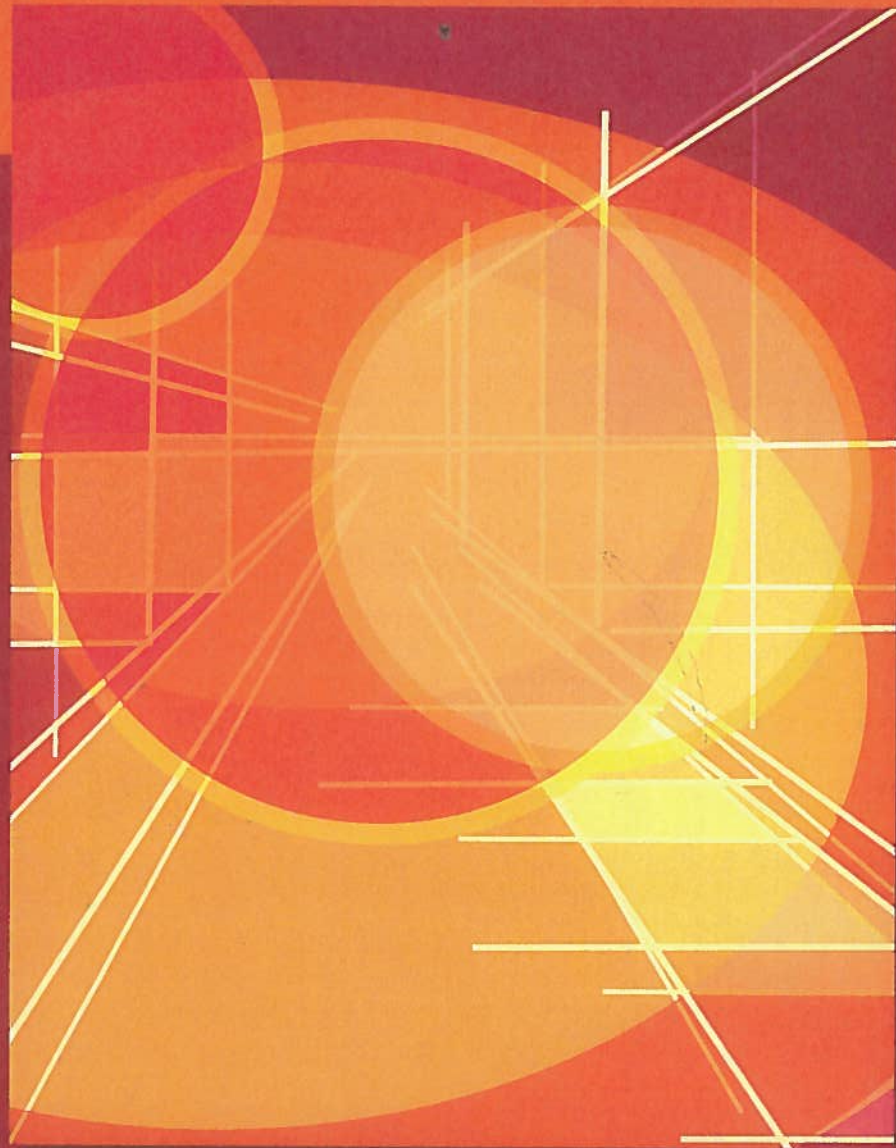


Kenneth S. Bordens • Bruce B. Abbott



Research Design and Methods

A Process Approach Eighth Edition

9

CHAPTER

CHAPTER OUTLINE

Survey Research

Designing Your Questionnaire

Writing Questionnaire Items

Assembling Your Questionnaire

Administering Your Questionnaire

Mail Surveys

Internet Surveys

Telephone Surveys

Group-Administered Surveys

Face-to-Face Interviews

A Final Note on Survey Techniques

Assessing the Reliability of Your Questionnaire

Assessing Reliability by Repeated Administration

Assessing Reliability With a Single Administration

Increasing Reliability

Assessing the Validity of Your Questionnaire

Acquiring a Sample for Your Survey

Representativeness

Sampling Techniques

Random and Nonrandom

Sampling Revisited

Sample Size

Summary

Key Terms

Using Survey Research

Gordon Allport (1954) characterized an attitude as “probably the most distinctive and indispensable concept in contemporary social psychology” (p. 43). Since Allport’s assessment, attitudes have transcended social psychology to become important in our everyday lives. We are surrounded by issues related to attitudes and their measurement. Pollsters and politicians are constantly measuring and trying to change our attitudes about a wide range of issues (such as abortion, the war on terrorism, and tax cuts). How and where we obtain information on these issues are also changing.

On November 4, 2008, a historic election took place in the United States. For the first time in history an African American was elected to the office of President of the United States. Not only did the 2008 election reflect a change in America’s willingness to vote for an African American candidate, it also reflected a change in how many citizens obtained their information on the candidates and the important political issues underlying the election. According to a 2009 survey conducted by the Pew Research Center, 74% of Internet users relied on the Internet to participate in or get information about the presidential election. More interestingly, there was a major increase in the percentage of adults in general as well as Internet users who obtain political news over the Internet (see Figure 9-1 for these trends).

The increased reliance on Internet sources for political news was true for a wide range of demographic groups. For example, the percentage of adults who sought political information online increased among all age groups from 2004 to 2008, with the greatest net increase among 18 to 24 year olds (a 21% increase). The increase was evident among all income groups measured (with the greatest increase among those earning less than \$30,000 per year) and among Democrats (a 10% increase), Republicans (a 9% increase), and independents (a 3% increase). Additionally, the Pew survey found that Obama supporters were more likely than opponent McCain supporters to engage in a variety of online political activities. For example, Obama supporters were more likely to use social networks (25%)

FIGURE 9

SOURCE: <http://www.pewresearch.org>
at the Web

than McCain (26% and

Survey
survey you
results. If
election t
on warran
sorts. You
you subscri
you took p

Even
tered surv
and candi
president’s
the course

Beacu
outs” of th
your own r
istration to
survey tech
out of the

SURVEY

Before we
the observ
and partici
about them

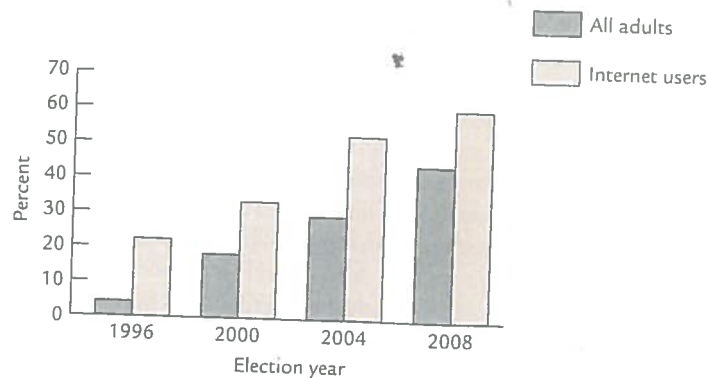


FIGURE 9-1 Trends in the use of the Internet to obtain political news.

SOURCE: <http://pewresearch.org/pubs/1192/internet-politics-campaign-2008>. Based on data provided at the Web site.

than McCain supporters (16%), and were more likely to post political content online (26% and 15% for Obama and McCain supporters, respectively).

Surveys are a widely used research technique. You may have participated in a survey yourself, or (perhaps more likely) you may have been the recipient of survey results. If you have answered a few questions from a local political party during election time, you have participated in a survey. Even those annoying questions on warranty registration cards that come with most products qualify as a survey of sorts. You are typically asked about your age, income, interests, magazines to which you subscribe, and so on. If you answered those questions and mailed back the card, you took part in a survey.

Even if you rarely participate in surveys, you are still likely to have encountered survey results. Political polls designed to gauge people's attitudes on key issues and candidates come out almost daily during election time. Polls about the U.S. president's approval rating, wars, and health care issues come out several times over the course of a year.

Because survey research is highly visible, you should understand the "ins and outs" of this important research technique. If you plan to use a survey technique in your own research, you should know about proper questionnaire construction, administration techniques, sampling techniques, and data analysis. Even if you never use survey techniques, understanding something about them will help you make sense out of the surveys that you are exposed to every day.

SURVEY RESEARCH

Before we discuss survey techniques, note the difference between the *field survey* and the observational techniques described in Chapter 8. In both naturalistic observation and participant observation, you simply observe behaviors and make copious notes about them. You do not administer any measures to your participants. Consequently,

you can only speculate about the motives, attitudes, and beliefs underlying the observed behaviors. In a field survey, you directly question your participants about their behavior (past, present, or future) and their underlying attitudes, beliefs, and intentions. From the data collected, you can draw inferences about the factors underlying behavior.

The inferences that you can draw from a field survey are limited by the fact that you do not manipulate independent variables. Instead, you acquire several (perhaps hundreds of) measures of the behaviors of interest. This purely correlational research strategy usually does not permit you to draw causal inferences from your data (see Chapter 4). For example, finding that political conservatism is a good predictor of voter choices does not justify concluding that political conservatism *causes* voter choices.

Instead, you use the field survey to evaluate specific attitudes such as those concerning issues surrounding nuclear disarmament, political candidates, or foreign imports. You also can use the field survey to evaluate behaviors. For example, you could design a questionnaire to determine which household products people use.

Surveys also have another important use: predicting behavior. Political polls often seek to predict behavior. Attitudes about political candidates are assessed, and then projections are made about subsequent voter behavior.

When you conduct survey research, you must ensure that your participants are treated ethically. One major ethical issue concerns whether and how you will maintain the *anonymity* of your participants and the *confidentiality* of their responses. Maintaining anonymity means that you guarantee there will be no way for the participants' names to be associated with their answers. This might be accomplished by instructing participants to mail back their questionnaires and informed-consent forms separately. No coding scheme would be used that would allow you to match up individual participants and their questionnaires. However, sometimes you may wish to code the questionnaires and informed-consent forms so that you can match them up later. You might do this, for example, if a participant has second thoughts about participating after the questionnaire has been returned. If so and you have promised your participants that their responses will remain anonymous, you must take steps to ensure that only authorized personnel associated with the research project can gain access to the code and only for the stated purpose.

