arranged them vertically underneath the question stem so that they would no longer blend together. To help create the impression that the response options were all part of one group, we placed them in close vertical proximity to one another and spaced them equally. We also indented them a few spaces to the right underneath the question stem to reinforce the subgrouping we were creating. Grouping and subgrouping of multiple questions is discussed in more depth in Chapter 6.

Guideline 4.12: Visually Standardize All Answer Spaces or Response Options

Another problem with the poor design that we needed to address was that some response options stood out visually more than others, making them more likely to be seen and selected. "Shopping" stood out because it was bolded, "dining"

because it was in reverse print, and "hanging out with friends" because it was underlined. Our solution to this problem was to standardize the design properties of all of the response options. The first thing we did was make sure they were all the same readable size with the same character spacing. We then changed them all to the same font.

We chose Times New Roman because of its readability and professionalism (i.e., compared to the Disney Script used for "conducting business"). Finally, we removed variations due to color, contrast (bolding), underlining, and reverse print. The resulting uniformity makes the response options easier to process and helps ensure that they will be processed equally. Incidentally, the similarity across response options also helps them appear as a subgroup within the larger question group (i.e., the Gestalt psychology principle of *similarity* says that items that appear regular and similar will be perceived as belonging together). In addition to making these changes, we reordered the response options so the "other" option was located at the end of the list so that respondents would process all of the response options before getting to it.

Guideline 4.13: Use Visual Design Properties to Emphasize Elements that Are Important to the Respondent and to De-Emphasize Those that Are Not

In the poor design, the words "describes" and "recent visit" in the question stem are emphasized with bolding and italics, respectively. However, these words seem no more important to the respondents' understanding of the response task or the

question than any others in the stem. As a result, the bolding and italics were removed in the revised design. Instead of emphasizing these words, we opted to use underlining to emphasize the word "one" in order to draw the respondents' attention to the fact that they should select only one of the response options. The choice of underlining for this purpose works quite well for paper surveys but should be carefully considered for web surveys because underlining already has a predefined meaning on the Internet, especially when combined with the color blue. Underlining on the Web often denotes a clickable link, although many web designers use underlining inconsistently, and sometimes links are not underlined.

Nevertheless, we use underlining in the same way in the merit-based pay example in the bottom of the figure to draw respondents' attention to the fact that this question is about elementary school teachers only. In this example we also face another common problem: the need to include extra information for survey processing reasons. In this case, the extra information is numbers located inside the check boxes to assist with data entry. Because they are unimportant to the respondent, these numbers should be deemphasized if they cannot be eliminated altogether. We do this by manipulating the properties of size, contrast, and location. The numbers are made smaller and lighter to make them less obvious but still visible to the astute data enterer. They are then relocated from the center to outside the check box, where they are less likely to be noticed by respondents but can still easily be used for data entry.

Guideline 4.14: Use Design Properties with Consistency and Regularity

This general guideline may be the most important. Even if the meaning of a design element or property is not immediately intuitive, the respondent has a better chance of learning its meaning and applying it throughout the questionnaire if it is used consistently, both within and across questions. However, if design elements and properties are used inconsistently, like the use of the bolding in the shopping mall example, the respondent has to relearn their meaning at each use. Doing so may require more patience and mental energy than some respondents are willing to expend.

A good rule of thumb is to use each design element or property for only one purpose. For example, no matter what question or what part of the question it is used in, underlining is only used to draw attention to important words, white square boxes are only used as answer spaces, bolding is only used to distinguish between the question stem and the answer options depending on the particular needs of one's survey. One can choose to use these design properties to convey different meanings than in this example, but the important thing is that they be used with consistency and regularity.

Taken together, these visual design guidelines (4.10–4.14) helped us organize all of the information in the questions in Figure 4.19 to make these questions more easily perceived and processed and, most important, to make it easier for respondents to provide a response. In addition to easing the response task, these changes also added an air of professionalism to the questions, thereby increasing

the likelihood that they would be taken seriously by potential respondents (i.e., perhaps increasing rewards and trust). In the remainder of this section we present several more general guidelines dealing with instructions, the response task, and organization.

Guideline 4.15: Make Sure the Words and Visual Elements that Make Up the Question Send Consistent Messages

One of the biggest lessons we have learned since the last edition of this book was published is just how influential visual design elements can be. We know of multiple cases where surveyors made design decisions with the intention of easing the response task but with the result that the design features they introduced biased the results. In one national survey, for example, the designers sorted related response options into two groups based on their content. Pretesting revealed that the subgrouping caused respondents to make mistakes in answering the question. Many respondents attempted to mark multiple answers, but because the survey was only designed to accept one answer, they inadvertently erased their first response when they entered their second.

Figure 4.20 shows examples of this type of design from a set of web survey experiments we undertook to examine the effects of subgrouping response options in this way (Smyth, Dillman, Christian, & Stern, 2006b). When the response options were subgrouped with no instruction to select the best answer (not shown in figure), 70% of respondents marked answers within both groups

compared to 41% when the response options were not subgrouped. Adding the instruction "Please select the best answer" to the subgrouped version reduced the number of respondents marking answers within both groups to 66%. The instruction also reduced the mean number of options respondents selected, as more respondents limited themselves to selecting one option from each subgroup instead of multiple options from each subgroup. Overall, these findings suggest that the subgrouping communicated to respondents that they should select answers from both subgroups. Within this context the instruction to select the best answer appears to have been interpreted as "select the best answer *from each subgroup.*" In other words, the visual information provided by the subgrouping influenced how the instruction was interpreted.

Figure 4.20 Make sure verbal and visual design elements send a consistent message.

Instruction and visual arrangement of response options contradict one another

Q18. What best describes the benefit of the Student Recreation Center? Please select the best answer.

Health Benefits
- ☐ The variety of physical fitness offerings
- ☐ The health and wellness offerings
- ☐ Helps reduce stress

Academic Benefits
- ☐ Improves academic productivity
- ☐ Enhances learning experience
- ☐ Provides information for students to learn about their health

- ☐ Don't Know

66% of respondents marked an answer in both the top and the bottom half of the response options

A revision with instruction and unnecessary subgrouping of response options removed

Q18. What best describes the benefit of the Student Recreation Center?

- ☐ The variety of physical fitness offerings
- ☐ The health and wellness offerings
- ☐ Helps reduce stress
- ☐ Improves academic productivity
- ☐ Enhances learning experience
- ☐ Provides information for students to learn about their health
- ☐ Don't Know

41% of respondents marked an answer in both the top and the bottom half of the response options

Another possible revision reincorporating instruction

Q18. What best describes the benefit of the Student Recreation Center? Please select the best answer.

- ☐ The variety of physical fitness offerings
- ☐ The health and wellness offerings
- ☐ Helps reduce stress
- ☐ Improves academic productivity
- ☐ Enhances learning experience
- ☐ Provides information for students to learn about their health
- ☐ Don't Know

Source: "Effects of Using Visual Design Principles to Group Response Options in Web Surveys," by J. D. Smyth, D. A. Dillman, L. M. Christian, and M. J. Stern, 2006b, *International Journal of Internet Science, 1*(1), pp. 6–16.

This experiment provides an excellent example of how verbal and visual elements can contradict one another, leading to errors in responses. Perhaps more important, though, it demonstrates the importance of stepping back and looking at question construction holistically to ensure that both the words and the visual design of the question are sending a consistent message about the meaning of the question and the response task. In many ways the guidelines we present in this chapter and the next provide the tools for doing just that.

Guideline 4.16: Integrate Special Instructions into the Question Where They Will Be Used Rather than Including Them as Free-Standing Entities

Frequently it is necessary to provide a special instruction to clarify a question. This leads to the undesirable practice of placing instructions outside of the question and emphasizing them with boxes or perhaps a different color. The problem with this practice is that once people have gotten into the routine of completing a questionnaire, the marking of an answer leads to the immediate search for the next question. As a result, free-standing instructions tend to be skipped entirely. The example in Figure 4.21 shows that such instructions are most likely to be properly applied if they are expressed as part of the query itself rather than placed as a separate entity (Christian & Dillman, 2004).

In the first layout, an instruction to skip this question and move on to the next if it does not apply is located below both the question stem and the response

options. In this design, 40% of respondents marked "no" and only 5% left the question blank and moved on. In the second layout, the instruction is moved up to a more integrated location between the question stem and the response options. Placing the instruction here resulted in 19% of respondents marking "no" and a full 26% leaving the question blank and moving on. When the instruction was located where it was needed to help respondents decide whether and how they should answer the question, more people were able to successfully apply it. In contrast, when the instruction was located as a free-standing entity outside the question stem and answer categories, many respondents had probably already marked an answer before they even noticed it. The fact that 11% of respondents to this version (compared to 3% when the instruction was integrated) left the next question (Question 9) blank suggests that when some respondents got to the instruction located below the response options, they applied it to the wrong question altogether. The third layout, which has shown some promise in cognitive interviews, is another possible way the instruction could be integrated with the question stem.