Maintaining confidentiality means that you do not disclose any data in individual form, even if you know which participants filled out which questionnaires. If you promise your participants that their responses will remain confidential, ethical practice dictates that you report only aggregate results.

QUESTIONS TO PONDER

1. What are some of the applications of survey research?
2. Why is it important to know about survey methods, even if you do not intend to conduct surveys?
3. How does a field survey differ from other observational methods?
4. What are anonymity and confidentiality and why are they important?

DESIGN

The first
A clear,
interpre
area ma
nitions
are not
interpre

Have
question
1972). Y
too muc
confuse
to sum
should i
behavio
naire sh
informat

The
Howeve
the parti
Such ch
variables
correlate
graphic
tude tow
toward a

In a
designed
dicting v
specifica
date pref
criterion

The
tionnaire
evant to
will pay
interpret
item for

Writing

Writing
care and
first-draft

DESIGNING YOUR QUESTIONNAIRE

The first step in designing a questionnaire is to clearly define the topic of your study. A clear, concise definition of what you are studying will yield results that can be interpreted unambiguously. Results from surveys that do not clearly define the topic area may be confusing. It is also important to have clear, precise operational definitions for the attitudes or behaviors being studied. Behaviors and attitudes that are not defined precisely also may yield results that are confusing and difficult to interpret.

Having a clearly defined topic has another important advantage: It keeps your questionnaire focused on the behavior or attitude chosen for study (Moser & Kalton, 1972). You should avoid the temptation to do too much in a single survey. Tackling too much in a single survey leads to an inordinately long questionnaire that may confuse or overburden your participants. It also may make it more difficult for you to summarize and analyze your data (Moser & Kalton, 1972). Your questionnaire should include a broad enough range of questions so that you can thoroughly assess behavior but not so broad as to lose focus and become confusing. Your questionnaire should elicit the responses you are most interested in without much extraneous information.

The type of information gathered in a questionnaire depends on its purpose. However, most questionnaires include items designed to assess the characteristics of the participants, such as age, sex, marital status, occupation, income, and education. Such characteristics are called *demographics*. Demographics are often used as *predictor variables* during analysis of the data to determine whether participant characteristics correlate with or predict responses to other items in the survey. Other, nondemographic items also can be included to provide predictor variables. For example, attitude toward abortion might be used to predict voter preference. In this case, attitude toward abortion would be used as a predictor variable.

In addition to demographics and predictor variables, you will have items designed to assess the behavior of interest. For example, if you were interested in predicting voter preference, you would include an item or items on your questionnaire specifically to measure voter preference (e.g., asking participants to indicate candidate preferences). That item, or a combination of several items, would constitute the *criterion variable*.

The questions to which your participants will respond are the heart of your questionnaire. Take great care to develop questions that are clear, to the point, and relevant to the aims of your research. The time spent in this early phase of your research will pay dividends later. Well-constructed items are easier to summarize, analyze, and interpret than poorly constructed ones. The next section introduces several popular item formats and offers suggestions for writing good questionnaire items.

Writing Questionnaire Items

Writing effective questionnaire items that obtain the information you want requires care and skill. You cannot simply sit down, write several questions, and use those first-draft questions on your final questionnaire. Writing questionnaire items involves

writing and rewriting items until they are clear and succinct. In fact, having written your items and assembled your questionnaire, you should administer it to a pilot group of participants matching your main sample in order to ensure that the items are reliable and valid.

When writing questionnaire items, you may choose among several popular types. Here we discuss the open-ended, restricted, partially open-ended, and rating-scale item types.

Open-Ended Items Open-ended items allow the participant to respond in his or her own words. The following example might appear in a survey like the Pew Internet use survey:

How often did you use the Internet to get political news for the 2008 presidential election?

The participant writes an answer to the question in the space provided immediately below. Such information may be more complete and accurate than the information obtained with a restricted item (discussed next). A drawback to the open-ended item is that participants may not understand exactly what you are looking for or may inadvertently omit some answers. Thus, participants may fail to provide the needed information. Another drawback to the open-ended item is that it can make summarizing your data difficult. Essentially, you must perform a content analysis on open-ended answers. All of the methods and rules that we discussed in Chapter 8 would come into play. It may be tempting to interpret open-ended responses rather than just summarize them, running the risk of misclassifying the answers.

Restricted Items Restricted items (also called *closed-ended items*) provide a limited number of specific response alternatives. A restricted item with ordered alternatives lists these alternatives in a logical order, as shown in this item adapted from the Pew survey:

How often did you use the Internet to get political news during the 2008 presidential election campaign?

- ☐ Very often
- ☐ Sometimes
- ☐ Not too often
- ☐ Never

Note how the alternatives for this question go from very often to never. Participants would respond by checking the blank space to the left of the desired answer. However, other methods for recording choices can be used with restricted items. For example, you could use a number to the right of each alternative and have participants circle the numbers corresponding to their choices.

Use unordered alternatives whenever there is no logical basis for choosing a given order, as shown in this example from the Pew survey:

Do you think that the political information you obtained from the Internet during the 2008 presidential election campaign was generally accurate or inaccurate?

- ☐ Accurate
- ☐ Inaccurate
- ☐ Neither
- ☐ Don't know

Because there is no inherent order to the alternatives, other orders would serve just as well. For example, you just as easily could have put "Inaccurate" before "Accurate."

By offering only specific response alternatives, restricted items control the participant's range of responses. The responses made to restricted items are therefore easier to summarize and analyze than the responses made to open-ended items. However, the information that you obtain from a restricted item is not as rich as the information from an open-ended item. Participants cannot qualify or otherwise elaborate on their responses. Also, you may fail to include an alternative that correctly describes the participant's opinion, thus forcing the participant to choose an alternative that does not really fit.

Partially Open-Ended Items Partially open-ended items resemble restricted items but provide an additional, "other" category and an opportunity to give an answer not listed among the specific alternatives, as shown in this example adapted from the Pew survey:

In what capacity did you most use the Internet during the 2008 presidential election campaign?

- ☐ Post political content online
- ☐ Engage politically on an online social network
- ☐ Share political videos, pictures, or audio content
- ☐ Sign up for online political updates
- ☐ Donate money online
- ☐ Other (Specify) _____

Dillman (2000) offers several suggestions for formatting restricted and partially open-ended items. First, use a boldface font for the stem of a question and a normal font for response category labels (as we have done in the previous examples). This helps respondents separate the question from the response categories that follow. Second, make any special instructions intended to clarify a question a part of the question itself. Third, put check boxes, blank spaces, or numbers in a consistent position throughout your questionnaire (e.g., to the left of the response alternatives). Fourth, place all alternatives in a single column. Other tips offered by Dillman (2000) for constructing and formatting questionnaire items are summarized in Table 9-1.

TABLE 9-1 Suggestions for Writing Good Survey Items

SUGGESTION	EXAMPLE
Use simple rather than complex words.	Use "work" rather than "employment."
Make the stem of a question as short and easy to understand as possible, but use complete sentences.	"Would you like to study in America?"
Avoid vague questions in favor of more precise ones.	Use "How many years have you lived in your current house?" rather than "Years in your house."
Avoid asking for too much information. Respondents may not have an answer readily available.	Use a list of ordered alternatives rather than an open-ended question when asking how often the respondent does something.
Avoid "check all that apply" questions.	Instead of "check all that apply," list each item separately and have respondent indicate liking/disliking for each.
Avoid questions that ask for more than one thing.	Instead of asking "Would you like to study and then live in America?" ask "Would you like to study in America?" and "Would you like to live in America?" separately.
Soften the impact of potentially sensitive questions.	Instead of asking "Have you ever stolen anything?" ask "Have you ever taken anything without paying for it?"

SOURCE: After Dillman, 2000.

Rating Scales A variation on the restricted question uses a rating scale rather than response alternatives. A rating scale provides a graded response to a question:

How much confidence do you have that the political news you obtained from the Internet during the 2008 presidential campaign was accurate?



There is no set number of points that a rating scale must have. A rating scale can have as few as 3 and as many as 100 points. However, rating scales commonly do not exceed 10 points. A 10-point scale has enough points to allow a wide range of choice while not overburdening the participant. Scales with fewer than 10 points also are used frequently, but you should not go below 5 points. Many participants may not

want to u
and the p
Scales rar
among, ev

You a
might do
is told the
because th

With
of the rest
ticipant k
accurate i
the midpc
reasonable
points.

You n
participan
develop a
non under
the nature
that your s
scale. In fa
scale trans

1
Very
Weak
(a)

1
Very
Weak
(b)

1
Very
Weak
(c)

FIGURE 9-2 labeled, and

want to use the extreme values on a scale. Consequently, if you have a 5-point scale and the participant excludes the end points, you really have only three usable points. Scales ranging from 7 to 10 points leave several points for the participants to choose among, even if participants do avoid the extreme values.

You also must decide how to label your scale. Figure 9-2 shows three ways that you might do this. In panel (a), only the end points are labeled. In this case, the participant is told the upper and lower limits of the scale. Such labeled points are called *anchors* because they keep the participant's interpretation of the scale values from drifting.

With only the end points anchored, the participant must interpret the meaning of the rest of the points. In Figure 9-2(b), all points are labeled. In this case, the participant knows exactly what each point means and may consequently provide more accurate information. In Figure 9-2(c), the scale is labeled at the end points and at the midpoint. This scale provides three anchors for the participant. This scale is a reasonable compromise between labeling only the end points and labeling all the points.