As this example demonstrates, it is not enough to simply move instructions from the front of the questionnaire or from a separate booklet into the appropriate question subgrouping. Rather, even within a single question, in order to be effective, instructions need to be strategically located where they will be used (the location of instructions is discussed in more depth with respect to establishment surveys in Chapter 12).

Figure 4.21 Integrate special instructions into the question stem.

Instruction placed outside of the navigational path

8. Have one-on-one meetings with professors contributed significantly to your WSU education?

☐ Yes
☐ No

If you haven't had many one-on-one meetings, just skip to Question 9.

40% marked No
5% left Question 8 blank
11% left Question 9 blank

A revision with the instruction placed within the navigational path

8. Have one-on-one meetings with professors contributed significantly to your WSU education?
If you haven't had many one-on-one meetings, just skip to Question 9.

☐ Yes
☐ No

19% marked No
26% left Question 8 blank
3% left Question 9 blank

Another possible revision with the instruction integrated with the question stem and visually distinguished using italics

8. Have one-on-one meetings with professors contributed significantly to your WSU education?
If you haven't had many one-on-one meetings, just skip to Question 9.

☐ Yes
☐ No

Source: "The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions," by L. M. Christian and D. A. Dillman, 2004, *Public Opinion Quarterly, 68*(1), pp. 58–81.

Guideline 4.17: Separate Optional or Occasionally Needed Instructions from the Question Stem by Font or Symbol Variation

When respondents begin to fill out a questionnaire, they are learning how the questionnaire works, including what must be read and what can be skipped. Requiring them to read through a great deal of material that does not apply or that can be skipped without negative consequences encourages the habit of skipping words and phrases. For these reasons a distinction should be made between words that are essential for every person to read and those that may be needed by only some respondents. There are many different reasons that reading a particular instruction may be optional. Perhaps it is because the instruction "put an X in the appropriate box" is the same instruction used for a previous question, and many respondents will remember that. It may also be that only a few respondents need the information, such as in the case of the instruction used in Figure 4.21 ("If you haven't had many one-on-one meetings, just skip to Question 9"). To avoid presenting information that respondents already know, or that applies to relatively few of them, distinguish this information from the query by the use of either italics (as shown for the second revision in Figure 4.21) or a symbol variation (e.g., putting it in parentheses).

Guideline 4.18: Organize Each Question in a Way that Minimizes the Need to Reread Portions in Order to Comprehend the Response Task

The goal underlying this guideline is efficiency for the respondent. In Figure 4.22, a recreated excerpt from the 1993 U.S. Census of Agriculture, it is inefficient for respondents to read in great depth about what land to consider in their answer before even knowing what the question is asking. The inevitable result is that it will be necessary to reread the information after discovering what the question is asking. The drawback to such inefficiency is that respondents may become frustrated and unwilling to retrace those steps and therefore may give a wrong answer or no answer at all. In this case, the problem is confounded by a visual layout that makes it somewhat unclear what navigational path is to be followed (i.e., what information is to be read in what order).

A more effective organization of the information is shown in the revision in the bottom panel of Figure 4.22. The revision allows respondents to know at the beginning that they are being asked to report the number of acres they own; they are then given instructions on what to include and exclude. The important implication of this principle is that no amount of visual redesign can compensate for poorly worded questions or unorganized information, which, once read, leave the respondent unclear about precisely what to do.

Figure 4.22 Poor information organization with unclear navigational path.

Poor information organization and lack of navigational path

CENSUS USE ONLY	035	036	037	038	039	040	041	042

SECTION 1 **ACREAGE IN 1992** – Report land owned, rented, or used by you, your spouse, or by the partnership, corporation, or organization for which you are reporting. Include ALL LAND, REGARDLESS OF LOCATION OR USE – cropland, pastureland, *rangeland, woodland, idle land, house lots, etc.*

S1

If the acres you operated in 1992 changed during the year, refer to the INFORMATION SHEET, section 1.