You may be wondering whether labeling each point changes the way that the participant responds on the scale. The answer seems to be a qualified no. When you develop a measurement scale, you are dealing with (1) the psychological phenomenon underlying the scale and (2) the scale itself. Labeling each point does not change the nature of the psychological phenomenon underlying the scale. You can assume that your scale, labeled at each point, still represents the phenomenon underlying the scale. In fact, researchers have sometimes expressed a misguided concern about such scale transformations (Nunnally, 1967). Minor transformations of a measurement

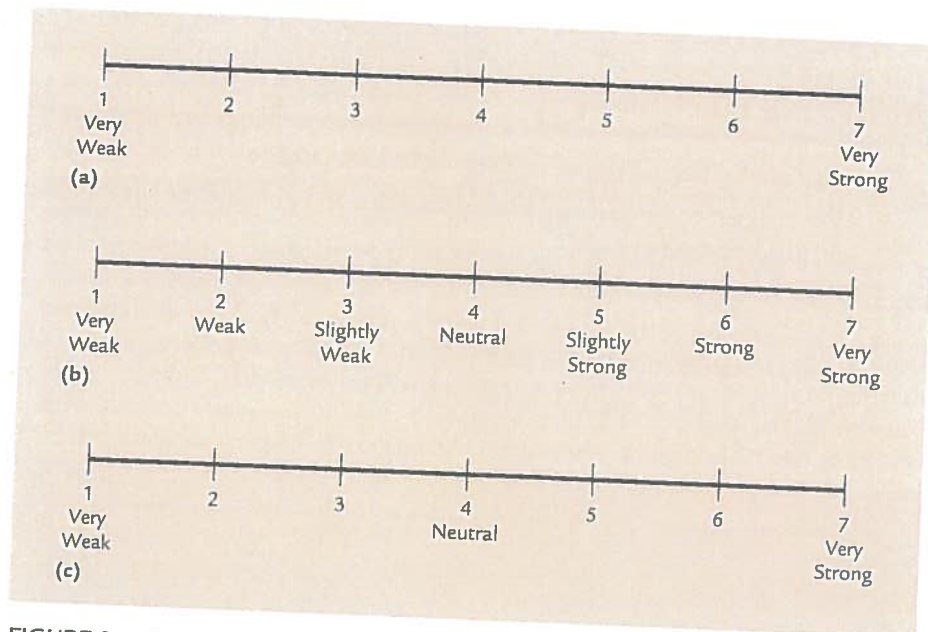


FIGURE 9-2 Three ways of labeling a rating scale: (a) end points only, (b) each point labeled, and (c) end points and midpoint labeled.

scale (such as labeling each point) probably do not affect its measurement properties or how well it represents the underlying psychological phenomenon being studied.

In the previous examples, participants respond by checking or circling the scale value that best represents their judgments. Alternative ways to format your scale give participants more flexibility in their responses. Figure 9-3 shows an example in which the end points are anchored and the participants are instructed to place a check or perpendicular line on the scale to indicate how they feel. To quantify the responses, you use a ruler to measure from an end point to the participant's mark. Your scale is then expressed in terms of inches or centimeters, and the resulting numbers are treated just like the numbers on a numbered scale.

Another variation on the rating scale is the *Likert scale*, which is widely used in attitude measurement research. A Likert scale provides a series of statements to which participants can indicate degrees of agreement or disagreement. Figure 9-4 shows two examples of formatting a Likert-scale item. In the first example, the attitude statement is followed by five blank spaces labeled from "Strongly Agree" to "Strongly Disagree." The participant simply checks the space that best reflects the degree of agreement or disagreement with each statement. The second example provides consecutive numbers rather than blank spaces and includes descriptive anchors only at the ends. Participants are instructed to circle the number that best reflects how much they agree or disagree with each statement. (For further information on Likert scaling, see Edwards, 1953).

A final note on rating scales is in order. Although rating scales have been presented in the context of survey research, be aware that rating scales are widely used in experimental research as well. Adapting rating scales to your particular research needs is a relatively simple affair. Anytime that your research calls for the use of rating scales, you can apply the suggestions presented here.

QUESTIONS TO PONDER

1. What are the steps involved in designing a questionnaire?
2. How do open-ended and restricted items differ, and what are the advantages and disadvantages of each?
3. What are the ways in which questionnaire items can be formatted?
4. What are some of the factors that you should pay attention to when constructing questionnaire items?
5. How do you design effective rating scales?



FIGURE 9-3 Rating scale formatted with no numbers. End points are labeled, and participants place marks on the line to indicate their responses.

Most political information on the Internet is accurate.

Strongly Agree Agree Neutral Disagree Strongly Disagree

(a)

Most political information on the Internet is accurate.

Strongly Agree Strongly Disagree

1 2 3 4 5

(b)

FIGURE 9-4 Samples showing Likert scales: (a) a standard Likert item on which the participant places a check in the blank under the statement that best reflects how he or she feels; (b) a five-point Likert scale using numbers that the participant circles.

Assembling Your Questionnaire

If your questionnaire is to be effective, its items must be organized into a coherent, visually pleasing format. This process involves paying attention to the order in which the items are included and to the way in which they are presented.

Dillman (2000) and Moser and Kalton (1972) agree that demographic items should not be presented first on the questionnaire. These questions, although easy to complete, may lead participants to believe that the questionnaire is boring. Dillman emphasizes the importance of the first question on a questionnaire. A good first question should be interesting and engaging so that the respondent will be motivated to continue. According to Dillman, the first question should apply to everybody completing the questionnaire, be easy so that it takes only a few seconds to answer, and be interesting. Of course, these rules are not carved in stone. If your research needs require a certain question to be presented first, that consideration should take precedence (Dillman, 2000).

Your questionnaire should have continuity; that is, related items should be presented together. This keeps your participant's attention on one issue at a time rather than jumping from issue to issue. Your questionnaire will have greater continuity if related items are grouped. An organized questionnaire is much easier and more enjoyable for the participant to complete, factors that may increase the completion rate. Continuity also means that groups of related questions should be logically ordered. Your questionnaire should read like a book. Avoid the temptation to skip around from topic to topic in an attempt to hold the attention of the participant. Rather, strive to build "cognitive ties" between related groups of items (Dillman, 2000).

The order in which questions are included on a questionnaire has been shown to affect the responses of participants. For example, McFarland (1981) presented

questions on a questionnaire ordered in two ways. Some participants answered a general question before specific questions, whereas others answered the specific questions first. McFarland found that participants expressed more interest in politics and religion when the specific questions were asked first than when the general questions were asked first. Sigelman (1981) found that question order affected whether or not participants expressed an opinion (about the popularity of the president), but only if the participants were poorly educated. Hence, question order may play a greater role for some participants than for others. Carefully consider your sample and the chosen topic when deciding on the order in which questions are asked.

The placement of items asking for sensitive information (such as sexual preferences or illegal behavior) is an important factor. Dillman (2000) suggests placing objectionable questions after less objectionable ones, perhaps even at the end of the questionnaire. Once your participants are committed to answering your questions, they may be more willing to answer some sensitive questions. Additionally, a question may not seem as objectionable after the respondent has answered previous items than if the objectionable item is placed earlier in the questionnaire (Dillman, 2000). You also should pay attention to the way that each page of your questionnaire is set up. There should be a logical “navigational path” (Dillman, 2000) that your respondent can follow. This path should lead the respondent through the questionnaire as if he or she were reading a book.

One way to accomplish this is to use appropriate graphics (e.g., arrows and other symbols) to guide respondents through the questionnaire. In fact, Dillman talks about two “languages” of a questionnaire. One language is verbal and relates to how your questions are worded. The other language is graphical and relates to the symbols and graphics used to guide respondents through the items on your questionnaire. Symbols and graphics can be used to separate groups of items, direct respondents where to go in the event of a certain answer (e.g., “If you answered ‘No’ to item 5, skip to item 7”) could be accompanied by an arrow pointing to item 7), or direct respondents to certain pages on the questionnaire. Dillman suggests the following three steps for integrating the verbal and graphical languages into an effective questionnaire:

1. Design a navigational path directing respondents to read all the information on a page.
2. Create effective visual navigational guides to help respondents stay on the navigational path.
3. Develop alternate navigational guides to help with situations where the normal navigational guide will be interrupted (e.g., skipping items or sections).

QUESTIONS TO PONDER

1. Why is the first question on a questionnaire so important?
2. What does it mean that a questionnaire should have continuity? Why is continuity important?
3. What is a questionnaire's navigational path, and why is it important?

ADM

After mail y
post i
ister y
Each

Mail

In a
comp
metho
mail t
propo
ticipa
who c
the in

Comb

strate
tive s
dents.
letter s
letter
in the
about
tionna
follow
plete
useful
an off
questi
T
week
for co
tionna
sent 2
ing th
the re
Y
token
Some
questi
rather
found

ADMINISTERING YOUR QUESTIONNAIRE

After you develop your questionnaire, you must decide how to administer it. You could mail your questionnaire to your participants, deliver your questionnaire via e-mail or post it on the Internet, telephone participants to ask the questions directly, administer your questionnaire to a large group at once, or conduct face-to-face interviews. Each method has advantages and disadvantages and makes its own special demands.

Mail Surveys

In a **mail survey**, you mail your questionnaire directly to your participants. They complete and return the questionnaire at their leisure. This is a rather convenient method. All you need to do is put your questionnaires into addressed envelopes and mail them. However, a serious problem called **nonresponse bias** occurs when a large proportion of participants fail to complete and return your questionnaire. If the participants who fail to return the questionnaire differ in significant ways from those who do return it, your survey may yield answers that do not represent the opinions of the intended population.

Combating Nonresponse Bias To reduce nonresponse bias, you should develop strategies to increase your return rate. Dillman (2000) notes that the single most effective strategy for increasing response rate is to make multiple contacts with respondents. Dillman suggests making four contacts via mail. The first consists of a *prenotice letter* sent to the respondent a few days before the questionnaire is sent. The prenotice letter should inform the respondent that an important questionnaire will be coming in the mail in a few days. It also should inform the respondent what the survey is about and why the survey will be useful. The second mailing would deliver the questionnaire itself, accompanied by a cover letter. The cover letter should include the following elements in the order listed (Dillman, 2000): the specific request to complete the questionnaire, why the respondent was selected to receive the survey, the usefulness of the survey, a statement of confidentiality of the respondent's answers, an offer of a token of appreciation (if such an offer is to be made), an offer to answer questions, and a real signature.