		None	Number of acres
1.	All land owned	☐	043

Better information organization and creation of clear navigational path

1. **How many acres of land did you own in 1990? You should report all land (crop land, pasture land, rangeland, woodland, idle land, house lots, etc.) regardless of location, owned by you, your spouse, or by the partnership, corporation, or organization for which you are reporting.** ***(If the acres you operated in 1990 changed during the year, refer to the information sheet, Section I.)***

_____Number of acres owned

Guideline 4.19: Choose Line Spacing, Font, and Text Size to Ensure the Legibility of the Text

Even a very well-worded question can be difficult for respondents to process if it is not designed in a legible way. Enhancing legibility means choosing an appropriate font, font size, and line length. With respect to fonts, one should avoid script fonts because they can be very difficult to read (e.g., *Brush Script MT*, *Walt Disney Script*, *Edwardian Script ITC*, *Freestyle Script*). Instead, serif or sans serif fonts should be used. Examples of serif fonts are Times New Roman, Garamond, Century, and Georgia. Each of these fonts has added detail, or *serifs*, at the end of the strokes that make up the structure of the letters. In contrast, sans serif fonts do not have the added details (e.g., Arial, Verdana, Tahoma, and Latha). Although both serif and sans serif fonts work well on paper, sans serif fonts are commonly preferred for web readability. Generally, one should also choose proportionally spaced fonts (e.g., Arial or Times New Roman) rather than monospace fonts (e.g., Courier New).

To some degree, the font size one chooses will depend on the survey population. A good rule of thumb is to use 10- to 12-point fonts for most populations but larger fonts for older populations. An additional consideration for web surveyors is that font preferences are often set on the user's computer, giving the designer little control. Additionally, font sizes appear different on screen than on paper, and the same font size may appear larger or smaller depending on factors such as the user's screen resolution. Thus, web designers are advised to seek additional resources on how to ensure legibility of text (more on this in Chapter 6).

With respect to line length, readers may have difficulty tracking along the lines, reading evenly, and finding their place at the beginning of the next line on the return sweep when text lines are too long. In comparison,

excessively short

lines of text require

almost constant eye

motion and frequent

return sweeps that

can become overly

burdensome.

Thus, a more moderate line length of 3 to 5 inches is recommended.

CONCLUSION

Although crafting survey questions may seem simple, we have demonstrated in this chapter how it requires attending to many details at once to help ensure that respondents process all of the component parts and comprehend the question as intended so that they can report an accurate answer. We have also discussed a

holistic approach that highlights the many aspects surveyors need to think about when crafting survey questions. This approach considers whether an open- or closed-ended question is best for each concept of interest in the survey and requires thinking about how all of the components of a question (the question stem, any additional instructions, and answer spaces or response options) work together to form the entire question stimulus.

Drawing on the considerable amount of research on the importance of the visual design and presentation of survey questions, we have demonstrated how crafting effective survey questions for mail and web surveys involves not only choosing words to form clear questions but also deciding how the components of the question are presented visually to respondents. Within this holistic framework, we have offered general guidelines to help surveyors as they choose the words to form questions and visually design and present the components of the question. This second step is an important one that is often neglected in other guides for crafting questions.

The guidelines we have presented in this chapter apply to nearly all survey questions. However, this is only the first of two chapters devoted to crafting survey questions. In the next chapter we shift our focus to crafting specific types of open-ended and closed-ended questions. Because each of these different question types has a different goal and a slightly different configuration of question components, each is subject to its own strengths and challenges. Although the general guidelines presented in the current chapter still apply to

these question types, we take them a step further to focus on both the wording and visual presentation of the question stem, any additional instructions, the answer spaces, and response options that are unique to these question types.

List of Guidelines

Guidelines for Choosing Words and Forming Questions

Guideline 4.1: Make sure the question applies to the respondent

Guideline 4.2: Make sure the question is technically accurate

(continued)

Guideline 4.3: Ask one question at a time

Guideline 4.4: Use simple and familiar words

Guideline 4.5: Use specific and concrete words to specify the concepts clearly

Guideline 4.6: Use as few words as possible to pose the question

Guideline 4.7: Use complete sentences with simple sentence structures

Guideline 4.8: Make sure "yes" means yes and "no" means no

Guideline 4.9: Be sure the question specifies the response task

Guidelines for the Visual Presentation of Survey Questions

Guideline 4.10: Use darker and/or larger print for the question and lighter and/or smaller print for answer choices and answer spaces

Guideline 4.11: Use spacing to help create subgrouping within a question

Guideline 4.12: Visually standardize all answer spaces or response options

Guideline 4.13: Use visual design properties to emphasize elements that are important to the respondent and to deemphasize those that are not

Guideline 4.14: Use design properties with consistency and regularity

Guideline 4.15: Make sure the words and visual elements that make up the question send consistent messages

Guideline 4.16: Integrate special instructions into the question where they will be used rather than including them as free-standing entities

Guideline 4.17: Separate optional or occasionally needed instructions from the question stem by font or symbol variation

Guideline 4.18: Organize each question in a way that minimizes the need to reread portions in order to comprehend the response task

Guideline 4.19: Choose line spacing, font, and text size to ensure the legibility of the text