The third mailing would take the form of a *thank you postcard* sent a few days or a week after the questionnaire was mailed. The postcard should thank the respondent for completing the questionnaire and remind the respondent to complete the questionnaire if not already done. The fourth contact provides a *replacement questionnaire*, sent 2 to 4 weeks after the original questionnaire and accompanied by a letter indicating that the original questionnaire had not been received. The letter also should urge the respondent to complete the replacement questionnaire and return it.

You may be able to increase your return rate somewhat by including a small token of your appreciation, such as a pen or pencil that the participant can keep. Some researchers include a small amount of money as an incentive to complete the questionnaire. As a rule, it is better to send the token along with the questionnaire rather than make the token contingent upon returning the questionnaire. One study found that 57% of respondents returned a survey questionnaire when promised \$50

for its return whereas 64% returned the questionnaire when \$1 was included with it (James & Bolstein, 1990).

Ironically, smaller rewards seem to produce better results than larger ones (Kanuk & Berenson, 1975; Warner, Berman, Weyant, & Ciarlo, 1983). Dillman (2000) suggests that a \$1 token is preferred because it is easy to mail and seems to produce the desired results. Finally, monetary incentives work better than tangible rewards (Church, 1993).

A few factors that do *not* significantly affect response rate include questionnaire length, personalization, promise of anonymity, and inclusion of a deadline (Kanuk & Berenson, 1975). (For reviews of the research supporting these findings, see Kanuk & Berenson, 1975, and Warner et al., 1983.)

Internet Surveys

An increasingly popular method of administering questionnaires is to post them on the Internet. **Internet surveys** can be distributed via e-mail or listserves or posted on a Web site. Which method you use depends on the nature and purpose of your survey. E-mail surveys are easy to distribute but do not permit complex navigational designs (Dillman, 2000). Consequently, e-mail surveys are best for relatively short, simple questionnaires. Web-based surveys allow you to create and present more complex questionnaires that incorporate many of the design features discussed previously (Dillman, 2000). To aid you in the task of implementing a Web-based survey, commercial software packages are available that allow you to design sophisticated questionnaires for posting on a Web site. There is significant advantage to using the Internet to conduct a survey or recruit participants: You can reach a large body of potential participants with relative ease. Data can be collected quickly and easily, resulting in a large data set. You still need to consider the problem of nonresponse bias. As with the mail survey, you can combat this problem with prenotification. For an Internet survey a short text message to potential respondents is more effective than an e-mail notice (Bosnjak, Neubarth, Couper, Bandilla, & Kaczmirek, 2008).

There are also disadvantages to Internet surveys. As discussed in Chapter 6, a sample of respondents from the Internet may not be representative of the general population. According to a 2007 study by the U.S. Department of Commerce (2008), only 61.7% of households had access to the Internet in the home. Further, households with higher levels of education and income were more likely to have Internet access. Additionally, access was greater for Asians (75.5%) and Whites (67.0%) than Blacks (44.9%). Another disadvantage is that one must have the resources available to post a survey on the Internet. This requires computer space on a server and the ability to create the necessary Web pages or the resources to pay someone to create your net survey for you.

Despite the potential for biased samples in Internet surveys, there is evidence that the results obtained from Internet surveys are equivalent to the results obtained from paper-and-pencil surveys. Alan De Beuckeleer and Flip Lievens (2009) conducted a survey across 16 countries using both Internet and paper-and-pencil deliveries. The results showed that in all of the countries the Internet and paper-and-pencil surveys returned equivalent results. De Beuckeleer and Lievens (2009) concluded that data collected with the two methods could be combined because the two methods

produced such highly similar data. In another study, Christopher Fleming and Mark Bowden (2009) found that the sample demographics of an Internet and a mail survey on travel preferences did not differ significantly.

In both of the studies just cited, the topics of the surveys were not sensitive or controversial. There is some evidence that the equivalence of Internet and conventional methods may not apply to more sensitive topics (DiNitto, Busch-Armendariz, Bender, Woo, Tackett-Gibson, & Dyer, 2009). DiNitto, et al. conducted a survey over the Internet and by telephone asking men about sexual assault behaviors. The results showed that respondents in both types of survey reported sexual assault behavior. However, a wider variety of sexual assault behaviors were reported by respondents to the telephone survey.

So, where does this leave us? It would appear that Internet surveys may produce comparable results to other survey methods for nonsensitive issues. You can be reasonably confident that your Internet survey on such issues will yield data that are highly similar to data collected with more conventional methods. However, you must exercise more caution when surveying about sensitive behaviors. In the latter case, an Internet survey may produce results that differ from more conventional methods.

Telephone Surveys

In a **telephone survey**, you contact participants by telephone rather than by mail or via the Internet. You can ask some questions more easily over the telephone than you can in written form. Telephone surveys can be done by having an interviewer ask respondents a series of questions or by interactive voice response (IVR). Telephone surveys using live interviewers have lost popularity as new technologies have become available. IVR surveys involve respondents using a touch-tone telephone to respond to a series of prerecorded questions. Modern IVR technologies also allow respondents to provide verbal answers in addition to numeric responses.

Telephone surveys may not be the best way to administer a questionnaire. The plethora of “junk calls” to which the population is exposed has given rise to a backlash against telephone intrusions. Laws have been passed on the state and federal level protecting people from unwanted calls, making it more difficult to reach prospective respondents. These laws, combined with caller ID and answering machines (which allow residents to screen their calls), make the telephone a less attractive medium for surveys now than in the past.

Group-Administered Surveys

Sometimes you may have at your disposal a large group of individuals to whom you can administer your questionnaire. In such a case, you design your questionnaire as you would for a mail survey but administer it to the assembled group. For example, you might distribute to a first-year college class a questionnaire on attitudes toward premarital sex. Using such a captive audience permits you to collect large amounts of data in a relatively short time. You do not have to worry about participants misplacing or forgetting about your questionnaire. You also may be able to reduce any volunteer bias, especially if you administer your questionnaire during a class period. People may participate because very little effort is required.

As usual, this method has some drawbacks. Participants may not treat the questionnaire as seriously when they fill it out as a group as when they fill it out alone. Also, you may not be able to ensure anonymity in the large group if you are asking for sensitive information. Participants may feel that other participants are looking at their answers. (You may be able to overcome this problem by giving adjacently seated participants alternate forms of the questionnaire.) Also, a few participants may express hostility about the questionnaire by purposely providing false information.

A final drawback to group administration concerns the participant's right to decline participation. A participant may feel pressure to participate in your survey. This pressure arises from the participant's observation that just about everyone else is participating. In essence, a conformity effect occurs because completing your survey becomes the norm defined by the behavior of your other participants. Make special efforts to reinforce the understanding that participants should not feel compelled to participate.

Face-to-Face Interviews

Still another method for obtaining survey data is the **face-to-face interview**. In this method, you talk to each participant directly. This can be done in the participant's home or place of employment, in your office, or in any other suitable place. If you decide to use a face-to-face interview, keep several things in mind. First, decide whether to use a structured interview or an unstructured interview. In a *structured interview*, you ask prepared questions. This is similar to the telephone survey in that you prepare a questionnaire in advance and simply read the ordered questions to your participants. In the *unstructured interview*, you have a general idea about the issues to discuss. However, you do not have a predetermined sequence of questions.

An advantage of the structured interview is that all participants are asked the same questions in the same order. This eliminates fluctuations in the data that result from differences in when and how questions are asked. Responses from a structured interview are therefore easier to summarize and analyze. However, the structured interview tends to be inflexible. You may miss some important information by having a highly structured interview. The unstructured interview is superior in this respect. By asking general questions and having participants provide answers in their own words, you may gain more complete (although perhaps less accurate) information. However, responses from an unstructured interview may be more difficult to code and analyze later on. You can gain some advantages of each method by combining them in one interview. For example, begin the interview with a structured format by asking prepared questions; later in the interview, switch to an unstructured format.

Using the face-to-face interview strategy leads to a problem that is not present in mail or Internet surveys but is present to some extent in telephone surveys: The appearance and demeanor of the interviewer may affect the responses of the participants. Experimenter bias and demand characteristics become a problem. Subtle changes in the way in which an interviewer asks a question may elicit different answers. Also, your interviewer may not respond similarly to all participants (e.g., an interviewer may react differently to an attractive participant than to an unattractive one). This, too, can affect the results.

The best way to combat this problem is to use interviewers who have received extensive training in interview techniques. Interviewers must be trained to ask questions in the same way for each participant. They also must be trained not to emphasize any particular words in the stem of a question or in the response list. The questions should be read in a neutral manner. Also, try to anticipate any questions that participants may have and provide your interviewers with standardized responses. This can be accomplished by running a small pilot version of your survey before running the actual survey. During this pilot study, try out the interview procedure on a small sample of participants. (This can be done with just about anyone, such as friends, colleagues, or students.) Correct any problems that arise.

Another problem with the interview method is that the social context in which the interview takes place may affect a participant's responses. For example, in a survey of sexual attitudes known as the "Sex in America" survey (Michael, Gagnon, Laumann, & Kolata, 1994), some questions were asked during a face-to-face interview. Some participants were interviewed alone whereas others were interviewed with a spouse or other sex partner present. Having the sex partner present changed the responses to some questions. For example, when asked a question about the number of sex partners one had over the past year, 17% of the participants interviewed alone reported two or more. When interviewed with their sex partner present, only 5% said they had two or more sex partners. It would be most desirable to conduct the interviews in a standardized fashion with only the participant present.

A Final Note on Survey Techniques

Although each of the discussed techniques has advantages, the mail survey has been the most popular. The mail survey can reach large numbers of participants at a lower cost than either the telephone survey or the face-to-face interview (Warner et al., 1983) and produces data that are less affected by *social desirability effects* (answering in a way that seems socially desirable). For these reasons, consider mail surveys first.

After designing your questionnaire and choosing a method of administration, the next step is to assess the reliability and validity of your questionnaire. This is typically done by administering your questionnaire to a small but representative sample of participants. Based on the results, you may have to rework your questionnaire to meet acceptable levels of reliability and validity. In the next sections, we introduce you to the processes of evaluating the reliability and validity of your questionnaire.

QUESTIONS TO PONDER

1. What are the different ways of administering a questionnaire?
2. What are the advantages and disadvantages of the different ways of administering a questionnaire?
3. What is nonresponse bias and what can you do to combat it?
4. How do social desirability effects affect your decision about how to administer a questionnaire?

ASSESSING THE RELIABILITY OF YOUR QUESTIONNAIRE

Constructing a questionnaire is typically not a one-shot deal. That is, you don't just sit down and write some questions and magically produce a high-quality questionnaire. Developing a quality questionnaire usually involves designing the questionnaire, administering it, and then evaluating it to see if it does the job.

One dimension you must pay attention to is the reliability of your questionnaire. In Chapter 5, we defined *reliability* as the ability of a measure to produce the same or highly similar results on repeated administrations. This definition extends to a questionnaire. If, on testing and retesting, your questionnaire produces highly similar results, you have a reliable instrument. In contrast, if the responses vary widely, your instrument is not reliable (Rogers, 1995).

In Chapter 5, we described two ways to assess the reliability of a measure: the test-retest method and the split-half method. In the next sections, we discuss the application of these two methods when assessing the reliability of a questionnaire.

Assessing Reliability by Repeated Administration

Evaluating test-retest reliability is the oldest and conceptually simplest way of establishing the reliability of your questionnaire. You simply administer your questionnaire, allow some time to elapse, and then administer the questionnaire (or a parallel form of it) again to the same group of participants. Although this method is relatively simple to execute, you need to consider some issues before using it.

First, you must consider how long to wait between administrations of your questionnaire. An intertest interval that is too short may result in participants remembering your questions and the answers they gave. This could lead to an artificially high level of test-retest reliability. If, however, you wait too long, test-retest reliability may be artificially low. According to Tim Rogers (1995), the intertest interval should depend on the nature of the variables being measured, with an interval of a few weeks being sufficient for most applications. Rogers suggests that test-retest methods may be particularly problematic when applied to the following:

1. *Measuring ideas that fluctuate with time.* For example, an instrument to measure attitudes toward universal health care should not be evaluated with the test-retest method because attitudes on this topic seem to shift quickly.
2. *Issues for which individuals are likely to remember their answers on the first testing.*
3. *Questionnaires that are very long and boring.* The problem here is that participants may not be highly motivated to accurately complete an overly long questionnaire and therefore may give answers that reduce reliability.

Some of the problems inherent in using the *same* measure on multiple occasions can be avoided by using alternate or parallel forms of your questionnaire for multiple testing sessions. As noted in Chapter 5, the type of reliability being assessed with this technique is known as parallel-forms reliability (Rogers, 1995).

For the parallel-forms method to work, the two (or more) forms of your questionnaire must be equivalent so that direct comparison is meaningful. According to Rogers (1995), parallel forms should have the same number of items and the same response format, cover the same issues with different items, be equally difficult, use the same instructions, and have the same time limits. In short, the parallel versions of a test must be as equivalent as possible (Rogers, 1995).

Although the parallel-forms method improves on the test-retest method, it does not solve all the problems associated with multiple testing. Using parallel forms does not eliminate the possibility that rapidly changing attitudes will result in low reliability. As with the test-retest method, such changes make the questionnaire appear less reliable than it actually is. In addition, practice effects may occur even when alternate forms are used (Rogers, 1995). Even though you use different questions on the parallel form, participants may respond similarly on the second test because they are familiar with your question format.

Assessing Reliability With a Single Administration

Because of the problems associated with repeated testing, you might consider assessing reliability by means of a single administration of your questionnaire. As noted in Chapter 5, this approach involves splitting the questionnaire into equivalent halves and deriving a score for each half; the correlation between scores from the two halves is known as split-half reliability (Rogers, 1995). This technique works best when your survey is limited to a single specific area (e.g., sexual behavior) as opposed to multiple areas (sexual behavior and sexual attitudes).

Although the split-half method circumvents the problems associated with repeated testing, it introduces others. First, when you split a questionnaire, each score is based on a limited set of items, which can reduce reliability (Rogers, 1995). Consequently, the split-half method may underestimate reliability. Second, it is not clear how splitting should be done. If you simply do a first-half/second-half split, artificially low reliability may occur if the two halves of the form are not equivalent or if participants are less motivated to answer questions accurately on the second half of your questionnaire and therefore give inconsistent answers to your questions. One remedy for this is to use an odd-even split. In this case, you derive a score for the odd items and a score for the even items.

Perhaps the most desirable way to assess the split-half reliability of your questionnaire is to apply the Kuder-Richardson formula. This formula yields the average of all the split-half reliabilities that could be derived from splitting your questionnaire into two halves in every possible way. The resulting number (designated KR20) will lie between 0 and 1; the higher the number, the greater the reliability of your questionnaire. A KR20 of .75 indicates a "moderate" level of reliability (Rogers, 1995).

In cases in which your questionnaire uses a Likert format, a variation on the Kuder-Richardson formula known as *coefficient alpha* is used (Rogers, 1995). Like KR20, coefficient alpha is a score between 0 and 1, with higher numbers indicating greater reliability. Computation of this formula can be complex. For details, see a text on psychological testing (e.g., see Cohen & Swerdlik, 2010; Rogers, 1995).

Increasing Reliability

Regardless of the method you use to assess the reliability, there are steps you can take to increase the reliability of your questionnaire (Rogers, 1995):

1. Increase the number of items on your questionnaire. Generally, higher reliability is associated with increasing numbers of items. Of course, if your instrument becomes too long, participants may become angry, tired, or bored. You must weigh the benefits of increasing questionnaire length against possible liabilities.
2. Standardize administration procedures. Reliability will be enhanced if you treat all participants alike when administering your questionnaire. Make sure that timing procedures, lighting, ventilation, instructions to participants, and instructions to administrators are kept constant.
3. Score your questionnaire carefully. Scoring errors can reduce reliability.
4. Make sure that the items on your questionnaire are clear, well written, and appropriate for your sample (see our previous discussion on writing items).

QUESTIONS TO PONDER

1. What is meant by the reliability of a questionnaire and why is it important?
2. How do you assess reliability with repeated administrations?
3. How do you assess reliability with a single administration?
4. What steps can be taken to increase reliability?

ASSESSING THE VALIDITY OF YOUR QUESTIONNAIRE

In Chapter 5, we discussed the validity of a measure and described several forms of validity that differ in their method of assessment: content validity, criterion-related validity, construct validity, and face validity. As with other measures, a questionnaire must have validity if it is to be useful; that is, it must measure what it is intended to measure. For example, if you are designing a questionnaire to assess political attitudes, the questions on your test should tap into political attitudes and not, say, religious attitudes.

Here we review content validity, construct validity, and criterion-related validity as applied to a questionnaire (Rogers, 1995). In a questionnaire, *content validity* assesses whether the questions cover the range of behaviors normally considered to be part of the dimension that you are assessing. To have content validity, your questionnaire on political attitudes should include items relevant to all the major issues relating to such attitudes (e.g., abortion, health care, the economy, and defense). The *construct validity* of a questionnaire can be established by showing that the questionnaire's results agree with predictions based on theory.

Establishing the *criterion-related validity* of a questionnaire involves correlating the questionnaire's results with those from another, established measure. There are two ways to do this. First, you can establish *concurrent validity* by correlating your questionnaire's results with those of another measure of the *same* dimension administered at the same time. In the case of your questionnaire on political attitudes, you would correlate its results with those of another, established measure of political attitudes. Second, you can establish *predictive validity* by correlating the questionnaire's results with some behavior that would be expected to occur, given the results. For example, your questionnaire on political attitudes would be shown to have predictive validity if the questionnaire's results correctly predicted election outcomes.

The validity of a questionnaire may be affected by a variety of factors. For example, as noted earlier, how you define the behavior or attitude that you are measuring can affect validity. Validity also can be affected by the methods used to gather your data. In the "Sex in America" survey, some respondents were interviewed alone and others with someone else present. One cannot be sure that the responses given with another person present represent an accurate reflection of one's sexual behavior (Stevenson, 1995). Generally, methodological flaws, poor conceptualization, and unclear questions can all contribute to lowered levels of validity.

QUESTIONS TO PONDER

1. What is the validity of a questionnaire and why is it important?
2. What are the different types of validity you should consider?
3. What factors can affect the validity of your questionnaire?

ACQUIRING A SAMPLE FOR YOUR SURVEY

In Chapter 6, we distinguished between a population (all individuals in a well-defined group) and a sample (a smaller number of individuals selected from the population). Once you have designed and pretested your questionnaire, you then administer it to a group of participants. It is usually impractical to have everyone in the population (however that may be defined) complete your survey. Instead, you administer your questionnaire to a small sample of that population.

Proper sampling is a crucial aspect of sound survey research methodology. Without proper sampling, you can't generalize your results to your target population (e.g., accurately predict voter behavior in an election). Three sampling-related issues you must consider are representativeness, sampling technique, and sample size.

Representativeness

Regardless of the technique you use to acquire your sample, your sample should be representative of the population of interest. A **representative sample** closely matches the characteristics of the population. Imagine that you have a bag containing 300 golf balls: 100 are white, 100 are orange, and 100 are yellow. You then select a sample of

30 golf balls. A representative sample would have 10 balls of each color. A sample having 25 white and 5 orange would not be representative (the ratio of colors does not approximate that of the population) and would constitute a nonrepresentative or **biased sample**.

The importance of representative sampling is shown by the failure of a political poll taken during the 1936 presidential election. In that election, Alf Landon was opposing Franklin Roosevelt. The editors of the *Literary Digest* (a now-defunct magazine) conducted a poll by using telephone directories and vehicle registration lists to draw their sample. The final sample consisted of nearly 10 million people! The results showed that Landon would beat Roosevelt by a landslide. Quite to the contrary, Roosevelt soundly defeated Landon. Why was the poll so wrong?

The problem stemmed from the method used to obtain the sample. Fewer people owned a car or telephone in the 1930s than do today. In fact, very few owned either. Those who did own a telephone or car tended to be relatively wealthy and Republican. Consequently, most of the participants polled favored the Republican candidate. Unfortunately for the *Literary Digest*, this sample did not represent the population of voters, and the prediction failed. How could the editors have been so stupid? In fact, they weren't stupid. Such sampling techniques had been used before and worked. It was only in that particular election (in which people were clearly split along party lines) that the problem emerged (Hooke, 1983).

The *Literary Digest* poll failed because it used a biased source (car registration and telephone listings). Whatever source you choose, you should make an effort to determine whether it includes members from all segments of the population in which you have an interest. A good way to overcome the problem of biased source lists is to use multiple lists. For example, you could use the telephone book and vehicle registration and voter registration lists to select your sample.

Sampling Techniques

At the heart of all sampling techniques is the concept of *random sampling*. In random sampling, every member of the population has an equal chance of appearing in your sample. Whether or not a participant is included in your sample is based on chance alone. Sampling is typically done without replacement. Once an individual is chosen for your sample, he or she cannot be chosen a second time for that sample.

Random sampling eliminates the possibility that the sample is biased by the preferences of the person selecting the sample. In addition, random sampling affords some assurance that the sample does not bias itself. As an example of self-biasing, consider the following case. In 1976, Shere Hite published *The Hite Report: A Nationwide Study on Female Sexuality*, which was a survey of women's sexual attitudes and behaviors. Hite's sample was obtained by initially distributing questionnaires through national mailings to women's groups (the National Organization for Women, abortion rights groups, university women's centers, and others). Later, advertisements were placed in several magazines (the *Village Voice*, *Mademoiselle*, *Brides*, and *Ms.*) informing women where they could write for a copy of the questionnaire. Finally, the questionnaire was reprinted in *Oui* magazine in its entirety (253 women returned the questionnaire from *Oui*).

The random sample of several people may not include all women, and further bias is possible.

If you ask different questions, you may get different answers. In fact, Hite's study of women's sexuality was due for a similar problem. As with the women's group, the sample probably included a disproportionate number of women.

When you select a population, try to make sure that you are not selecting those who are more likely to respond.

In 1976, Hite's study was to gather responses from a sample of women. When the data were analyzed, the response rate was high. In the situation of nonresponse, men reported a higher population than the actual population. Extramarital sex was a common topic.

The random sample of women was a random sample. You can't expect a response rate of 100%.

Using a random sample, you can get a good idea of the population. Sampling, stratified sampling, and random sampling are all research methods.

Simple Random Sampling is a method of selecting a sample from a population. In this method, every member of the population has an equal chance of being selected. This is the simplest method of sampling.

The question that you should ask yourself at this point is, "Did Hite obtain a random sample of the population of women?" The answer is no. Hite's method had several problems. First, the memberships of the organizations that Hite contacted may not represent the population of women. For example, you cannot assume that members of NOW hold similar views, on the average, to those of the population of all women. Second, asking people through magazine ads to write in for questionnaires further biases the sample. Can you figure out why?

If you said that the people who write in for the questionnaires may be somehow different from those who do not, you are correct. Who would write in to obtain a questionnaire on sexuality? Obviously, women who have an interest in such an issue. In fact, Hite indicates that many of her participants expressed such an interest. One woman wrote, "I answered this questionnaire because I think the time is long overdue for women to speak out about their feelings about sex" (Hite, 1976, p. xxxii). As with the members of the women's organizations, you could question whether the women who wrote in for questionnaires are representative of all women. They probably are not.

When a sample is biased, the data obtained may not indicate the attitudes of the population as a whole. Hite concluded from her sample that women in this country were experiencing a "new sexuality." However, that new sexuality was limited to those women whose attitudes were similar to those who answered her questionnaires.

In 1983, Hite published *The Hite Report on Male Sexuality*. The method she used to gather data was similar to the one used in her earlier study of women. In this book, Hite responded to the criticisms of her method. She presented evidence that her sample of men was similar in age, religion, and education to the most recent census data. What was not clear, however, was whether or not the attitudes of the men who responded to her questionnaire were similar to those of the general population. As in the survey of women, the data obtained may not be representative of the population of men. Some evidence suggests they were not. Hite said that 72% of married men reported having had an extramarital affair. Is this an accurate estimate of the population or an estimate of a special subsection of the population? Apparently, it is the latter. Other surveys have found that about 25% of men report having had extramarital affairs.

The lesson of the Hite example is that you should make every effort to obtain a random sample. This may be difficult, especially if you are dealing with a sensitive topic. You could use some of the strategies previously suggested for reducing non-response bias (such as including a small reward or using follow-ups). If your sample turns out to be nonrandom and nonrepresentative, temper any conclusions you draw.

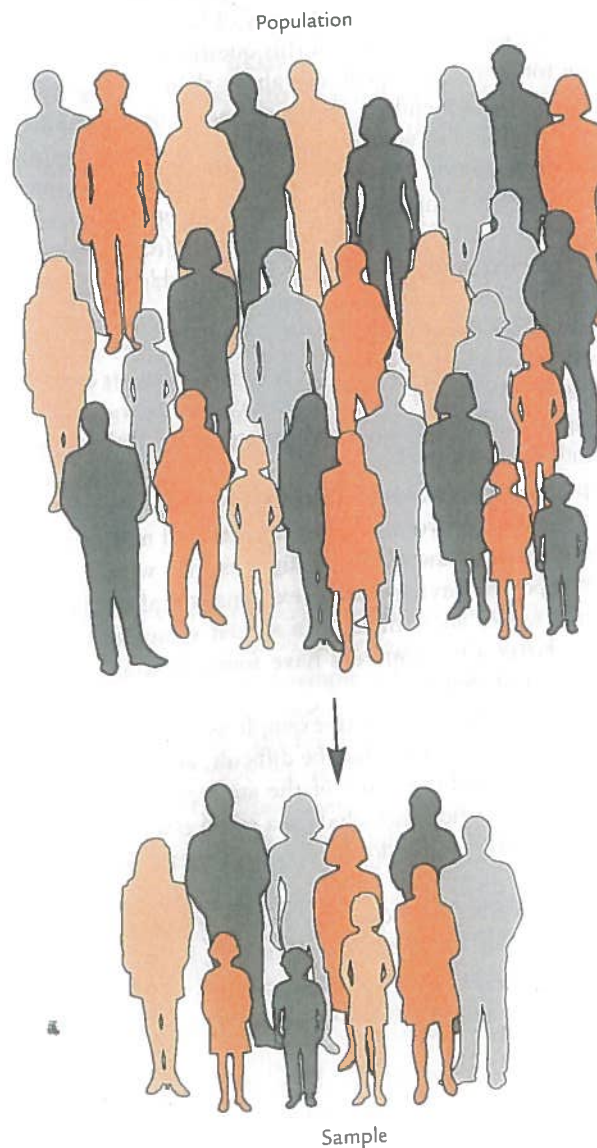
Using the proper sampling technique is one way to obtain a representative sample. Several techniques are available to you. Five of them (simple random sampling, stratified sampling, proportionate sampling, systematic sampling, and cluster sampling) are discussed next. These techniques are not mutually exclusive. Often researchers combine them to help ensure a representative sample of the population.

Simple Random Sampling Randomly selecting a certain number of individuals from the population is a technique called **simple random sampling**. Remember the golf ball example? A simple random sample of 50 would involve dipping your hand

into the bag 50 times, each time withdrawing a single ball. Figure 9-5 illustrates the simple random sampling strategy. From the population illustrated at the top of the figure, 10 participants are selected at random for inclusion in your survey.

In practice, selecting a random sample for a survey is more involved than pulling golf balls from a bag. Often it involves consulting a table of random numbers. The numbers in such a table have been chosen at random and then subjected to a number of statistical tests to ensure that they have the expected properties of random numbers. You can find a table of random numbers in the Appendix (Table 1A).

FIGURE 9-5 Example of simple random sampling. The people at the top of the figure represent the population, and the people at the bottom represent the randomly selected sample.



As
sample,
page of
Open y
235,035
that pag
sample.

A
survey i
area (th
then us
(e.g., 5,
called. (

unlistec
Eve
sample,
select p
ball exa
though
this pro
sample
Howev
ate to r
and on
unlikely
needed
random

QUES

- 1.
- 2.
- 3.
- 4.

Stratifi
sample
For exa
and H
stratun
segmer
fied sar
(gray a

As an example of how to use the table of random numbers to select a random sample, imagine you are using the telephone book as a source list. Starting on any page of the random number table, close your eyes and drop your finger on the page. Open your eyes and read the number under your finger. Assume that the number is 235,035. Then go to page 235 in the telephone book and select the 35th name on that page. Repeat this process until you select all the participants constituting the sample.

A variant of random sampling that can be used when conducting a telephone survey is *random digit dialing* (Dillman, 2000). List all the exchanges in a particular area (the first three digits of the phone numbers, not including the area code). You then use the table of random numbers or a computer to select four-digit numbers (e.g., 5,891). The exchange plus the four-digit number provides the number to be called. (Any nonworking numbers are discarded.) This technique allows you to reach unlisted as well as listed numbers.

Even though random sampling reduces the possibility of systematic bias in your sample, it does not guarantee a representative sample. You could, quite at random, select participants who represent only a small segment of the population. In the golf ball example, you might select 50 orange golf balls. White and yellow golf balls, even though represented in the population, are not in your sample. One way to combat this problem is to select a large sample (such as 200 rather than just 50 balls). A large sample is more likely to represent all segments of the population than a small one. However, it does not *guarantee* that representation in your sample will be proportionate to representation in the population. You may end up with 90 white, 90 orange, and only 20 yellow golf balls in a sample of 200, although such a result is highly unlikely. In addition, as you increase sample size, you also increase the cost and time needed to complete the survey. Fortunately, more sophisticated techniques provide a random yet representative sample without requiring a large number of participants.

QUESTIONS TO PONDER

1. What is a representative sample and why is it important to have one for a survey?
2. What is a biased sample and how can a biased sample affect your results?
3. What is a random sample and why is it important to do random sampling?
4. What is simple random sampling?

Stratified Sampling Stratified sampling provides one way to obtain a representative sample. You begin by dividing the population into segments, or *strata* (Kish, 1965). For example, you could divide the population of a particular town into Whites, Blacks, and Hispanics. Next, you select a separate random sample of equal size from each stratum. Because individuals are selected from each stratum, you guarantee that each segment of the population is represented in your sample. Figure 9-6 shows the stratified sampling strategy. Notice that the population has been divided into two segments (gray and colored figures). A random sample is then selected from each segment.

illustrates the
e top of the
han pulling
mbers. The
to a number
ndom num-
).



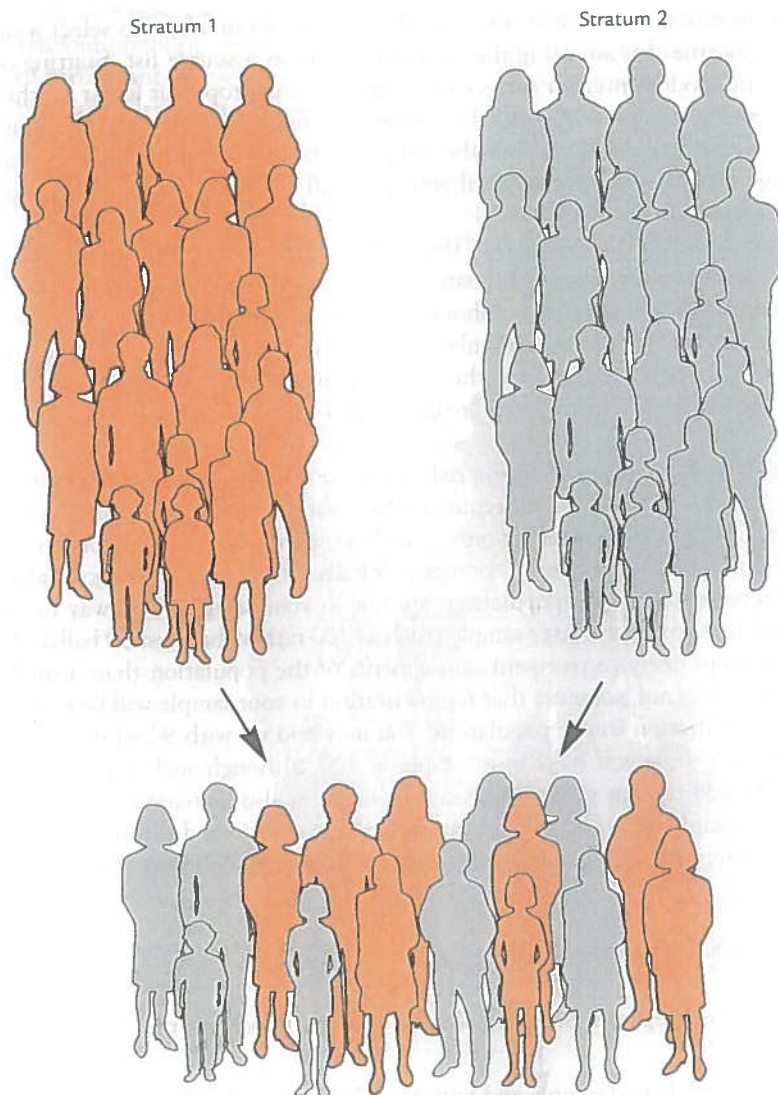


FIGURE 9-6 Example of stratified sampling. The population is divided into two strata from which independent random samples are drawn.

Proportionate Sampling Simple stratified sampling ensures a degree of representativeness, but it may lead to a segment of the population being overrepresented in your sample. For example, consider a community of 5,000 that has 500 Hispanics, 1,500 Blacks, and 3,000 Whites. If you used a simple stratification technique in which you randomly selected 400 people from each stratum, Hispanics would be overrepresented in your sample relative to Blacks and Whites, and Blacks would be overrepresented relative to Whites. You could avoid this problem by using a variant of simple stratified sampling called **proportionate sampling**.

In p
reflected
of 10%
60% Wh
have 120
techniqu
By t
has been
responde
populatio

Systematic
in conjun
tematic s
Acc
after a re
telephon
item (wh
effective
fourth it

FIGURE 9
included i

In proportionate sampling, the proportions of people in the population are reflected in your sample. In the population example, your sample would consist of 10% Hispanics ($500/5,000 = 10\%$), 30% Blacks ($1,500/5,000 = 30\%$), and 60% Whites ($3,000/5,000 = 60\%$). So, if you draw a sample of 1,200, you would have 120 Hispanics, 360 Blacks, and 720 Whites. According to Kish (1965), this technique is the most popular method of sampling.

By the way, stratification and proportionate sampling can be done after a sample has been obtained (Kish, 1965). You randomly select from the participants who responded the number from each stratum needed to match the characteristics of the population.

Systematic Sampling Systematic sampling is a popular technique that is often used in conjunction with stratified sampling (Kish, 1965). Figure 9-7 illustrates the systematic sampling technique.

According to Kish (1965), this technique involves sampling every k th element after a random start. For example, once you have randomly chosen the page of the telephone book from which you are going to sample, you then might pick every fourth item (where $k = 4$). Systematic sampling is much less time consuming and more cost effective than simple random sampling. For example, it is much easier to select every fourth item from a page than to select randomly from an entire list.

Richardson, E.	555-6396*
Richardson, J. B.	555-6789
Richardson, L. R.	555-2311
Richardson, M.	555-9902
Richardson, V.	555-7822*
Richeson, A. P.	555-8211
Richeson, T.	555-3762
Richey, B. B.	555-9943
Richey, C. L.	555-1470*
Richey, G. J.	555-8218
Richhart, W.	555-6539
Richman, A.	555-8902
Richman, B. I.	555-0076*
Richman, H. H.	555-9215
Richman, Z. L.	555-1093
Richmond, A.	555-7634
Richmond, B. B.	555-7890*
Richmond, C.	555-2609
Rideman, L.	555-7245
Ritchey, A. K.	555-6790

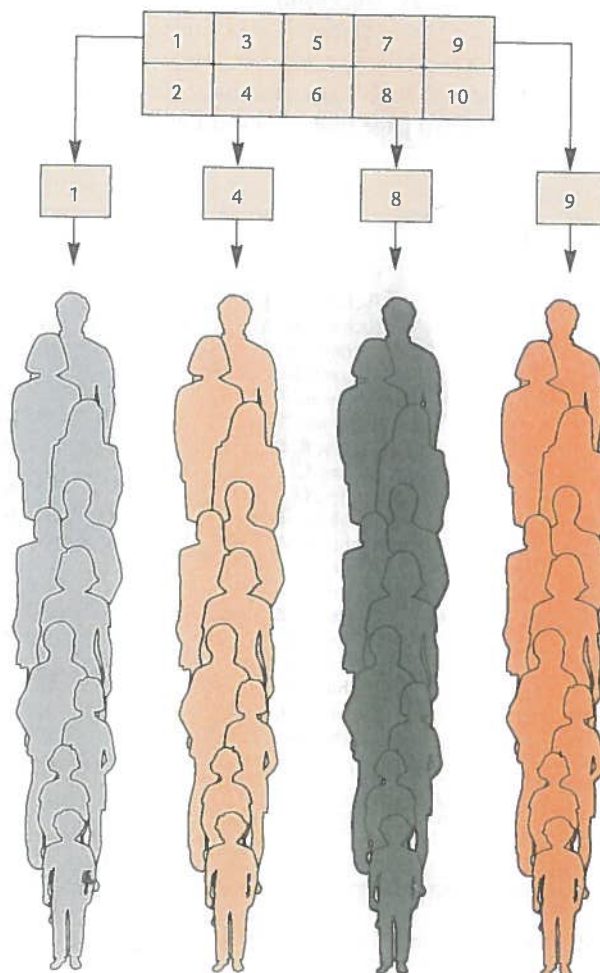
Each of the names with a star (*) would be included in your sample.

FIGURE 9-7 Example of systemic sampling. After a random start, every selected name is included in the sample (indicated with an asterisk).

Cluster Sampling In some cases, populations may be too large to allow cost-effective random sampling or even systematic sampling. You might be interested in surveying children in a large school district. To make sampling more manageable, you could identify naturally occurring groups of participants (clusters) and randomly select certain clusters. For example, you could randomly select certain departments or classes from which to sample. Once the clusters have been selected, you would then survey all participants within the clusters. **Cluster sampling** differs from the other forms of sampling already discussed in that the basic sampling unit is a group of participants (the cluster) rather than the individual participant (Kish, 1965). Figure 9-8 illustrates cluster sampling. This figure shows how you select four groups from a larger pool of groups.

An obvious advantage to cluster sampling is that it saves time. It is not always feasible to select random samples that focus on single elements (individuals, families, etc.). Cluster sampling provides an acceptable, cost-effective method of acquiring a

FIGURE 9-8 Example of cluster sampling. After selecting subgroups of the population, all participants in each subgroup are surveyed.



sample. On the negative side, cluster sampling does limit your sample to those participants found in the chosen clusters. If participants within clusters are fairly similar to one another but differ from those in other clusters, the sample will leave out important elements of the population. For example, clusters consisting of geographical areas of the United States (e.g., East, Midwest, South, Southwest, and West) may differ widely in political opinion. If only East and Midwest are selected for the sample, the opinions collected may not reflect the opinions of the country as a whole. Thus, cluster sampling does have drawbacks.

A variant of cluster sampling is **multistage sampling**. You begin by identifying large clusters and randomly selecting from among them (first stage). From the selected clusters, you then randomly select individual elements (rather than selecting all elements in the cluster). This method can be combined with stratification procedures to ensure a representative sample.

Other sophisticated sampling techniques are available to the survey researcher, but to explore them all would require a whole book. If you are interested in learning about these techniques, read Kish (1965).

Random and Nonrandom Sampling Revisited

In Chapter 6, we distinguished between random sampling (in which each member of a population has an equal chance of being selected) and nonrandom sampling (in which a limited group of potential participants is tapped). The sampling techniques we have just discussed may be used in the context of random or nonrandom sampling. Ideally, you would want to use random sampling. This is especially true, as noted in Chapter 6, if you want to make specific predictions about specific behaviors. However, as a practical matter, it may not always be possible to use a true random sample. Instead, you may have to administer your questionnaire to a convenience sample, such as students at a particular university, which is a nonrandom sample. Similarly, surveys conducted via the Internet use nonrandom samples, consisting only of those with computers who know how to access the Internet and have the ability to complete the survey. Of course, using a nonrandom sample limits the generality of your results, and making specific predictions about behavior may not be possible. However, a nonrandom sample (as noted in Chapter 6) is perfectly acceptable for most research interests in psychology. If you use nonrandom sampling, you should include a discussion of possible limitations of your results in the discussion section of any report that you write.

QUESTIONS TO PONDER

1. What are the various sampling techniques that represent modifications of simple random sampling?
2. Under what conditions would you use each of the sampling techniques discussed above?
3. What are the implications of using a nonrandom sample?

cost-
ed in
able,
only
nts or
then
other
f par-
e 9-8
larger

lways
nities,
ring a



Sample Size

One factor you must contend with if you perform a survey is the size of your sample. You should try to select an *economic sample*—one that includes enough participants to ensure a valid survey and no more. You must take into account two factors when considering the size of the sample needed to ensure a valid survey: the amount of acceptable error and the expected magnitude of the population proportions.

The question of acceptable error arises because most samples deviate to some degree from the population. If you conduct a political poll on a sample of 1,500 registered voters and find that 62% of the sample favors Smith and 38% Jones, you would like to say that 62% of the population favors Smith. However, these sample proportions do not exactly match those of the population (the population proportions may be 59% and 41%). This deviation of sample characteristics from those of the population is called **sampling error**.

When determining sample size, you must decide the acceptable amount of sampling error. Unfortunately, there are no broad rules of thumb as to the acceptable margin of error. It depends in part on the use to which you will put your results (Moser & Kalton, 1972). If you plan to apply your results to implement changes in behavior, you may want a small margin of error. If you are interested simply in describing a set of characteristics, you may tolerate a larger margin of error. A good way to determine the acceptable margin of error is to look at literature describing similar surveys to see what margin of error was used.

The second component you need to consider when determining sample size is the magnitude of the differences you expect to find. Here again, there is no broad rule of thumb to guide you. You can make use of previous surveys to get an estimate of the magnitude of the differences. Or you can conduct a small pilot survey to gain some insight into the magnitudes.

Once you have determined the acceptable error and the expected magnitude of differences, you can calculate the size of the sample needed. The calculation is relatively easy for simple random sampling. Moser and Kalton (1972) suggested the following formula:

$$n' = \frac{P'(1 - P')}{(SE_p)^2}$$

where P is the estimate of the proportion of the population that has a particular characteristic and SE_p is the acceptable margin of error. For example, if you expect 62% of the population to favor Smith in an election and your acceptable margin of error is 2% (0.02), then the formula gives $n = 589$. Thus, you should have 589 participants in your sample.

When the size of the population is large, you do not need to consider population size when calculating sample size. If the population is small, however, then you must use the *finite population correction* (fpc) when calculating sample size. Crano and Brewer (1986) suggest using the following formula when the sample size is more than 10% of the population size:

$$n = N \times n' / (N + n')$$

where $n =$
vious for
drawn. For

Thus, if the
participar

For s
random s
ability in
(the varia
for comp
plex. If yo
(1972) ar

QUEST

1. W
2. W
3. H
do
4. W
se

SUMM

Survey re
tudes of
research.
survey d
variety o
uct, to p

The
tionnaire
goal for
and burd
issues of

Ofte
characte
interest
when yo

Que
participa

where n = the corrected sample size, n' = the sample size calculated with the previous formula, and N = the size of the population from which the sample is to be drawn. For example, using the previous numbers and $N = 2,000$, you have

$$n = 2000 \times 589 / (2000 + 589) = 455$$

Thus, if the population from which your sample will be drawn consists of only 2,000 participants, you would use a sample size of 455 rather than 589.

For stratified sampling, determining sample size is more difficult than for simple random sampling. You must take into account the between-strata error (the variability in the scores of participants in different strata) and the within-strata error (the variability in the scores of participants within the same stratum). The formulas for computing sample size with the more sophisticated sampling techniques are complex. If you pursue survey research using these techniques, consult Moser and Kalton (1972) and Kish (1965) for more information.

QUESTIONS TO PONDER

1. What is meant by an "economic sample"?
2. What is sampling error and how do you know if you have an acceptable level?
3. How does the magnitude of the differences you expect to observe affect your decision about sample size?
4. What are some of the sample size issues you need to consider for different sampling techniques?

SUMMARY

Survey research is used to evaluate the behavior (past, present, and future) and attitudes of your participants. Survey research falls into the category of correlational research. Therefore, you cannot draw causal inferences about behavior from your survey data, no matter how compelling the data look. Surveys are used in a wide variety of situations. They can be used to research the marketability of a new product, to predict voter behavior, or to measure existing attitudes on a variety of issues.

The first step in a survey is to clearly define the goals of your research. Your questionnaire is then designed around those goals. You should have a reasonably focused goal for your survey. A questionnaire that tries to do too much may be confusing and burdensome to your participants. Keep your questionnaire focused on the central issues of your research.

Often a questionnaire is organized so that questions about your participants' characteristics (demographic items) and questions about the behavior or attitude of interest are included. The demographic items can later be used as predictor variables when you look for relationships among the variables that you measured.

Questionnaire items can be of several types. Open-ended questions allow your participants to answer in their own words. A major advantage of this type of question

is the richness of the information obtained. A drawback is that responses are difficult to summarize and analyze. A restricted question provides response categories for participants. A variation on the restricted item is a rating scale on which participants circle a number reflecting how they feel. This type of item yields data that are easier to summarize and analyze. However, the responses made to restricted items are not as rich as those obtained with an open-ended item. A partially open-ended item gives participants not only clearly defined response alternatives but also a space to write in their own response category.

Once you have decided what types of items to include on your questionnaire, you must then actually write your questions. When writing items, you should avoid using overly complex words when simpler words will suffice. Your questions should be precise. Vague or overly precise wording yields inconsistent data. In addition, you should avoid using words that are biased or judgmental.

A questionnaire is more than just a collection of questions. Questions should be presented in a logical order so that your questionnaire has continuity. Also, it is a good idea to place demographic items at the end. These questions tend to be boring, and participants may be turned off if you have demographic items at the beginning of your questionnaire. Sensitive questions should be placed toward the middle. Your participants may be more willing to answer such questions after answering several other, more innocuous questions. Sensitive items should be carefully worded. Your questionnaire should have a logical "navigational path." This path should lead the respondent through the questionnaire as if he or she were reading a book.

Constructing a questionnaire involves more than sitting down and writing a set of items. Developing a good questionnaire involves several steps, including assessing its reliability, or your questionnaire's ability to produce consistent results. One way to assess reliability is to administer your questionnaire (or parallel forms of the questionnaire) more than once. If the results are highly similar, the questionnaire is reliable. Another way to assess reliability is with a single administration of your questionnaire. The most common way to do this is to use a split-half method by which you divide your questionnaire in half (e.g., odd versus even items) and correlate the two halves. Two statistics used to evaluate split-half reliability are the Kuder-Richardson formula and coefficient alpha.

If you find low reliability, you can do several things to increase it. You can increase the number of items on your questionnaire, standardize administration procedures, make sure that you score questions carefully, and ensure that your items are clear, well written, and appropriate for your sample.

In addition to assessing reliability, you should evaluate the validity of your questionnaire. The term *validity* in this context refers to whether your questionnaire actually measures what you intend it to measure. There are three ways to assess validity. First, you can establish content validity by making sure that items on your questionnaire cover the full range of issues relevant to the phenomenon you are studying. Second, criterion-related validity can be established by correlating the results from your questionnaire with one of established validity. Third, you can establish construct validity by establishing that the results from your questionnaire match well with predictions made by a theory. No one of these methods is best. Perhaps the best approach is to establish validity using more than one of the three methods.

Five ways to administer your questionnaire are the mail survey, group administration, telephone survey, face-to-face interview and Internet survey. The mail survey is easiest. You simply mail your questionnaires and wait for a response. However, this method is plagued by nonresponse bias. Return rates can be increased with effective cover letters, follow-up reminders tailored to the nature of your participant population, and small rewards. In group administration, you give your questionnaire to a large number of participants at once. The advantage of group administration is that you can collect large amounts of data quickly. Surveys also can be conducted over the telephone. Questionnaires designed for telephone surveys should be relatively short, with clearly worded, short questions. Because your questions will be read to your participants, make sure that the person reading the questions speaks clearly and slowly. In an interview, you ask your questions to your participants in a face-to-face session. Interviews can be either structured (questions asked from a prepared questionnaire in a fixed order) or unstructured (each interview is different). Finally, you can conduct your survey on the Internet, which allows you to reach large numbers of potential respondents. Data can be collected quickly and easily via the Internet. However, the sample obtained from the Internet may not be representative, and you must have the equipment, resources, and knowledge necessary to post a questionnaire this way.

One of the most crucial stages of survey research is acquiring a sample of participants. Because you want to make statements about how people think on an issue, be sure your sample represents the population. Biased samples lead to invalid data and ultimately incorrect conclusions. Sampling techniques include simple random sampling (in which every participant has an equal chance of being in your survey) and stratified sampling (in which your population is broken into smaller segments and random samples are then drawn from those smaller segments). Other sampling techniques are proportionate sampling, multistage sampling, and cluster sampling. The sampling technique you use depends on the needs of your survey.

Whichever sampling technique you choose, you must consider the issue of sample size. Your sample should be large enough to be representative of the population, yet not too large. Try to acquire an economic sample that has just enough participants to adequately assess behavior or attitudes. The size of the most economic sample is determined with a special formula.

KEY TERMS

open-ended item	biased sample
restricted item	simple random sampling
partially open-ended item	stratified sampling
mail survey	proportionate sampling
nonresponse bias	systematic sampling
Internet survey	cluster sampling
telephone survey	multistage sampling
face-to-face interview	sampling error
representative sample	